Lipophilicity in Drug Action and Toxicology

Edited by
V. Pliška, B. Testa and H. van de Waterbeemd

Volume 4

VCH

Lipophilicity in Drug Action and Toxicology

edited by Vladimir Pliška, Bernard Testa and
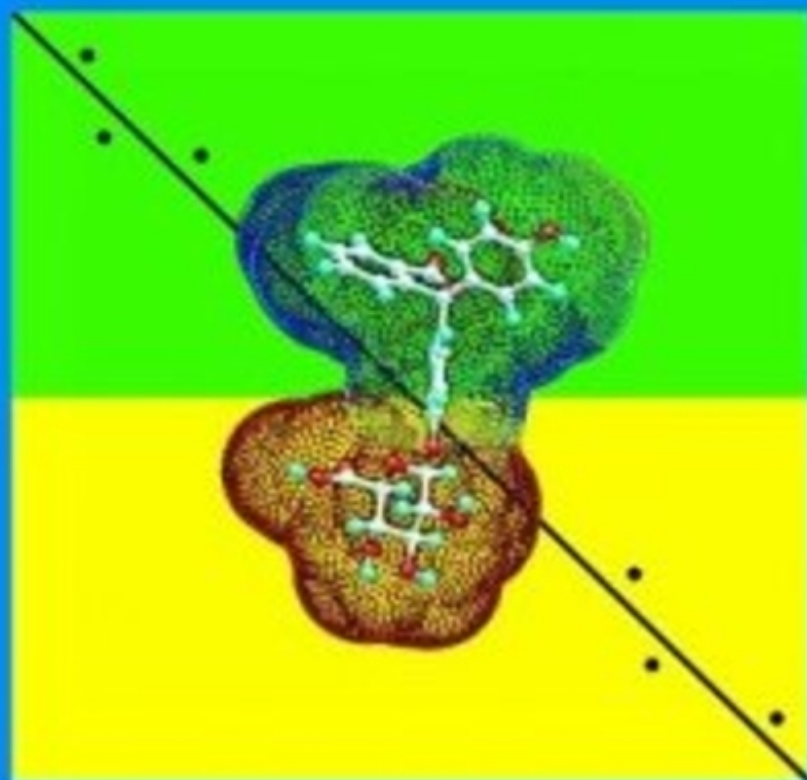Han van de Waterbeemd

VCH

Methods and Principles in Medicinal Chemistry

Edited by
R. Mannhold
H. Kubinyi
H. Timmerman

Lipophilicity in Drug Action and Toxicology

edited by
Vladimir Pliška
Bernard Testa
Han van de Waterbeemd

VCH

Weinheim · New York · Basel · Cambridge · Tokyo

Preface

After three volumes covering the various techniques quantifying the relations of biological activity and chemical properties of drug molecules, the fourth volume in the series "Methods and Principles in Medicinal Chemistry" focuses on the role of lipophilicity in drug action and toxicology.

Lipophilicity is well known as a prime physico-chemical descriptor of xenobiotics with relevance to their biological properties. The hydrophobic interactions of drugs with their receptors, the pharmacokinetic behaviour of drug molecules, toxicological properties and pharmaceutical aspects like solubility are examples of a steadily increasing number of topics in which lipophilicity plays an important role.

In keeping with the outstanding importance of lipophilicity in biosciences, this topic is treated in the present volume by more than twenty leading experts. The first out of five sections covers the physico-chemical background of molecular interactions and partitioning. The following two sections deal with the various experimental and computational approaches to quantifying lipophilicity. Experimental assessment includes partitioning as well as chromatographic alternatives. Computational procedures range from the classical approach employing hydrophobic fragmental constants to three-dimensional concepts which reflect the impact of conformational aspects of lipophilic behaviour. The last two sections reflect the relevance of lipophilicity in biological responses to xenobiotics and in drug design. Inter alia, the dependence of pharmacokinetic processes, like membrane transport and biotransformation on lipophilicity as well as environmental hazard assessment using lipophilicity data, deserve mention here. Lipophilicity scales for peptides and amino acids are discussed in their relation to drug design.

The present volume convincingly achieves its main objective, to put emphasis on lipopilicity as an important property for a vast number of biological processes.

December 1995

Düsseldorf   Raimund Mannhold
Ludwigshafen   Hugo Kubinyi
Amsterdam   Hendrik Timmerman

Methods and Principles in Medicinal Chemistry

Edited by
R. Mannhold
H. Kubinyi
H. Timmerman

A Personal Foreword

The idea to write a state-of-the-art monograph examining manifold aspects of lipophilicity was born at the 7th European QSAR Symposium in Interlaken (1988). The Organizing Committee suggested that the spared funds of the Symposium be utilized for organizing a workshop on lipophilicity which would establish a solid base for such a monograph, and challenged us to undertake the task. It took us six long years to fulfill the first part of our commitment. The *Symposium on Lipophilicity in Drug Research and Toxicology* was held in Lausanne in March 1995, and was only possible thanks to an additional support by numerous companies and organizations. The remaining task turned out to be even more difficult. To publish a book *so* specified meant to ask a number of authors for collaboration. The our great joy, the majority of the contacted persons accepted, wrote their chapters, and even delivered their manuscript in time. We thank them for their collaboration.

However, editing is a thankless task. The text of any book of this series should be generally comprehensible, thus assuming a more or less consistent style. In trying this, editors are put under pressure by both authors and publisher, for different reasons. On top of it, there are also series editors who have firm and generally justified notion about the style of the entire book series. There is also a different degree of bias by individual partners: in our estimate the highest one with the authors, who usually present their favorite child with a great deal of an understandable enthusiasm, and the lowest one with the publishers who ist obliged to consider – and to foresee – the general success of the publication. Editors and series editors may find themselves somewhere half-way in between. Provided that the final outcome does *not* have a character of conference proceedings, the editors are compelled to set up basic style rules, and to exercise a certain pressure on the authors to follow them. In most instances, this was benevolently understood and respected.

The book contains, besides purely methodological contributions and established physico-chemical concepts, also chapters which my seemingly touch the problem of lipophilicity only from afar, or which may rather be considered as dreams of the future. However, we are conviced that they have a rightful place in this book.

There are many persons, in additions to the participating authors, to whom we owe our thanks. To name only very few of them, Professor Jean-Luc Fauchère gave the spiritual, and also material, impulse to this book. The VCH editors, Dr. Thomas Mager and Dr. Michael Bär carried out all the burdens asociated with the preparation for printing and manufacturing. And our colleagues, the series editors Professors Raimund Mannhold, Hugo Kubinyi, and Hendrik Timmerman, were most helpful with their critical comments.

February 1996

Zürich   Vladimir Pliška
Lausanne   Bernard Testa
Basel   Han van de Waterbeemd

List of Contributors

Michael H. Abraham
The Department of Chemistry
University College London
20 Gordon Street
London WC1H 0AJ
UK

Alex Avdeef
Sirius Analytical Instruments Ltd.
Riverside
Forest Row Business Park
Forest Row
East Sussex RH18 5DW
UK

Štefan Baláz
Department of Biochemical Technology
Slovak Technical University
Radlinského 9
SK-81237 Bratislava
Slovakia

Frédéric Billois
Institut de Chimie Therapeutique
BEP
Université de Lausanne
CH-1015 Lausanne-Dorigny
Switzerland

Ronald T. Borchardt
Pharmaceutical Chemistry Department
University of Kansas
Lawrence, KS
USA

Philip S. Burton
Drug Delivery Research Systems
Upjohn Laboratories
Kalamazoo, MI
USA

Pierre-Alain Carrupt
Institut de Chimie Thérapeutique
BEP
Université de Lausanne
CH-1015 Lausanne-Dorigny
Switzerland

Harpreet S. Chadha
The Department of Chemistry
University College London
20 Gordon Street
London WC1H 0AJ
UK

Marvin Charton
Chemistry Department
Pratt Institute
Brooklyn, NY 11205
USA

Chao-Kun Cheng
Department of Mathematical Sciences
Virginia Commonwealth University
Richmond, VA
USA

Robert A. Conradi
Drug Delivery Research Systems
Upjohn laboratories
Kalamazoo, MI
USA

Karl Dross
C. and O. Vogt Institute for Brain Re-
search
Heinrich-Heine-Universität
Universitätstrasse 1
40225 Düsseldorf
Germany

Emanuel Escher
Département de Pharmacologie
Faculté de Médecine
Université de Sherbrooke
Sherbrooke (Quebec)
J1H 5N4
Canada

Jean-Luc Fauchère
Institut de Recherches SERVIER
11 rue des Moulineaux
92150 Suresnes
France

Holger Fischer
F. Hoffmann-La Roche Ltd.
Pharma Research New Technologies
Structure-Property Correlations Group
CH-4002 Basel
Switzerland

Gerd Folkers
Department of Pharmacy
ETH Zürich
Winterthurerstrasse 190
CH-8057 Zürich
Switzerland

Patrick Guillard
Institut de Chimie thérapeutique
BEP
Université de Lausanne
CH-1015 Lausanne-Dorigny
Switzerland

Pavel Hobza
J. Heyrovský Institute of Physical
Chemistry
Academy of Sciences of the Czech Re-
public
Dolejškova 3
182 23 Prague 8
Czech Republic

Manfred Kansy
F. Hoffmann-La Roche Ltd.
Pharma Research New Technologies
Structure-Property Correlations Group
CH-4002 Basel
Switzerland

Lemont B. Kier
Department of Medicinal Chemistry
Virginia Commonwealth University
Richmond, VA
USA

Albert J. Leo
Pomona College and BioByte Corpora-
tion
Claremont
California, CA 91711–0517
USA

Robert L. Lipnick
Office of Pollution Prevention and Tox-
ics
U.S. Environmental Protection Agency
Washington, DC 20460
USA

Giuseppe Lisa
Institut de Chimie Thérapeutique
BEP
Université de Lausanne
CH-1015 Lausanne-Dorigny
Switzerland

Raimund Mannhold
Department of Laser Medicine
Molecular Drug Research Group
Heinrich-Heine-Universität
Universitätstrasse 1
D-40225 Düsseldorf
Germany

Alfred Merz
Department of Pharmacy
ETH Zürich
Winterthurerstrasse 190
CH-8057 Zürich
Switzerland

Christophe Meyer
Institut National de la Recherche Agro-
nomique
BP 527
F-44026 Nantes Cedex
France

Serge Pérez
Institut National de la Recherche Agro-
nomique
BP 527
F-44026 Nantes Cedex
France

Vladimir Pliška
Department of Animal Science
Swiss Federal Institute of Technology
ETH Zürich
CH-8092 Zürich
Switzerland

W. Graham Richards
Physical and Theoretical Chemistry
Oxford University
Oxford
UK

Christoph Sonntag
C. and O. Vogt Institute for Brain Re-
search
Heinrich-Heine-Universität
Universitätstrasse 1
D-40225 Düsseldorf
Germany

Albert Taylor
Servier Research and Development
Fulmer Hall
Windmill Road
Fulmer
Slough SL3 6HH
UK

Bernard Testa
Institut de Chimie Thérapeutique
BEP
Université de Lausanne
CH-1015 Lausanne-Dorigny
Switzerland

Ruey-Shiuan Tsai
Institut de Chimie Thérapeutique
BEP
Université de Lausanne
CH-1015 Lausanne-Dorigny
Switzerland

Michael S. Tute
Chemistry Department
University of Kent at Canterbury
Canterbury
Kent
UK

Peter Vis
Servier Research and Development
Fulmer Hall
Windmill Road
Fulmer
Slough SL3 6HH
UK

Björn Wagner
F. Hoffmann-La Roche Ltd.
Pharma Research New Technologies
Structure-Property Correlations Group
CH-4002 Basel
Switzerland

Bernard Walther
Servier Research and Development
Fulmer Hall
Windmill Road
Fulmer
Slough SL3 6HH
UK

Han van de Waterbeemd
F. Hoffmann-La Roche Ltd.
Pharma Research New Technologies
Structure-Property Correlations Group
CH-4002 Basel
Switzerland

Peter Weber
Institut de Chimie Thérapeutique
BEP
Université de Lausanne
CH-1015 Lausanne-Dorigny
Switzerland

Rudolf Zahradník
J. Heyrovský Institute of Physical
Chemistry
Academy of Sciences of the Czech Re-
public
Dolejškova 3
182 23 Prague 8
Czech Republic

Contents

6   Centrifugal Partition Chromatography for Lipophilicity Measurements   89

R.-S. Tsai, G. Lisa, P.-A. Carrupt and B. Testa

17   Membrane Transport and Cellular Distribution 295
      Š. Baláž

18   Applications of a Solvation Equation to Drug Transport
      Properties 311
      M. H. Abraham and H. S. Chadra

1 Lipophilicity: The Empirical Tool and the Fundamental Objective. An Introduction

Vladimir Pliška, Bernard Testa and Han van de Waterbeemd

1.1 Setting the Scene

At the end of the 7th QSAR Symposium held in Interlaken in 1988, the organisers asked a number of participants which topics they felt should require greater attention in future meetings. The list of suggestions was indeed long and diverse. One subject, however, was mentioned almost unanimously, namely the pharmacological, toxicological, and pharmacokinetic significance of weak interactions in general and lipophilicity in particular.

This interest is understandable and legitimate. Weak interactions such as hydrogen bonds, van der Waals forces, hydrophobic effects, and charge transfer interactions are absolutely essential for molecular recognition and interactions in living systems. They underlie the formation of firmly determined molecular and supramolecular structures (for instance in biological macromolecules, membranes, etc.) and, at the same time, enable their amazing flexibility and adaptability. As a rule, several weak forces participate in any interaction occurring in a biological system. Due to their superposition, intermolecular and intramolecular complexes may exhibit a broad range of association constants from about 10^4 mol L^{-1} (enzyme-substrate complexes) to 10^{14} mol L^{-1} (polyvalent antibody-antigen complexes). Since biologically important macromolecules always contain a variety of polar and nonpolar sites, the role of polar and hydrophobic forces is of utmost significance in all processes of biological recognition.

Before going any further, it appears appropriate to comment on the words "hydrophobicity" and "lipophilicity" since they are used rather loosely and inconsistently in the literature. As discussed in greater details below, lipophilicity is a molecular property expressing the relative affinity of solutes for an aqueous phase and an organic, water-immiscible solvent. As such, lipophilicity encodes most of the intermolecular forces that can take place between a solute and a solvent. In contrast, hydrophobicity is a consequence of attractive forces between nonpolar groups (e.g., hydrocarbon chains and rings) and therefore is but one component of lipophilicity. Factorization of lipophilicity into its polar and hydrophobic components contributes considerably to our understanding of the nature of lipophilicity and its role in the biological world [1, 2].

1.2 Biological Aspects

The relation between lipid solubility and biological effects of drugs was recognized almost a century ago by Meyer [3] and by Overton [4]. Some decades later, Pauling discovered a relationship between lipophilicity and anesthetic potency in a series of chem-

ically heterogeneous compounds [5]. It soon became evident that a quantification, or even a description of lipophilicity in thermodynamic terms, is not practicable. Until now, only empirical scales of lipophilicity have been of importance in practice, some expressing the changes in free energy associated with solute transfer between two phases, others being dimensionless indices relating partitioning data of given solutes to a general standard. This latter approach is based on the assumption of linear free-energy changes and is represented by the Leffler-Grunwald operators [6]. It was, in fact, first employed by Hammett in 1935 to describe electronic properties of substituent groups attached to a fixed molecular backbone [7]. Later, Zahradnik and coworkers used responses obtained in two related biological systems to derive what is in fact, but not by name, a set of lipophilicity constants [8, 9]. Such attempts were not unique during the late 1950s and early 1960s. However, it is to the great credit of Hansch, Fujita and Leo that empirical constants can be readily used in pharmacology and toxicology [10, 11]. Besides deriving an extensive set of lipophilicity descriptors, the so-called π-values, Hansch and colleagues proved their apparent additive nature, thus establishing them as genuine substituent constants.

The structure and function of any biological system are closely related to the lipophilic properties of its component molecules. First, lipid-lipid interactions strongly influence the structure of biological membranes, and thereby the compartmentation of compounds within cell organelles. Second, transport and distribution processes within biological systems are to a large extent controlled by the lipophilicity of the system components. The highly hydrophobic interior of a bilayer membrane enables or facilitates the passage of lipophilic substances and prevents the free diffusion of polar molecules except water in and out of cells and organelles. By controlling both transport and compartmentation processes with some degree of selectivity, lipophilicity imposes an adjustable resistance to free diffusion, thus becoming the major obstacle to a random distribution of substances in biological systems, which would be entirely incompatible with life. The same is true for distribution within an organism where several physiological barriers control the access of endogenous and exogenous compounds to various organs and tissues. It is well established that the hemato-encephalic (blood-brain), placental and hemato-mammary (blood-mammary gland) barriers are of a very selective nature, so that specific transporter systems have to mediate the passage of vital compounds, the hydrophilicity of which prevents their passive membrane permeation.

Last, but not least, lipophilicity plays a dominating role in ligand-receptor interactions, e.g., in the binding of hormones, neurotransmitters, modifiers of cellular processes (e.g., growth, initiation, or repression factors) and drugs to their receptors. The same applies for enzyme-substrate, enzyme-inhibitor, antigen-antibody and other ligand-macromolecule interactions.

1.3 The Molecule in the Background

While molecular pharmacology deals with the response of a cell to a substance recognized as a message, medicinal chemistry attempts to unveil the semantics, and perhaps also the syntax, of the molecular language which encodes these messages. In order to achieve this, molecular structure has to be described in a pharmacologically relevant

way; adoption of a multilevel description of molecular structure [12] appears to be the best approach to this end. Such a description starts at a simple geometrical level, continues with a stereoelectronic one, and ends up at levels of intermolecular interactions. It is at these latter levels that one encounters properties like solubility and lipophilicity whose high content in structural information remains difficult to understand fully.

Lipophilicity, however, is far from being only an empirical tool in structure-activity analysis. It is also a unique probe that can be used to unravel the complex and dynamic interplay between intermolecular forces and intramolecular interactions in solutes of interest. The former comprise interactions between a solute and the aqueous and organic phases, namely [1, 2, 13]:

- Ion-ion and ion-dipole (permanent, induced) interactions (for ionic solutes);
- Charge transfer interactions;
- Hydrogen bonds (normal, reinforced);
- Van der Waals interactions (forces of orientation, induction, and dispersion);
- Hydrophobic bonds.

Intramolecular interactions that influence lipophilicity can be classified as follows:
- Through-bond electronic effects a) in aromatic systems, and b) across aliphatic segments;
- Through-space electronic/polar effects comprising a) internal electrostatic bonds (ionic bonds, H-bonds, and other electrostatic bonds), b) internal electrostatic repulsion, and c) collision of hydration spheres due to proximity effects between polar groups;
- Through-space steric/hydrophobic effects comprising a) internal hydrophobic bonds (hydrophobic collapse), and b) internal steric hindrance.

Intramolecular interactions can explain differences in lipophilicity seen between regioisomers and between configurational diastereomers. The interplay between conformational diastereomerism and lipophilicity, which is particularly manifest in molecular chameleons, is gaining increasing recognition in compounds of sufficient size and functional complexity [14]. Furthermore, intramolecular interactions affecting lipophilicity represent a major and incompletely understood challenge to the accuracy of current fragmental systems.

1.4 Some Pragmatic Aspects

1.4.1 Definitions and Symbols

At this point, we should make an explanatory comment concerning the expressions "hydrophobicity" and "lipophilicity". Their usage is not uniform. Semantically, they seem to stand for the same feature or object, and are therefore frequently considered to be synonymous. In the scientific use, however, their meaning is quite different. The following operational definitions have been suggested by the IUPAC [15, 16]:

- **Hydrophobicity** is the association of nonpolar groups or molecules in an aqueous environment which arises from the tendency of water to exclude nonpolar molecules.

● **Lipophilicity** represents the affinity of a molecule or a moiety for a lipophilic environment. It is commonly measured by its distribution behavior in a biphasic system, either liquid-liquid (e.g., partition coefficient in 1-octanol/water) or solid-liquid (e.g. retention on RP-HPLC or TLC, see section 1.4.2) systems.

Such definitions are by no means unambiguous and noncontroversial, as our knowledge of molecular mechanisms underlying these phenomena, although continuously growing, is still far from being complete. There are, however, pragmatic reasons for their (albeit tentative) differentiation, and it is therefore not astonishing that they occur with different frequency in languages used in different research disciplines. The term "hydrophobicity" is familiar to biophysicists working with X-ray diffraction, NMR spectroscopy and molecular models. It is used in connection with the description of the molecular surface of a compound in contact with an aqueous environment. "Lipophilicity" is a term mainly employed by medicinal chemists to describe transport processes of a compound in biological systems. Much confusion also exists in the symbols of lipophilicity parameters. To bring some clarity, we offer in Table 1 a compilation of useful symbols.

1.4.2 Experimental Techniques

A great step forward has been achieved since the pioneering work of Meyer [3] and of Overton on the partitioning of anesthetics in olive oil/water [4]. Hansch, Fujita and their coworkers chose the 1-octanol/water solvent system as an arbitrary standard for expression of lipophilicity [10, 11], and pioneered its measurement by the shake-flask technique. Most of the available data refer to this partitioning system. This standard technique, sometimes laborious and precarious, can efficiently and quite safely be sub-

Table 1. Lipophilicity parameters and their recommended symbols

Referred Symbol	Parameter	Alternatives
$\log P$	partition coefficient of neutral species[a]	$\log K$, $\log PC$
$\log P_{oct}$	$\log P$ for 1-octanol/water	
$\log P_{alk}$	$\log P$ for alkane/water	
$\log D$	distribution coefficient: "apparent" partition[a] coefficient	$\log P'$, $\log P_{app}$
$\log D_{oct}$	$\log D$ for 1-octanol/water	
$\log D_{alk}$	$\log D$ for alkane/water	
$\log P^{+}$	partition coefficient of cationic form	
$\log P^{-}$	partition coefficient of anionic form	
$\log P^{+/-}$	partition coefficient of zwitterionic form	
CLOGP	$\log P$ calculated by the CLOGP program[b]	
$\log k$	log of capacity factor in RP-HPLC	
$\log k_w$	$\log k$ extrapolated to 100 % aqueous eluent	

[a] $\log P$ and $\log D$ can be calculated one from the other using the appropriate correction for ionization [1].

[b] Reference [19].

stituted by various chromatographic techniques: thin-layer chromatography (TLC), reversed-phase high performance liquid chromatography (RP-HPLC or RPLC), and centrifugal partition chromatography (CPC), which are all employed routinely.

1.4.3 Computational Procedures

Lipophilicity has been expressed by means of manifold descriptors mainly based on partition coefficients or similar thermodynamic features. Relationships between individual scales are, apart from some exceptions, very close. These descriptors can be obtained by a number of computational routines; they receive attention in this book. Fragmental constants, i.e., contributions of individual molecular fragments to the overall value of a descriptor, are roughly additive and thus afford quick predictions of lipophilicity from molecular structures. In this way, some problems and limitations associated with the experimental assessment of substituent constants can be overcome [17]. From a visual point of view, this property has been simulated as a dynamic process, exhibiting the characteristics observed experimentally [18].

From a practical point of view, lipophilicity descriptors are important for at least two reasons. First, they may predict unsatisfactory drug candidates and avoid, in a simple way, an extensive experimentation. This relates to both transport properties and intrinsic activity of the potentially interesting substances. Second, they enable to investigate structure-property relationships, in particular intermolecular forces and intramolecular interactions. These relationships are of utmost importance in drug design. It would be, for example, of little use to design a highly hydrophilic substance if it is targeted to the central nervous system.

1.5 Objectives of the Book

In summary, lipophilicity is an essential property of molecules whose roles in biological systems are numerous and essential. Above all, it is intimately connected with regulatory pathways in living systems, and allows them to exist away from equilibrium. In so far as medicinal chemists and pharmacologists aim at sending messages (i.e., drugs) to ailing cells, they cannot avoid viewing lipophilicity as one of the most significant properties controlling both the delivery and the reception of the message.

The aim of this monograph is therefore rather straightforward, namely, to present the state-of-the-art of the area, to bring about a current insight necessary for interpretation of lipophilicity data, and to demonstrate how research in cell biology, pharmacology, medicinal chemistry, toxicology, and related fields can benefit from them. Our main interest, however, is to put emphasis on lipophilicity as an important property controlling a great many processes in living organisms.

References

[1] van de Waterbeemd, H., and Testa, B., *Adv. Drug Res.* **16**, 85–225 (1987)

[2] El Tayar, N., Testa, B. and Carrupt, P. A., *J. Phys. Chem.* **96**, 1455–1459 (1992)

[3] Meyer, H., *Arch. Exp. Pathol. Pharmakol.* **42**, 109–118 (1899)

[4] Overton, E., *Studien über die Narkose.* Fisher: Jena 1901

[5] Pauling, L., *Science* **139**, 15–21 (1961)

[6] Leffler, J. E., and Grunwald, E., *Rates and Equilibria of Organic Reactions, As Treated by Statistical, Thermodynamic and Extrathermodynamic Method* Wiley: New York 1963, p. 22–27

[7] Hammett, L. P., *Chem. Rev.* **17**, 125–136 (1935)

[8] Zahradnik, R. and Chvapil, M., *Experientia,* **16**, 511–512 (1960)

[9] Zahradnik, R., *Arch. Int. Pharmacodyn.* **85**, 311–329 (1962)

[10] Hansch, C., and Fujita, T., *J. Am. Chem. Soc.* **86**, 1616–1626 (1964)

[11] Hansch, C., and Leo, A., *Substituent Constants for Correlation Analysis in Chemistry and Biology.* Wiley: New York 1979

[12] Testa, B., and Kier, L. B., *Med. Res. Rev.* **11**, 35–48 (1991)

[13] El Tayar, N., Tsai, R. S., Testa, B., Carrupt, P. A., and Leo, A., *J. Pharm. Sci.* **80**, 590–598 (1991)

[14] Carrupt, P. A., Testa, B., Bechalany, A., El Tayar, N., Descas, P. and Perrissoud, D., *J. Med. Chem.* **34**, 1272–1275 (1991)

[15] *Glossary of Terms Used in Medicinal Chemistry,* IUPAC, in press

[16] *Glossary of Terms Used in Computational Drug Design,* IUPAC, in press

[17] Rekker, R. F., *The Hydrophobic Fragmental Constant. Its Derivation and Applications. A Means of Characterizing Membrane Systems.* Elsevier: Amsterdam 1977

[18] Kier, L. B., Chen, C.-K., Testa, B., and Carrupt, P.-A., *Pharmaceut. Res.* **12**, 1–6 (1995)

[19] Leo, A. J., *Chem. Revs.* **93**, 1281–1306 (1993)

2 Lipophilicity: A History

Michael S. Tute

Abbreviations

AM1	Molecular orbital program (Dewar)
CLOGP	Program for lipophilicity calculation
HPLC	High performance liquid chromatography
HYDRO	Program for lipophilicity calculation on flexible solutes
ISA	Isotropic (non-polar) surface area
SASA	Solvent accessible surface area
SCAP	Program for lipophilicity calculation using solvent-dependent conformations
TLC	Thin layer chromatography

Symbols

D	Distribution coefficient
f	Hydrophobic fragmental constant
P	Partition coefficient, refers usually to octanol/water
S	Molar solubility
S_A	Surface area
S_i	Atomic surface area
T_m	Melting point
V	Solute volume
α_H	Solute H-bond donor strength
β_H	Solute H-bond acceptor strength
Δq_i	Atomic partial charge
π	Hydrophobic substituent constant
π^*	Solute polarity/polarisability
π_H	Enthalpic component of hydrophobic substituent constant
π_s	Entropic component of hydrophobic substituent constant
σ	Hammett electronic substituent constant

2.1 Introduction

Lipophilicity is usually expressed by the partition coefficient ($\log P$), a molecular parameter which describes the partitioning equilibrium of a solute molecule between water and an immiscible lipid-like organic solvent. By convention, the ratio of concentrations in the two phases is given with the organic phase as numerator, so that a positive value for $\log P$ reflects a preference for the lipid phase, and a negative value reflects a

relative affinity for water. Also by convention, where ionizable molecules are concerned, log$P$ refers to the neutral species whereas what is actually measured may be the distribution coefficient, log$D$. The distribution coefficient refers to the ratio of total concentrations of ionized and unionized species across both phases.

Many workers have emphasized that the value of log$P$ depends largely on interactions made by the solute with the water phase, either being repelled by water (hydrophobic effect) or solvated by water through hydrogen bonds or other polar forces (hydrophilic effect). Such emphasis has encouraged use of the term hydrophobicity, and in medicinal chemistry and particularly for QSAR the substituent constants, π, and fragment constants, f, are almost universally described as hydrophobic substituent parameters or hydrophobic fragmental constants.

Use of the term hydrophobicity has also been dependent on a perception of the thermodynamics of partitioning of strictly nonpolar solutes such as the aliphatic and aromatic hydrocarbons between water and a lipid phase, and on a particular use of the term "hydrophobic bonding" to describe the tendency of nonpolar groups to associate in aqueous solution, thereby reducing the extent of contact with neighboring water molecules. As discussed by Némethy [1], the formation of such "hydrophobic bonds" has long been considered to be driven by an entropy effect: the water molecules become more ordered around exposed nonpolar residues, and when the hydrophobic "bond" is formed, the order decreases, resulting in a favorable entropy and hence free energy of formation. For over 30 years, it has been commonly supposed that the "hydrophobic" interaction between nonpolar side chains of a protein, associated with formation and breakdown of layers of abnormal water, makes a prominent contribution to the stability of the native, folded form. The existence, nature, and effect of "hydrophobic hydration" is today a subject of intense controversy (see sction 2.4.2).

Use of the term hydrophobicity by the Hansch group [2], who in 1964 pioneered the use of octanol/water as the standard solvent pair for measurement, could also be justified on the grounds that this particular solvent pair is such that polar effects are similar in each phase. Both water and octanol have hydroxyl groups that can participate in polar interactions with the solute molecule, and moreover there is a considerable amount of water within the octanol phase. So, an octanol/water log$P$ value will emphasize differences in hydrocarbon interactions with water and with lipid, but tend to hide differences in the interaction of polar and hydrogen-bonding groups.

Recent studies have clearly and repeatedly shown that log$P$ in general incorporates two major contributions, namely a "bulk" term reflecting both hydrophobic (entropic) and dispersion (enthalpic) effects, and electrostatic terms reflecting hydrogen bonds and other dipole-dipole effects. Moreover, the emphasis on interaction of the solute with the water phase has been challenged; more emphasis has now been placed on enthalpic interactions within the lipid phase; free energy simulations have been carried out and thermodynamic measurements have been made to better understand fundamental interactions of the solute with each phase. As a result, traditional explanations of partitioning in terms of "hydrophobic bonding" have had to be reconsidered.

2.2 Measurement of Lipophilicity

The partition coefficient was first defined in 1872 by Berthelot and Jungfleisch [3], who wrote *"On the Laws that Operate for the Partition of a Substance between two Solvents"*. It was first used to correlate and explain the potencies of biologically active substances at the turn of the century, by both Meyer [4] and Overton [5] in their studies of narcotic compounds. Overton's work stimulated other investigations of the use of partition coefficients for biological correlations, among them a study by Seidell [6] in 1912. Believing that the partition coefficient of thymol might be relevant to a study of the mode of action of thymol against hookworm, Seidell made measurements using a variety of lipid phases, including olive oil, castor oil, peanut oil, and linseed oil. In those days, measurement was particularly tedious: it was necessary to separate thymol from the oil by a steam distillation, and then to estimate thymol in water by treatment with bromine, titrating the resulting hydrobromic acid produced!

With the development of UV spectroscopy, measurement of the partition coefficient for compounds with strong absorption, nonextreme values, and sufficient solubility in the aqueous phase has become routine, using the "shake-flask" method, partitioning between one of a wide variety of lipid phases, and water or an appropriate buffer solution as the aqueous phase. For many ionizable compounds, compounds of low solubility, and compounds with low UV absorbance or extreme values of partition coefficient then special methods of measurement or alternative lipophilicity parameters have had to be devised.

In 1959, Gaudette and Brodie [7] realized both the possibility for using a partition coefficient to model lipophilic character, and the relevance of lipophilicity to pharmacokinetic processes. They found a parallel between the heptane/buffer partition coefficients of certain drugs, and their rate of entry into cerebrospinal fluid. However, generalised use of log$P$ as a lipophilicity parameter did not come about until after 1964, with the Hansch octanol/water system remaining to this day the standard for both experimental and theoretical investigations. In 1971, Leo, Hansch and Elkins [8] published the first comprehensive review of partition coefficients, with a tabulation of nearly 6000 values, including their own measurements on some 800 in the octanol/water system. The review incorporated an account of the shake-flask method of measurement, which was discussed more exhaustively in a 1973 monograph by Purcell, Bass and Clayton [9].

Octanol/water log$P$ has also been measured by high-performance liquid chromatography [10], and by using a filter-probe to sample selectively from the aqueous or lipid phase so there is no need to fully separate the phases [11, 12]. For ionizable compounds, Brandström in 1963 [13] was first to use a potentiometric titration technique. One aqueous phase titration, with a pH-meter probe, was carried out in the aqueous phase to determine pK_a. A second titration was carried out in the presence of octanol, when partition occurred and the pK_a shifted. The difference in pK_a was related to log$P$. In 1974 Seiler [14] modified this technique so as to determine pK_a and log$P$ from a single titration. The technique has now been refined to enable not only simultaneous pK_a and log$P$ determination, but to allow treatment of substances with multiple ionization constants, ion-pair partitioning, and self-association reactions leading to the formation of oligomers [15, 16].

CH$_2$OCOC$_8$H$_{17}$

CHOCOC$_8$H$_{17}$

Me

PGDP

CH$_2$OCOC$_n$H$_{2n+1}$

CHOCOC$_n$H$_{2n+1}$

CH$_2$OPOCH$_2$CH$_2$NMe$_3$
$\overset{O^-}{\underset{O}{||}}$
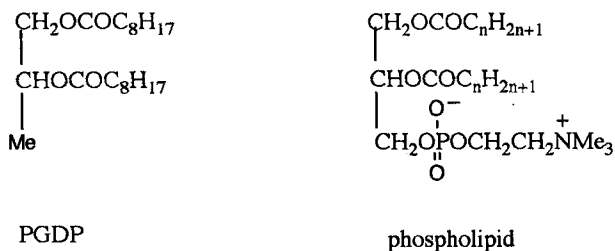
phospholipid

Figure 1. Structural similarities between propylene glycol dipelargonate (PGDP) and a phospholipid molecule.

Lipophilicity has, since 1964, been traditionally measured in the octanol/water system. However, for particular purposes and for particular sets of compounds, other solvent pairs have been used. Octanol/water values have been shown to be generally satisfactory for modeling serum protein binding and for modeling lipophilic interactions with biological membranes consisting largely of protein, but for other types of membrane then a different solvent system might be more appropriate. In 1989, Leahy et al. [17] suggested that membranes (or receptors) could exist with very different hydrogen bonding characteristics from those of octanol. Thus, membranes may contain neither acceptors nor donors (modeled by an alkane); or contain largely amphiprotic groups (as in a protein, modeled by octanol); largely proton donor groups (which may be modeled by chloroform); or largely proton acceptor groups (as in a phospholipid membrane). Leahy argued for the use of propylene glycol dipelargonate (PGDP) as lipid phase to model phospholipid membranes (Fig. 1) and have accordingly measured many partition coefficients in the PGDP/water system [18].

For many compounds, the traditional equilibrium method of partition coefficient measurement may be impossible, impractical, or inappropriate. As a practical alternative to log$P$, particularly for biological correlations, much use has been made of parameters derived from chromatographic retention. In 1941 Martin and Synge [19] showed that for reversed phase thin-layer chromatography, Eq. (1) relates partition coefficient, P, to the ratio, R_f, of distances moved by the compound spot and the solvent front in a given time, with K being a constant for the system. In 1950 Bate-Smith and Westall [20] defined the parameter R_m as in Eq. (2) from which Eq. (3) follows. In practice, excellent correlations have been found between R_m and log$P$ taking the form of Eq. (4). Kaliszan [21] has reviewed the use of lipophilicity parameters derived from HPLC, TLC, and paper chromatography.

$$P = K\left[\left(\frac{1}{R_f}\right) - 1\right] \tag{1}$$

$$R_m = \log\left[\left(\frac{1}{R_f}\right) - 1\right] \tag{2}$$

$$R_m = \log P - \log K \tag{3}$$

$$R_m = a\log P + b \tag{4}$$

2.3 Calculation of Lipophilicity

2.3.1 Substitution Method

The Hansch group were the first to point out [2] in their influential paper of 1964, that the octanol/water log$P$ value of simple benzenoid derivatives could be calculated by a method bearing close analogy to the Hammett [22] treatment of chemical reactivity, including ionization, of substituted benzene derivatives. Hammett had shown in the 1930s that the equilibrium or rate constant of parent (unsubstituted) molecule, K_H, and the equilibrium or rate constant for a substituted compound, K_X, could be correlated by

$$\log \left(\frac{K_X}{K_H} \right) = \varrho \sigma_X \tag{5}$$

which could be rewritten as

$$\log K = \varrho \sigma_X + \log K_H \tag{6}$$

The substituent constant σ_X refers to the electronic effect of the substituent and is a parameter applicable to many different reactions (characterized by different values of ϱ) whose rate depends on the degree of electron release or withdrawal by the substituent. For the derivation of σ constants, the ionization of benzoic acids was defined as the standard reaction for which ϱ was set to unity. In analogous fashion to the Hammett treatment, Hansch defined substituent constants, π, by Eq. (7), choosing octanol/water as the standard system. Then, by analogy to Eq. (6) for reaction rates or equilibria, Eq. (8) could be used to calculate log$P$.

$$\log \left(\frac{P_X}{P_H} \right) = \pi_X \tag{7}$$

$$\log P_X = \log P_H + \pi_X \tag{8}$$

Just as Hammett had found that different σ values were required for para- and for meta-substituents on a benzoic acid, because of differing contributions of field and resonance effects on reactivity, so the Hansch group immediately recognized that different π values would be required according to the environment of the substituent. Electronic effects in particular would alter the interaction of a polar substituent with the water phase: consider 4-nitrophenol, where neither the hydroxyl group nor the nitro group would behave towards water or towards octanol in like fashion to the hydroxyl or nitro group in phenol itself, or in nitrobenzene. It was rapidly appreciated that the lipophilicity parameter, log$P$, was only to a first approximation an additive property: it has considerable constitutive character. This at first proved to be a major difficulty for the calculation of lipophilicity, but in fact opened the way to using lipophilicity measurements to probe a variety of intramolecular effects, including not only electronic but steric effects, so-called proximity effects when polar groups share a solvation shell, hydrogen bonding, and conformation (sometimes called folding effects, see chapter 4).

2.3.2 Fragment Additivity Method

The π-system was used for some 15 years, but was destined to give way to a much more general fragmentation method of calculating log$P$. The substituent scheme was only applicable, in general, to substituted benzene derivatives. For other compounds, the problem immediately arose, what does one take as "parent" and what as substituent? Moreover, rather serious errors occurred in the application and interpretation of lipophilicity calculations using the substituent approach. Hansch and Anderson [23] in 1967 suggested that the difference in calculated log$P$ and in measured log$P$ (which was lower) in compounds of the type $C_6H_5CH_2CH_2CH_2X$ indicated a folding of the alkyl chain, so that substituent X interacted with the aromatic ring through "intramolecular hydrophobic bonding". In 1973 Nys and Rekker [24] suggested that the difference did not arise from any intramolecular folding, but in fact arose because of the implicit neglect of the lipophilicity of hydrogen. The application of Eq. (7) to calculate log$P$ for the compounds above requires the addition:

$$\log P(C_6H_5 - CH_2 - CH_2 - CH_2 - X) = \log P(C_6H_6) + 3\pi(CH_3) + \pi(X)$$

and makes no distinction between the lipophilicity of CH_3 or CH_2.

Nys and Rekker [24, 25] then suggested a totally different approach to log$P$ calculation, which was to transform our understanding. This approach was based on the assignment of "fragmental constants", f, to a selection of structural fragments, the calculated log$P$ then being simply the sum of fragment values appropriate to the molecule plus a number of interaction factors, F, that were necessary to correct for intramolecular electronic or steric interactions between fragments. The fragment system is expressed by Eq. (9):

$$\log P = \sum_{i=1}^{m} f_i + \sum_{i=1}^{n} F_i \tag{9}$$

Rekker used a large database of published log$P$ values to derive both fragment values and correction factors statistically. His first book on the method was published in 1977 [26] and refinements were later made by Rekker and de Kort in 1979 [27] using a database of over one thousand log$P$ measurements. A second book in 1992 by Rekker and Mannhold [28] includes further refinements and example calculations.

A feature of Rekker-type calculations as currently implemented is that many of the correction factors, F, are considered to be multiples of a so-called "magic constant", C_M, the latest value for which is 0.219 [28]. The calculation of lipophilicity therefore follows Eq. (10) with, for example, a proximity correction of kn (key number) equal to 2 for a two-carbon separation of polar groups:

$$\log P = \sum_{i=1}^{m} f_i + \sum kn.C_M \tag{10}$$

There has been much speculation as to whether the "magic constant" has any fundamental significance, Rekker having proposed that it might be related to a quantum displacement of water in the first solvation shell around the solute.

Not long after Rekker had published his hydrophobic fragmental system (based on a "reductionist" principle – the statistical analysis of a large database), Leo et al. [29] devised a fragmental system based on a "constructionist" principle, that is, based on the selection of just a small set of well-validated measurements on relatively simple molecules. In 1979, Hansch and Leo [30] went on to publish a book on their version of fragment additivity, outlining the method and giving sample calculations. By this time, though, the method (particularly the application of correction factors) had become extremely complex, and a computer program was clearly required. The program, CLOGP, was duly written to take a structural formula input, and output a $\log P$ value, along with some indication of confidence in the value, and a breakdown of fragments used and correction factors applied. The current version of this program is widely recognized as the "industry standard" for calculation of $\log P$. An account of the history of $\log P$ calculations was given in 1985 by Dearden [31] and in 1993 Leo [32] reviewed the status of $\log P$ calculations based on fragment additivity and empirical correction terms. By this time, the MASTERFILE database maintained by the Hansch group at Pomona College in Claremont, California, contained over 40 000 $\log P$ values measured in over 300 solvent systems, including over 18 000 in octanol/water. Details of the fragment additivity method, including the all important methods of assigning the various correction factors, have been published by Leo [32, 33].

2.3.3 Fragmentation into Atoms

For its fragments, the CLOGP method uses a basis set of "isolating" carbon atoms (aliphatic and aromatic taking different values), attached hydrogen atoms, and a variety of "polar" fragments, again having different values for an attachment to an aromatic or aliphatic structure. Of course, fragmentation of a molecule is somewhat arbitrary, and there are advantages and disadvantages of any fragmentation scheme. Fragments larger than a single atom can be selected, so that significant electronic interactions are contained within one fragment, and this is perceived as the main advantage of using fragments larger than single atoms. The advantage of using an atomic fragmentation approach is that ambiguities are avoided, but a disadvantage is that a very large number of atom types are needed to describe a reasonable range of molecules, unless atomic charges are calculated to distinguish between various electronic forms of the same, or similarly hybridised, atom. In 1984, Broto et al. [34], in 1986 Ghose and Crippen [35], and in 1989 Viswanadhan et al. [36] implemented atomic level schemes of additivity. Ghose and coworkers extended and refined their atom values in 1987, and suggested that fewer atom types may be needed, provided that a separate parameter for atomic charge were to be included [37]. They suggested that use of a molecular orbital, CNDO/2 calculation, may be appropriate.

Atom level fragment schemes work well in many instances, but a common shortcoming, pointed out by Leo, is the failure to deal with long-range interactions such as found in p-nitrophenol [32].

2.3.4 Molecular Orbital Calculations

Many workers have been attracted by the possibility of calculating molecular orbital indexes that should be relevant to the differences in solvation energy between water and octanol. The first such effort was by Rogers and Cammarata in 1969 [38], who developed a correlation equation for octanol/water log$P$ of just 30 simple aromatic solutes using charge density and induced polarization as calculated parameters, suggesting that solvation by the aqueous phase was charge-controlled, whereas solvation by the octanol phase was polarizability controlled. This idea was followed up by Klopman and Iroff [39] in 1981, and by Zavoruev and Bolotin [40] in 1982, who also found charge density calculations useful in simple series of solutes. The Klopman study of 61 solutes indicated that terms for bulk as well as charge were necessary. As many as 10 independent variables were required, including the number of hydrogen, carbon, nitrogen, and oxygen atoms; the sums of squared charges on carbon, nitrogen, and oxygen atoms, and indicator variables for the presence of functionalities such as acid or ester, nitrile, and amide.

Coming now to 1989, Bodor and coworkers [41] developed a regression model for prediction of octanol/water log$P$ in which charge density on nitrogen and oxygen atoms was found important, a dipole moment term was included, and "bulk" was represented by descriptors of surface area, volume, and molecular weight. They extended the Klopman set of solutes to include 118 miscellaneous aliphatic, aromatic and heteroaromatic compounds, with functional groups embracing alcohols, phenols, amines, ethers, and amides; and even included a few complex drug molecules such as atropine and tetracycline! The most important result to emerge from this work was that surface area (see section 2.3.5) and volume were the most significant descriptors. Although the standard error and correlation indicate a good predictive power, the method, needing the intermediate computation of 15 regression parameters, on a structure previously fully optimized using the AM1 procedure, must rank as too unwieldy for general predictive use.

2.3.5 Calculations Based on Surface Area

In 1949, Collander [42] showed that for solutes of a similar structure (homologous series, similar hydrogen bond donating, or accepting properties) log$P$ for a solute in one solvent, e.g., octanol, could be related to its log$P$ in another solvent, e.g., chloroform, by a linear equation as shown in Eq. (11).

$$\log P_{octanol} = a + b \log P_{chloroform} \tag{11}$$

Collander relationships have often been invoked to estimate log$P$ values in one system from values measured in another.

In 1978, Dunn and Wold [43] analyzed data for 26 solutes in 6 different solvent systems by principal components analysis, and established that there are two fundamental components which contribute to the partition coefficient. They suggested that one component is associated with the chain length in homologous series and may be a mo-

lar volume or surface area effect, while the second is clearly the result of a polar inter-action between solute and solvent.

In discussing the concept of "hydrophobic bonding", Hermann in 1972 suggested and later found a linear relationship between solubility of hydrocarbons in water, and the surface area of the cavity they form [44]. Earlier, in 1971, Lee and Richards had in-troduced the concept and showed how to calculate "solvent accessible surface area" (SASA) to describe quantitatively the relationship of proteins to solvent [45]. The SASA was defined as the area of the surface traced out by the center of a probe sphere, representing a water molecule, as it moves around the solute molecule, just touching its van der Waals surface. In 1974, Chothia [46] found linear relationships between SASA and free energy changes for the solvent transfer of protein residue side chains. In 1976, Yalkowsky and Valvani calculated SASA to estimate partition coefficients of hydrocarbons [47].

Following on from these studies, Moriguchi and coworkers [48] in 1985 made use of SASA (S_A) together with empirical correction terms for polar moieties (S_H), to core-late with log$P$ for 138 miscellaneous compounds with a correlation coefficient of 0.995! In Eq. (12), note that since S_A is the total surface area (but not including hydrogen atoms) and not the hydrophobic area, the S_H parameter implies both a correction for hydrophilic surface area, and the effect of specific hydration of the polar moiety. The method was able to reproduce differences of log$P$ between geometrical isomers, not calculable by other means. It therefore opened the way to the calculation of a log$P$ for different conformations.

$$\log P = 1.90 S_A - 1.00 \sum S_H - 1.06 \tag{12}$$

$$n = 138, r = 0.995, s = 0.13$$

In 1987, Dunn and coworkers [49, 50] made a highly significant contribution by defin-ing the possible structure of hydrated complexes of solute molecules containing polar groups. They then measured an isotropic surface area, ISA, as the SASA associated with the nonpolar portion of the hydrated solute, and as a separate parameter, the SASA associated with the hydrated surface area was calculated and expressed as a fraction, f(HSA) being the ratio of hydrated surface area and total surface area of the hydrated solute molecule. Principal component analysis on the log$P$ values of 69 sol-utes in 6 solvent systems extracted two factors explaining almost all variance in parti-tioning, and these factors were proposed to be ISA and f(HSA). Linear regression equations were then developed to estimate log$P$ in each solvent system. Interestingly, the coefficient in the ISA term was the same for each solvent system, the coefficient in the f(HSA) term differed considerably for each system, but was not statistically signifi-cant in predicting octanol/water or ether/water log$P$! This led to the suggestion that in these particular solvents, *the solute may partition as the hydrated molecule*. This sugges-tion is most reasonable, in the light of recent studies on the so-called "water-dragging" effect of many solutes, which can increase the concentration of water in the nonpolar phase at equilibrium [51].

Camilleri and coworkers in 1988 [52] explored the possibility of using surface area to predict partition, and considered over 200 benzene derivatives containing a variety of

substituents and functional groups, such as alkyl, hydroxyl, alkoxyl, amino, ester, ketone, etc. They chose to use the Connolly surface of the molecule [53], which is also obtained by rolling a sphere over the solute surface, but which is defined as the sum of contact and re-entrant surfaces, the re-entrant surface being the inner surface of the solvent probe as it comes in contact with two or more atoms of the solute. Each molecule was considered as a combination of fragments defined as, e.g., aromatic hydrocarbon, saturated hydrocarbon chain, OH group, NH_2 or NH group, carbonyl group, etc. The regression Eq. (13) was solved, to find coefficients a_n where A_n values are the measured surface areas of the various components of the molecule:

$$\log P = a_0 + a_1 A_1 + a_2 A_2 + \cdots\cdots a_n A_n \tag{13}$$

For interpolative prediction of $\log P$, Eq. (13) was deemed as accurate as the CLOGP program. Moreover, it is clearly applicable to different conformations, especially those involving a shielding of certain parts of a molecule from solvent interaction. When applied to predict $\log P$ for paracyclophane (Fig. 2), for example, CLOGP gives a value of 5.79, whereas a value of 4.83 was calculated by the Camilleri method, compared with experimental measurements of 4.33 by "shake-flask" and 4.61 by HPLC. Whereas the Camilleri method is clearly applicable to different conformations, it was parameterized using conformationally rigid compounds, to avoid any problem of having to estimate conformation, and to avoid making the invalid assumption that structures do not change their conformation on passing from one phase to another.

In 1991, Kantola and coworkers presented an atom-based parameterization, using atomic contributions to surface area, S_i, atomic numbers, N, and net charges Δq_i, associated with each atom and with the molecule in a defined conformation [54]. They thereby computed a conformationally dependent lipophilic quantity, p, which is only equal to the macroscopic property, $\log P$, if only one conformer (or a rigid compound) is involved in each phase. For Eq. (14), parameters, α, β, γ were obtained by regression analysis on a set of 90 rigid compounds:

$$p = \sum_i a_i(N)S_i + \beta_{1i}(N)S_i(\Delta q_i)^2 + \gamma_i(N)q_i \tag{14}$$

Considering all compounds, containing carbon, nitrogen, oxygen, halogen, and hydrogen atoms and by using AM1-computed geometries and charges, a quite reasonable correlation coefficient of 0.92 was achieved.
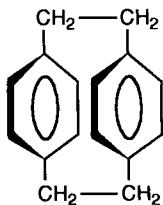
Figure 2. Paracyclophane

The next advance to be made in computing conformationally dependent lipophilicities will involve a determination of thepopulation of each conformation in both phases. Partition coefficients will then need to be computed by summation over all conformations. Some progress towards this goal has been made by Richards and coworkers who in 1992 developed the HYDRO program [55]. This program is intended to compute log$P$ for conformationally flexible molecules, but has so far been tested on only a few uncharged linear dipeptides. The method is the first to consider explicitly the effects of the population of accessible conformational minima in both phases. The partition coefficient for each dipeptide was calculated from the energy change on moving the relevant gas phase conformations into water and into octanol. These energies were calculated by using solvation contributions based upon solvent accessible surface area, and two sets of empirical parameters, the initial gas phase conformations being generated by systematic search, molecular mechanics.

2.4 The Nature of Lipophilicity

The physiochemical nature of lipophilicity can bediscerned from relations between log$P$ and other physical properties, whether measured or calculated. Perusal of relationships between the empirical correction factors used in fragment additivity schemes, and actual structure in solution as determined by spectroscopic methods, can also give clues.

Clearly, lipophilicity depends on the relative solvation energies between water and the lipid (octanol) phase. More can be learned by study of each phase separately, than by study of complex partition systems with solutes bearing both nonpolar and polar residues, where it is never easy to ascribe an effect to one or the other phase, or to a combination of effects.

One of the first computerized methods proposed for the estimation of log$P$ was SCAP (solvent-dependent conformational analysis) developed by Hopfinger and Buttershell [56] in 1976. For simple solutes, a solvation energy was computed for each phase separately, using solvation shell parameters and applying these to a molecular mechanics model of the solute.

A modern equivalent to a SCAP computation, but much more computationally intense, is the free energy perturbation method, whereby a partition coefficient difference can be calculated for simple solutes between pure solvents. Thus, Richards and coworkers in 1989 [57] used molecular dynamics simulations and the free energy perturbation method to compute the *difference* in log$P$ between methanol and ethanol, partitioned between water and carbon tetrachloride. Calculated and experimental values for this difference agreed to within 0.06 log$P$ units.

Unfortunately, the high content of water in the octanol phase (2 mol L^{-1} at equilibrium) does not lend it to detailed simulation. A simulation taking place in a "box" of solvent would have to contain enough solvent molecules to determine whether the water was evenly distributed or tended to cluster round, or hydrate, the solute. The flexibility of the octanol molecule (six rotatable bonds) would greatly increase the simulation time if using many explicit solvent molecules. Octanol was chosen for QSAR studies for a number of reasons, one of which was its high water content and its ability to hy-

drogen bond, so as to model a biological membrane better than a pure lipid solvent such as carbon tetrachloride or benzene.

2.4.1 Relation to Other Molecular Properties

Since $\log P$ reflects the difference in solvation energy between water and the lipid phase, it is to be expected that correlation would be found between $\log P$ and water solubility. Hansch, Quinlan and Lawrence in 1968 found linear relationships with $\log S_w$ for a wide range of liquids [58]. In 1980, Yalkowsky and Valvani extended this to solids, by including melting point in a linear relationship [59]. By 1991, Suzuki [60] had developed an estimating system for both partition coefficient and solubility. Eq. (15) covers a wide range of $\log P$ and molar solubility, S, values in a set of 348 liquids and 149 solids. T_m is the melting point in °C; for liquids T_m is set equal to 25.

$$\log 1/S = 1.050 \log P + 0.00956(T_m - 25) - 0.515 \tag{15}$$

$$n = 497, r = 0.976, s = 0.505$$

If the important feature of the partitioning process is the formation of "iceberg" or "ordered" water about a solute, one would expect linear relationships between the "bulk" of the solute (surface area, see section 2.3.5, or volume) and free energy of partitioning, with corrections for polar interactions between the solute and water. Many correlations with volume have been presented, often being given in support of the "iceberg" theory of the hydrophobic effect [61]. In a seminal paper published in 1977, Cramer [62] looked in detail at volume/solvation relationships for hydrocarbons and for rare gases. He pointed out that whereas solubility of hydrocarbons in water decreased linearly with molecular volume, the solubility of rare gases in water actually increases with volume, in direct contradiction to the then current views of "hydrophobic" bonding. He discussed some other discrepancies in the popular "iceberg" theory of hydrophobic bonding, pointing also to the oft-overlooked importance of interaction with the lipid phase in determining $\log P$. In the partitioning of a methylene moiety from water into octanol, Cramer gave the values of $\Delta G = -0.54$ kcal/mol for favorable solvation by octanol, and only $\Delta G = +0.18$ for unfavorable solvation by water. The overall process of transfer of a methylene is thus much better characterized energetically as lipophilic, than as hydrophobic – whatever the nature of hydrophobic forces might be.

Cramer offered an explanation of his results in terms of a cavity model of solvation. In a cavity model, energies are calculated first to create a solvent cavity, and then for interactions between the solvent and the solute molecule within the cavity. The first component is always a repulsive energy, which increases directly with solvent bulk (surface area or volume). This repulsive component is opposed by an attractive solvation energy which is proportional to solute polarizability, and polarizability in turn is proportional to molecular volume. Hence, it is possible to have both positive and negative slopes for a relationship between partition coefficient and volume.

The cavitation model is very attractive, and the focus on polarizability has been renewed recently by the "solvatochromic" approach to $\log P$, first proposed by Kamlet

and coworkers in 1977 [63–65] and discussed in the review by Leo [32] in 1993. Eq. (16) expresses the relationship between $\log P$ and a solute volume term, V, a polarity/polarizability term π^*, and independent measures of solute hydrogen-bond acceptor strength β_H and hydrogen-bond donor strength α_H.

$$\log P = aV + b\pi^* + c\beta_H + d\,\alpha_H + e \tag{16}$$

Leo has commented that once Eq. (16) has been well established, its real value will not be in calculating $\log P$, but in the understanding it affords us of the relative contributions of solute size, polarizability/polarity and H-bond acceptor strength. Solute H-bond donor strength does not seem important for the octanol/water system, since the H-bond acceptor strength of water and octanol are about equal.

Honig, Sharp and An-Suei Yang in 1993 reviewed macroscopic models of aqueous solutions [66]. Their discussion was in terms of the free energy of formation of cavity interfaces, which will be governed by the cohesive forces of the solvent. In this sense, all aqueous interfaces (solute-water interfaces), whether the solute is nonpolar or polar, are "hydrophobic" in that there will always be some force, surface tension, acting to minimize this interfacial area. This begs the question, what is the origin of surface tension?

2.4.2 Thermodynamics of Partitioning

2.4.2.1 Phase Transfer

Dearden [31] pointed out in his review in 1985 that a thermodynamic analysis of partitioning cannot explain the partitioning process fully, because each parameter (enthalpy, entropy) reflects the difference between behavior in each phase, and tells us nothing directly about the absolute contributions in each phase. It is only by the thermodynamic investigation of gas-water and gas-lipid (octanol) partition that the contribution of each phase can be properly assessed. In terms of the driving force for partitioning from water to octanol, Cramer's investigations discussed above [62], and a wealth of experimental solvation data collected since [67, 68], have indicated that the dominant contributions for nonpolar solutes come from interaction with the lipid phase. In partitioning, then, it is not "hydrophobicity" but lipophilicity which is the driving force.

Studies on the role of packing interactions in stabilizing the folded form of proteins have led to the same conclusion regarding the relative importance of interactions in the aqueous phase, and interactions (packing) in the folded state.

In 1990 Dill [69], in an excellent review of the forces responsible for protein folding, concluded that the dominant force is "hydrophobic" but only provided that the term is operationally defined in terms of the *transfer* of nonpolar side chains from water into a nonpolar environment.

In 1992, Sneddon and Tobias [70] carried out a molecular dynamics simulation of the thermodynamics of interconverting isoleucine and valine side chains in the core of the protein, ribonuclease T_1. They concluded that burial of nonpolar side chains in the in-

terior of the protein is favorable not so much because of the aversion of nonpolar groups towards water ("hydrophobicity") but rather because these groups can participate in favorable packing interactions (enthalpic, van der Waals interactions) within the core of the folded protein. So, with protein folding as with partitioning of solutes between aqueous and lipid phases, the emphasis is still on lipophilicity as the driving force.

Dearden's review [31] covered pertinent thermodynamic investigations of simple solute partitioning up to 1985. In 1990, Burgot and coworkers [71] investigated the thermodynamics of octanol/water partitioning of a series of β-blockers (Fig. 3), ranging in log$P$ from 0.16 (atenolol) to 3.37 (propranolol). They measured the enthalpy of transfer, and derived entropy from the independently measured partition coefficients. Apart from one compound, sotalol, which contains the highly polar sulfonamide group, transfer was dominated by the entropy term. This result was in accordance with the earlier thermodynamic investigations by Weiland *et al.* [72] of binding to the β-receptor itself, which lead them to suggest that antagonist binding is entropy-driven, whereas agonist binding is enthalpy-driven. The quite reasonable expectation of opposite behavior (in terms of enthalpy-versus entropy-driven binding) of agonists and antagonists is not, however, consistently observed in all studies [73, 74].

An intriguing study of the thermodynamics of partitioning of an extensive series of substituted benzoic acids, between octanol and water, was made in 1992 by Da, Ito and Fujiwara [75]. From van't Hoff plots (note that Beezer, Hunter and Storey [76] have pointed out that errors may be considerable when thermodynamic quantities are measured by use of van't Hoff plots), enthalpies and entropies of transfer were obtained,
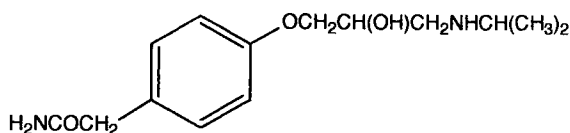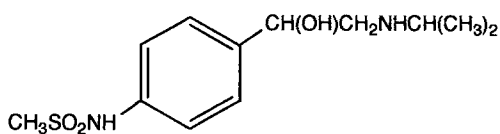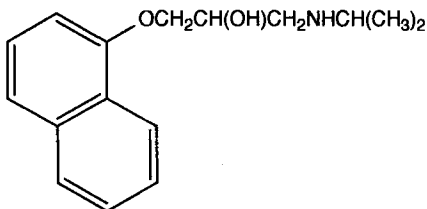
Figure 3. Selected β-blockers. log P values increase from top to bottom.

and the original Hansch π constant was separated into π_H (enthalpic) and π_S (entropic) parts according toEq. (17):

$$\pi = \pi_H + \pi_S \tag{17}$$

For the great majority of substituents, theenthalpic component is dominant, that is $| \pi_H | > | \pi_S |$. However, for alkyl and higher alkoxy substituents, entropy dominates, so $| \pi_S | > | \pi_H |$. There is no correlation between π_H and π_S. The study shows clearly that substituent constants are not determined exclusively by enthalpy or by entropy, but that the two terms contribute cooperatively. Furthermore, use of separate enthalpy or entropy parameters rather than free energy parameters was shown in several instances to enhance the correlation coefficient in QSAR equations.

2.4.2.2 The Aqueous Phase and the "Hydrophobic Bond"

The attraction of two nonpolar groups for one another in water has a strong parallel with the process of partitioning of a nonpolar solute between water and a nonpolar solvent, and it is widely believed that this force, the "hydrophobic effect" or "hydrophobic bonding" is due to some special properties of water. Controversy persists over both the physical origin of this attraction, over the properties and extent of "ordered" water that is widely believed to exist in the vicinity of nonpolar solutes, and over the use of the term "hydrophobic" or "hydrophobic bond" to describe the phenomenon. Kauzmann [77] in 1959 was the first to propose that the aversion of nonpolar groups for water is the most important factor in stabilizing the folded state of proteins, and he introduced the term "hydrophobic bond" to describe the apparent attraction between two nonpolar groups in water. Use of such a term was criticized by Hildebrand [78] in a letter to the *Journal of Physical Chemistry* in 1968. Hildebrand pointed out that the noun, "bond", was inappropriate because the apparent attraction between alkyl groups in water has none of the features which distinguish a chemical bond from van der Waals forces. Further, he regarded "hydrophobic" inappropriate, on the grounds that alkyl chains in micelles of soap are not bonded together because of phobia for surrounding water, for they stick together just as strongly in the absence of water. Hildebrand suggested one speaks simply of *alkyl interaction free energy*, or *entropy*. The net free energy of solution of hydrocarbons in water at room temperature is dominated by a large, negative entropy term, believed to arise from the increased ordering of water molecules in the immediate vicinity of the nonpolar solute.

Némethy, Scheraga and Kauzmann [79] replied to this criticism by saying that they did not wish to argue on a point of nomenclature, for the term "hydrophobic bond" had proved useful as shown by its frequent occurrence in physical, chemical, and biochemical nomenclature. Moreover, because of the entropy factor, the water-hydrocarbon system does differ in a unique manner from other systems of low miscibility – so why not use a unique term to describe it?

So what explanations have been offered, at the molecular level, for this unique behavior? The review by Dill [69] in 1990 can usefully be consulted, and also in 1990 there was published an excellent discussion by Taylor [80], who goes into the various interpretations that have been put forward to explain Cramer's [62] results.

The traditional explanation of Némethy, Scheraga and others is that water molecules around the nonpolar solute arrange themselves like an "iceberg" or "flickering cluster", an ordered state of low energy, low entropy, with good water-water hydrogen bonds. When two nonpolar residues approach one another, two solvent shells merge into one, with release of some ordered water molecules to "bulk" water, consequent increase in entropy, and decrease in free energy. This view has frequently been challenged as over-simplistic, and attention has been drawn to the role of attractive (enthalpic) forces in water which cannot be dismissed, and to the contribution to entropy changes made by flexible solutes (alkyl chains) which are more restricted in solution – any solution – than they are in the gas phase [80].

A new view of the hydrophobic effect was advanced by Muller in 1990 [81, 82] and in 1992 a review by Muller [83] summarized the way in which divergent opinions on the nature of the effect have arisen. According to Muller, all available data are consistent with the idea that some structural reorganization of water adjacent to nonpolar groups does occur, and that this is indeed responsible for the folding of proteins and drives the association of nonpolar groups. However, the organization of this "hydration shell water" is not iceberg-like. Muller's view, in essence, is that one may distinguish between "hydration shell" and "bulk" water molecules in the following way: hydration shell H-bonds are enthalpically stronger than bulk H-bonds, but a greater fraction of them is broken. This view leads to a neat explanation for the well-known effect of urea in increasing the solubility of hydrocarbons larger than ethane. It is postulated that urea reduces hydrophobic hydration by two mechanisms [82]. First, it occupies space in the "hydration shell" that could otherwise accommodate water molecules, and second, it alters the contribution of van der Waals interactions to the enthalpy of solvation.

2.5 Lipophilicity and Biological Activity

It is considered that our understanding of the role of lipophilicity in drug action is the single most significant result to come from the development of QSARs over the past three decades [84].

Quantitative correlations with lipophilicity have been established with regard to relative potencies at receptor sites, the regulation of drug transport, protein binding, pharmacokinetics, toxicity, drug metabolism, and enzyme induction.

Such correlations have been established in the fields of human drug research embracing both pharmacodynamics and chemotherapy. Many studies have been made that are of interest in the development of pesticides, weedkillers, and other agrochemicals. A field of quite recent interest is that of ecological toxicity, studies having been made of the role of lipophilicity in soil adsorption, environmental toxicity – particularly toxicity to species of fish – and environmental persistence (cf. Chapter 19).

Many histories of QSAR in drug research have been written [31, 85–88] and all refer back to Overton's studies at the turn of the century. Overton [5] noticed correlations between oil-water partition coefficients and narcotic potencies in tadpoles, and concluded that narcosis was due to physical changes affected in the lipid constituents of cells. In particular, Overton reasoned that in order to understand the action of an anesthetic in man, knowledge would be required of its ability to penetrate cells, the fat con-

tent of the animal, and the partition coefficient between water and the relevant lipid! Though there were various other investigations in the next 60 years, these seemed concerned only with narcotic or anesthetic activity. The much wider relevance and applicability of lipophilicity only became appreciated following the seminal studies of Corwin Hansch in the early 1960s.

In 1962, Hansch [89] defined the "hydrophobic" substituent constant, π, and showed for a series of substituted phenoxyacetic acids that variation in biological activity (concentration to induce a 10 % growth in a plant cell culture) could be described by Eq. (18), parabolic in π, and linear in the Hammett constant, σ. So started, in the unlikely field of botany, the era of QSAR.

$$\log(1/C) = 4.08\pi - 2.14\pi^2 + 2.78\sigma + 3.36 \tag{18}$$

By 1964, the Hansch group had settled on the octanol/water system as the standard for measuring partition coefficients, and had described further examples of parabolic relationships between relative biological activity and lipophilicity. This sometimes involved lipophilicity of the complete molecule, expressed as $\log P$, and sometimes involved variation in lipophilicity at a substituent position, as in the example above. Reasons for the parabolic relationship were put forward. Mathematical modeling, by Penniston and coworkers in 1969 [90] of the transport of molecules through a series of membranes, supported the expectation of a parabolic relationship between the probability of a molecule traversing a given number of lipid barriers in a given time, and its $\log P$ value. In 1977, Kubinyi [91] put forward both kinetic and equilibrium models to justify the expectation of bilinear relationships to describe drug transport in terms of $\log P$. Many bilinear equations have now been found. Investigations into the role of lipophilicity in drug transport were reviewed in 1990 by Dearden [92].

Both parabolic and bilinear relationships allow one to derive the optimum value of $\log P$ for transport to a given location, within the time of a biological assay. Evidence for an optimum in lipophilicity for CNS depressants was found by 1968 [93]. Hansch was then able to assert that in order for drugs to gain rapid access to the CNS, they should preferably have a $\log P$ value near 2.0. Subsequently, studies on anesthetics, hypnotics, and other CNS agents have been made and have given birth to the "Principle of Minimal Hydrophobicity in Drug Design" [87]. The thrust of this is that to keep drugs out of the CNS, and thereby avoid CNS-related side effects such as depression, weird dreams, and sedation, one should design drugs so that $\log P$ is considerably lower than 2.0. This ploy has been successful in the new generation of non-sedative antihistamines.

That we require drugs to have *lower* rather than higher lipophilicity depends also on other observations made over the past 30 years. Many studies on plants, animals, fish, various organelles such as liver microsomes, and enzymes have shown a linear increase in toxicity or inhibitory action in a series of compounds as $\log P$ or π increases [94].

A very high lipophilicity should also be avoided because of adverse effects on protein binding, and on drug absorption, including solubility [95].

Linear, and sometimes parabolic relationships have been found between lipophilicity and drug metabolism, either in whole animals, in liver microsomes, or by specific enzymes such as cytochrome P-450. Metabolism can be undesirable for two reasons: it may limit drug bioavailability, or it may produce toxic metabolites [95].

The ideal drug candidate, going into human studies, should have already been designed with the idea of keeping lipophilicity as low as possible, provided that this can be done without great loss of affinity to the target receptor. The receptor will commonly be the substrate binding site, or perhaps an allosteric site on an enzyme, or some control site on a cell surface. The receptor may therefore be part of a protein, and ligands may bind deep in "pockets" or on the surface of that protein. Just as the coefficients in a parabolic or bilinear QSAR can give information on the optimum lipophilicity for transport, the coefficient in the π term of a linear QSAR developed for a receptor (protein) can be diagnostic of the mode of receptor binding. Evidence from studies of enzyme inhibitors suggests that the coefficient in π is near 0.5 when a substituent "contacts" a surface, but near 1.0 when engulfed in a pocket. Many examples come from enzymes whose structures and binding sites have been established by X-ray crystallography [96].

Over the past 20 years, the Hansch group has collected into their database some 6000 sets of data, with attendant QSAR equations, from physical organic chemistry, medicinal chemistry, and toxicology. By 1993 [97] this database contained about 3000 biological QSARs, only 15 % of which lack a term for lipophilicity! Lipophilicity is clearly a major determinant of biological activity.

References

[1] Némethy, G., *Angew. Chem. Int. Ed. Engl.* **6**, 195–280 (1967)
[2] Fujita, T., Iwasa, J., and Hansch, C., *J. Am. Chem., Soc.* **86**, 5175–5180 (1964)
[3] Berthelot, M., and Jungfleisch, E., *Ann. Chim. Phys.* **26**, (4) 396–407 (1872)
[4] Meyer, H., *Arch. Exp. Pathol. Pharmacol.* **42**, 109–118 (1899)
[5] Overton, E., *Studien über die Narkose.* Fischer: Jena 1901
[6] Seidell, A., *Orig. Com. 8th Intern. Congr. Appl. Chem.* **17**, 85–89 (1912)
[7] Gaudette, L. E., and Brodie, B. B., *Biochem. Pharmacol.* **2**, 89–96 (1959)
[8] Leo, A., Hansch, C., and Elkins, D., *Chem. Rev.* **71**, 525–616 (1971)
[9] Purcell, W. P., Bass, G. E., and Clayton, J. M., *Strategy of Drug Design: A Guide to Biological Activity.* Wiley: New York 1973
[10] Mirlees, M. S., Moulton, S. J., Murphy, C. T., and Taylor, P. J. J., *J. Med. Chem.* **19**, 615–619 (1976)
[11] Tomlinson, E. J., *J. Pharm. Sci.* **71**, 602–604 (1982)
[12] Hersey, A., Hill, A. P., Hyde, R. M., and Livingstone, D. J., *Quant. Struct.-Act. Relat.* **8**, 288–296 (1989)
[13] Brandström, A., *Acta. Chem. Scand.* **17**, 1218–1224 (1963)
[14] Seiler, P., *Eur. J. Med. Chem.* **9**, 663–665 (1974)
[15] Comer, J., *Chem. in Britain* **30** (12), 983–986 (1994)
[16] Avdeef, A., *J. Pharm. Sci.* **82** (2), 1–8 (1993)
[17] Leahy, D. E., Taylor, P. J., and Wait, A. R., *Quant. Struct.-Act. Relat.* **8**, 17–31 (1989)
[18] Leahy, D. E., Morris, J. J., Taylor, P. J., and Wait, A. R., *J. Chem. Soc. Perkin Trans.* **2**, 723–731 (1992)
[19] Martin, A. J. P., and Synge, R. L. M., *Biochem. J.* 35, 1358–1386 (1941)
[20] Bate-Smith, E. C., and Westall, R. G., *Biochim. Biophys. Acta* **4**, 427–440 (1950)
[21] Kaliszan, R., *J. Chromatogr.* **220**, 21–83 (1981)
[22] Hammett, L. P., *Physical Organic Chemistry.* McGraw-Hill: New York 1940

[23] Hansch, C., and Anderson, S. M., *J. Org. Chem.* **32**, 2583–2586 (1967)

[24] Nys, G. G., and Rekker,R. F., *Chim. Ther.* **8**, 521–535 (1973)

[25] Nys, G. G., and Rekker, R. F., *Eur. J. Med. Chem.* **9**,361–375 (1974)

[26] Rekker, R. F., *The Hydrophobic Fragmental Constant,* Pharmacochemistry Library, Vol. I, W.Th. Nauta and R. F. Rekker, eds., Elsevier, Amsterdam, 1977

[27] Rekker, R. F., and de Kort, H. M., *Eur. J. Med. Chem.* **14**, 479–488 (1979)

[28] Rekker, R. F., and Mannhold, R., *Calculation of Drug Lipophilicity,* VCH, Weinheim, 1992

[29] Leo, A., Jow, P. Y. C., Silipo, C., and Hansch, C., *J. Med. Chem.* **18**, 865–868 (1975)

[30] Hansch, C., and Leo, A., *Substituent Constants for Correlation Analysis in Chemistry and Biology,* Wiley, New York, 1979

[31] Dearden, J. C., *Environmental Health Perspectives* **61**, 203–228 (1985)

[32] Leo, A., *Chem. Rev.* **93** (4), 1281–1306 (1993)

[33] Leo, A., Methods of calculating partition coefficients. In *Comprehensive Medicinal Chemistry,* Vol. **4**, C. Hansch, P. G. Sammes, J. B. Taylor (Eds.). Pergamon Press. Oxford; 295–319 (1990)

[34] Broto, P., Moreau, G., and Vandycke, C., *Eur. J. Med. Chem.* **19**, 71–78 (1984)

[35] Ghose, A., and Crippen, G. J., *J. Comput. Chem.* **7**, 565–577 (1986)

[36] Viswanadhan, V., Ghose, A., Revankar, G., and Robins, R. J., *J. Chem. Inf. Comput. Sci.* **29**, 163–172 (1989)

[37] Ghose, A. K., Pritchett, A., and Crippen, G. M., *J. Comput. Chem.* **9**, 80–90 (1987)

[38] Rogers, K. S., and Cammarata, A., *Biochim. Biophys. Acta* **193**, 22–29 (1969)

[39] Klopman, G., and Iroff, L. D., *J. Comput. Chem.* **2**, 157–160 (1981)

[40] Zavoruev, S. M., and Bolotin, V. A., *Khim. Farm. Zh.* **16**, 1361–1364 (1982)

[41] Bodor, N., Gabanyi, Z., and Wong, Chu-Kkuok., *J. Am. Chem. Soc.* **111**, 3783–3786 (1989)

[42] Collander, R., *Acta Chem. Scand.* **3**, 717–747 (1949)

[43] Dunn, W. J., III, and Wold, S., *Acta Chem. Scand.* **B 32**, 536–542 (1978)

[44] Hermann, R. B., *J. Phys. Chem.* **76**, 2754–2759 (1972)

[45] Lee, B., and Richards, F. M., *J. Mol. Biol.* **55**, 379–400 (1971)

[46] Chothia, C., *Nature* **248**, 338–339 (1974)

[47] Yalkowsky, S. H., and Valvani, S. C., *J. Med. Chem.* **19**, 727–728 (1976)

[48] Iwase, K., Komatsu, K., Hirono, S., Nakagawa, S., and Moriguchi, I., *Chem. Pharm. Bull.* **33**, 2114–2121 (1985)

[49] Dunn, W. J. III, Koehler, M. G., and Grigoras, S., *J. Med. Chem.* **30**, 1121–1126 (1987)

[50] Koehler, M. G., Grigoras, S., and Dunn, W. J., III, *Quant. Struct.-Act. Relat.* **7**, 150–159 (1988)

[51] Fan, W., El Tayar, N., Testa, B., and Kier, L. B., *J. Phys. Chem.* **94**, 4764–4766 (1990)

[52] Camilleri, P., Watts, S. A., and Boraston, J. A., *J. Chem. Soc. Perkins Trans.* **2**, 1699–1707 (1988)

[53] Connolly, M. L., *Science* **221**, 709–713 (1983)

[54] Kantola, A., Villar, H. O., and Loew, G. H., *J. Comput. Chem.* **12**, 681–689 (1991)

[55] Richards, N. G. J., Williams, P. B., and Tute, M. S., *Int. J. Quantum Chem.* **44**, 219–233 (1992)

[56] Hopfinger, A. J., and Battershell, R. D., *J. Med. Chem.* **19**, 569–575 (1976)

[57] Essex, J. W., Reynolds, C. A., and Richards, W. G., *J. Chem. Soc. Chem. Commun.,* 1152–1154 (1989)

[58] Hansch, C., Quinlan, J. E., and Lawrence, G. L., *J. Org. Chem.* **33**, 347–350 (1968)

[59] Yalkowski, S. H., and Valvani, S. C., *J. Pharm. Sci.* **69**, 912–922 (1980)

[60] Suzuki, T., *J. Computer-Aided Molecular Design* **5**, 149–166 (1991)

[61] Leo, A., Hansch, C., and Jow, P. Y. C., *J. Med. Chem.* **19**, 611–615 (1976)

[62] Cramer, R. D., III, *J. Am. Chem. Soc.* **99**, 5408–5412 (1977)

[63] Kamlet, M., Abboud, J. L., and Taft, R. J., *J. Am. Chem. Soc.* **99**, 6027–6038 (1977)

[64] Kamlet, M., Abboud, J. L., Abraham, M., and Taft, R. J., *J. Org. Chem.* **48**, 2877–2887 (1983)

[65] Kamlet, M., Doherty, R., Abraham, M., and Taft, R., *J. Phys. Chem.* **92**, 5244–5255 (1988)

[66] Honig,B., Sharp, K., and Yang, An-Suei, *J. Phys. Chem.* **97**, 1101–1109 (1993)

[67] Abraham, M. H., *J. Am. Chem. Soc.* **104**, 2085–2094 (1982)

[68] Ben-Naim, A., and Marcus, Y., *J. Chem. Phys.* **81**, 2016–2027 (1984)

[69] Dill, K. A., *Biochemistry* **29**, 7133–7155 (1990)

[70] Sneddon, S. F., and Tobias, D. J., *Biochemistry* **31**, 2842–2846 (1992)

[71] Burgot, G., Serrand, P., and Burgot, J.-L., *Int. J. Pharm* **63**, 73–76 (1990)

[72] Weiland, G. A., Minneman, K. P., and Molinoff, P. B., *Nature* **281**, 114–117 (1979)

[73] Bree, F., El Tayar, N., van de Waterbeemd, H., Testa, B., and Tillement, J.-P., *J. Receptor Res.* **6**, 381–409 (1986)

[74] Raffa, R. B., and Porreca, F., *Life Sciences* **44**, 245–258 (1989)

[75] Da, Y.-Zh., Ito, K., and Fujiwara, H., *J. Med. Chem.* **35**, 3382–3387 (1992)

[76] Beezer, A. E., Hunter, W. H., and Storey, D. E., *J. Pharm. Pharmacol.* **35**, 350–357 (1983)

[77] Kauzmann, W., *Adv. Protein Chem.* **14**, 1–63 (1959)

[78] Hildebrand, J. H., *J. Phys. Chem.* **72**, 1841–1842 (1968)

[79] Némethy, G., Scheraga, H. A., and Kauzmann, W., *J. Phys. Chem.* **72**, 1842 (1968)

[80] Taylor, P. J., Hydrophobic properties of drugs. In *Comprehensive Medicinal Chemistry*, Vol. 4, Hansch, C., Sammes, P. G., and Taylor, J. B. (Eds.). Pergamon Press: Oxford; 241–294 (1990)

[81] Muller, N., *Acc. Chem. Res.* **23**, 23–28 (1990)

[82] Muller, N., *J. Phys. Chem.* **94**, 3856–3859 (1990)

[83] Muller, N., *Trends Biol. Sci* **17**, 459–463 (1992)

[84] Topliss, J. G., Commentary. In *Quantitative Structure-Activity Relationships of Drugs*, Topliss, J. G. (Ed.). Academic Press: London; 497–504 (1983)

[85] Rekker, R. F., *Quant. Struct.-Act. Relat.* **11**, 195–199 (1992)

[86] Van de Waterbeemd, H., *Quant. Struct.-Act. Relat.* **11**, 200–204 (1992)

[87] Hansch, C., Björkroth, J. P., and Leo, A., *J. Pharm. Sci.* **76**, 663–687 (1987)

[88] Tute, M. S., History and objectives of quantitative drug design. In *Comprehensive Medicinal Chemistry*, Vol. 4, Hansch, C., Sammes, P. G., and Taylor, J. B. (Ed.). Pergamon Press: Oxford; 1–31 (1990)

[89] Hansch, C., Maloney, P. P., Fujita, T., and Muir, R. M., *Nature* **194**, 178–180 (1962)

[90] Penniston, J. T., Beckett, L., Bentley, D. L., and Hansch, C., *Mol. Pharmacol.* **5**, 333–341 (1969)

[91] Kubinyi, H., *J. Med. Chem.* **20**, 625–629 (1977)

[92] Dearden, J. C. Molecular structure and drug transport. In *Comprehensive Medicinal Chemistry*, Vol. 4, Hansch, C., Sammes, P. G., and Taylor, J. B. (Eds.). Pergamon Press: Oxford; 375–411 (1990)

[93] Hansch, C., Steward, A. R., Anderson, S. M., and Bentley, D. L., *J. Med. Chem.* **11**, 1–11 (1968)

[94] Hansch, C., and Dunn, W. J., III, *J. Pharm. Sci.* **61**, 1–19 (1972)

[95] Austel, V., and Kutter, E., Absorption, distribution, and metabolism of drug. In *Quantitative Structure-Activity Relationships of Drugs*, Topliss, J. G. ed., Academic Press: London; 437–496 (1983)

[96] Hansch, C., and Klein, T., *Acc. Chem. Res.* **19**(12), 392–400 (1986)

[97] Hansch, C., *Acc. Chem. Res.* **26**(4), 147–153 (1993)

3 Thermodynamics of van der Waals and Hydrophobic Interactions

Rudolf Zahradník and Pavel Hobza

Abbreviations

AO	Atomic orbital
AMBER	Empirical force field
AM1	Semiempirical method of quantum chemistry
BSSE	Basis set superposition error
CC	Coupled-cluster method for calculation of correlation energy
CFF91	Empirical force field
CHARMM	Empirical force field
CI	Configuration interaction method
CVFF	Empirical force field
DISCOVER	Program code
DFT	Density Functional Theory
DZ	Double zeta
E^{COR}	Correlation energy
E^{SCF}	Self Consistent Field energy
GROMOS	Program code
HF	Hartree-Fock quantum chemical method
HFR	Hartree-Fock-Roothaan quantum chemical method
LCAO	Linear combination of atomic orbitals
MC	Monte Carlo method
MD	Molecular dynamics method
MM2	Empirical force field
MM3	Empirical force field
MP	Møller-Plesset method for calculation of correlation energy
P.E.S.	Potential energy surface
PM3	Semiempirical method of quantum chemistry
QZ	Quadruple zeta
SCF	Self Consistent Field method
TZ	Triple zeta
vdW	van der Waals (e.g., forces, interactions)
ZPE	Zero-point energy

Symbols

a	Constant
d_x, d_y, d_z	Cartesian displacement coordinates
k	Boltzmann constant, parameter
m	Mass
r	Distance
ε	Energy
ν	Vibration frequency
A	Helmholtz energy
C	Parameter
C_v, C_p	Heat capacity at constant volume and pressure
D_e	Dissociation energy
E	Energy
F	Force, matrix of potential energy
G	Gibbs free energy, matrix of kinetic energy
H	Enthalpy, Hamiltonian
I	Moment of inertia
K	Thermodynamic equilibrium constant
M	Power constant
N	Number of particles, power constant
P	Pressure
Q	Partition function
R	Distance, gas constant
S	Entropy, overlap matrix
T	Absolute temperature (in K)
U	Internal energy
V	Volume, potential energy

3.1 Introduction

The purpose of this work is two-fold. First, we wish to present our opinion and feelings about that we call van der Waals (vdW) interactions (also called non-covalent or weak) in the last decade of the 20th century (section 3.3). These interactions play a very important role in chemistry, and in the realm of biodisciplines they assume an absolutely overwhelming role [1]. In contrast to covalent bond (decisive for chemistry), vdW interactions are essentially due to interactions between permanent, induced and time-dependent electric multipoles. It is fair to add that electric charges and dipoles cause the strongest interactions [1].

Second, we wish to offer to our colleagues approaching this area from the biological side a clear and simple review of the tools which are currently used in the area under study (sections 3.1 and 3.2) [2, 3].

Thermodynamics deals with work, heat, energy, entropy, Gibbs energy, and the equilibria of physical and chemical processes. This type of thermodynamics is called equilibrium thermodynamics. In fact, in biological systems processes which are far
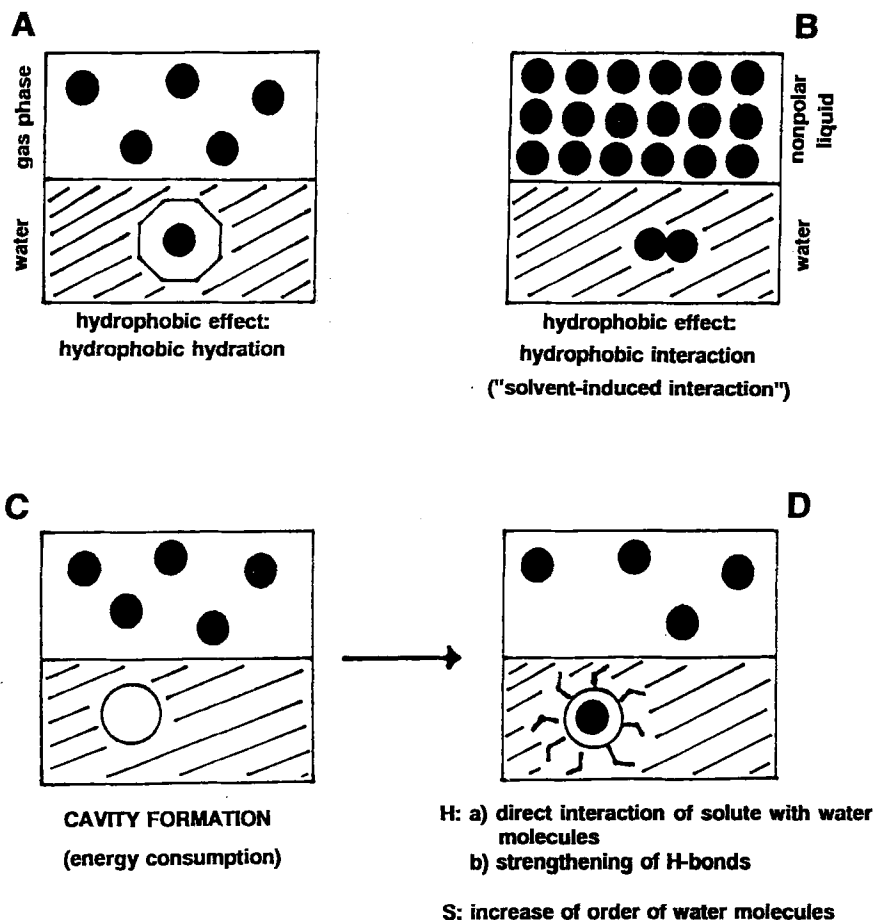
Figure 1. Two aspects of the hydrophobic effect: phenomenological description. A, **Hydrophobic hydration** of a nonpolar molecule (black circle) originating from a gas phase (nonpolar liquid phase also possible). The clathrate structure of the water molecules surrounding the nonpolar molecule is indicated by a regular octagon; B, Two nonpolar molecules, originating in this instance from a nonpolar liquid phase, prefer the associated state in water: **hydrophobic interaction** (or "solvent-induced interaction"). The hydrophobic hydration is described theoretically as a two-step process; C, **Cavity formation** connected with energy loss; D, The introduction of a nonpolar molecule into the cavity is connected with an enthalpy (H) decrease (interaction enthalpy) as well as an entropy (S) decrease (increase of order of water molecules surrounding the nonpolar molecule).

from equilibrium play an important role. In this connection, non-equilibrium (or irreversible) thermodynamics plays a decisive role. This exceedingly important discipline has not yet attained broader applicability but will definitely play a fundamental role in the future.

There is no uniform definition of *intermolecular complexes* [4]. In this respect, the situation in the literature is, indeed, chaotic. We prefer a rather broad and general de-

finition of these complexes which we call van der Waals systems (alternatively called vdW molecules, vdW clusters or vdW associates). The individual parts (subsystems) of these species are held together by forces other than covalent (chemical) forces. Among typical representatives of these species are, e.g., ionic complexes in which the ion-dipole or quadrupole force dominates, charge-transfer complexes, species with hydrogen bonds and London vdW molecules. In the last species the subsystems are linked together by London dispersion forces.

Hydrophilicity and *hydrophobicity* [5, 6] are "conjugated" terms which are complementary with lipophobicity and lipophilicity. Generalization of these terms is straightforward and leads to lyophilicity (i.e., solvent attraction) and lyophobicity (i.e., solvent repulsion). These terms are used for individual species, which are either lyophilic or lyophobic, or include groups possessing lyophilic and groups possessing lyophobic features. Pragmatically, these terms are most frequently used in connection with systems consisting of two immiscible liquids. The most topical systems in chemistry and especially in biology have water as one component. It is fair to admit that an essential feature of all these interactions was already known to alchemists; They said "*similia similibus dissolvuntur*".

More quantitatively, all these processes assume two common features: They depend on the solvent, water, and the ultimate equilibrium is greatly affected by the entropy change accompanying the process.

The tendency for dissolved species to aggregate spontaneously in aqueous solutions is apparently the main organizational process in the realm of biodisciplines. This is called the hydrophobic effect. It was described by Kauzmann [7] in 1959 and then intensively studied by Scheraga et al. [8] and other authors [9]. The hydrophobic effect has two aspects: hydrophobic hydration and hydrophobic interaction. The former is the hydration of a nonpolar system and the latter is the interaction between nonpolar molecules dissolved in water (Fig. 1A and 1B). For purposes of theoretical description, hydrophobic hydration is described as a two-step process (Fig. 1C and 1D).

Unique features of water are responsible for all these processes. Only one specific characteristic is presented in Table 1, namely the energy accompanying water dimer formation; for the sake of comparison related energy values are also given for neighboring hydrides.

Table 1. Important feature of water: strong H-bonding. Values indicate *ab initio* stabilization energies (MP2/6-31G* level) in kcal mol^{-1}.

Strong	H-bonding	Weak	H-bonding
$(NH_3)_2$	4.4	$(PH_3)_2$	0.4
$(H_2O)_2$	7.3	$(H_2S)_2$	1.6
$(HF)_2$	8.0	$(HCl)_2$	1.9

Concepts like electronegativity, aromaticity, or hydrophilicity are mostly introduced because we like (and sometimes really need) pictorial explanations. They permit us not only to "understand" but also to organize accumulated observations. It is also true, however, that sometimes these essentially useful tools run us into difficulties. In any case, it is desirable to realize that description of the phenomena under consideration

can be carried out in terms of physics; we do not really need those concepts. The main tools of theoretical physics currently used for computational investigation of atoms, molecules, biomolecules, polymers, biopolymers, and solids are presented in Table 2. All of them represent axiomatic disciplines.

Table 2. Tools of theoretical physics currently used for theoretical studies of atoms, molecules, biomolecules, polymers, biopolymers, and solids [2].

Discipline	Axiomatic basis
Classical reversible thermodynamics	Three laws of thermodynamics
Classical mechanics	Newton's principles
Statistical mechanics	Four axioms of statistical mechanics
Time-independent molecular quantum mechanics	Four axioms of quantum mechanics

3.2 Outline of Thermodynamics and Auxiliary Disciplines

This chapter contains several items of textbook knowledge. The reason for this is that some potential users of theoretical tools are, to a smaller or greater extent, familiar with the individual steps of the overall treatment but essential features of the overall procedure escape them (Fig. 2).
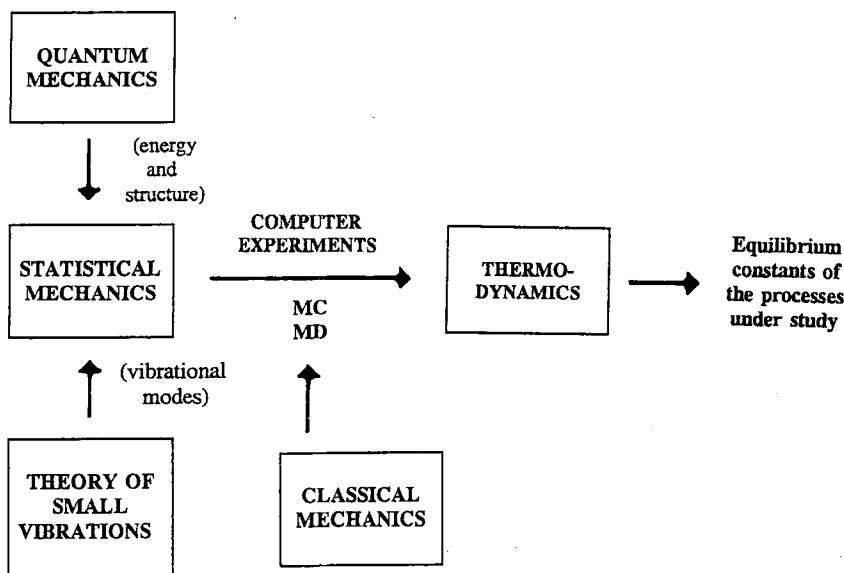
Figure 2. Theories and procedures used in computational treatment of equilibria (MC and MD are Monte Carlo-type computer experiments and molecular dynamics, respectively).

Some fundamental relationships of reversible thermodynamics are summarized in Appendix 1. Boltzmann introduced statistical mechanics a century ago. It is the partition function (a magnitude related to the Boltzmann distribution function) which assumes a central position in statistical mechanics and thermodynamics: its important feature is that it can be easily evaluated in terms of universal constants and molecular constants and characteristics (see Appendix 2). These magnitudes can either be obtained from the analysis of experimental molecular spectra (rotational, vibrational, and electronic) or can be generated by means of molecular quantum mechanics (solution of the Schrödinger equation) and by solving the vibrational problem [10] (Wilson's FG analysis). A fascinating feature of statistical thermodynamics is that it connects the structure of a substance with its thermodynamic behavior. Evaluation of the partition functions of molecules is a simple and straightforward task for processes taking place in an ideal gas phase. It is true that the fundamental equations of statistical mechanics have been available for many years also for condensed media, but their evaluation is hopelessly complicated. In spite of great differences between ideal gas phase and real condensed phase, the former represents mostly a valuable starting point.

The introduction of computer experiments about 30 years ago, represented a great step forward. That which is called the Monte Carlo (MC) procedure [11] is well suited for describing the structures of liquids and solutions and also equilibria in solutions. However, another procedure, based on classical mechanics, is capable of doing the same job but it also describes the dynamic features of processes in solutions [12], which is a very valuable feature. Two types of procedure are elaborated, which permit us to calculate a change in the Helmholtz energy (definition: $A = U - TS$; cf. Abbreviations) of the system under study when passing from state i to state j ($\Delta A_{i \to j}$). The first procedure is named the thermodynamic perturbation theory [13] and the second, the thermodynamic integration method [14] (Appendix 3). Calculations of ΔA are carried out for an NVT ensemble (i.e., constant number of particles, N, constant volume, V and temperature, T). Passing from the NVT ensemble to the NPT ensemble (P is constant pressure) is connected with passing from the Helmholtz to the Gibbs energy (definition: $G = H - TS$; cf. Abbreviations). A specific illustration of application of the relative free energy calculation method [15] to the evaluation of the difference in the solvation free energies of methanol and ethane [16] is given in section 3.5.

Molecular quantum mechanics represents a tool [3] that makes the molecular constants and total energy readily accessible (Appendix 4). All the necessary molecular constants can be obtained from the total wave function. In a majority of instances, we are interested in the lowest energy value (i.e., energy value of the ground state) resulting from solution of the determinant equation.

Although solving the Schrödinger equation now represents a standard and routine procedure, a good-quality level for bigger polyatomic molecules is still a rather demanding task. Therefore, it is expedient even now to have several levels of complexity at our disposal [3, 17]. Firstly, there are nonempirical and semiempirical quantum chemistry methods. Pragmatically speaking, the variation method is mostly a versatile tool; its combination with the Hartree-Fock-Roothaan (HFR) procedure is the most widespread technique. It is fair to admit that it is only rarely a reliable tool at a nonempirical level. The beyond-HFR procedure is always necessary when a covalent or van der Waals bond is formed or split. When investigating polyatomic species, complete

geometry optimization is a standard procedure. When stationary points are located on the potential energy surface of the system under study, a decision must be made as to whether they represent minima (stable isomers) or saddle points (Eyring's activated complexes). This is an easy task because we need to carry out vibrational analysis anyway (Appendix 5) and this analysis permits us to make the decision.

The remaining two procedures are based on classical mechanics. This is amazing because our treatment deals with particles of the microcosmos, which is, of course, a realm for quantum and not for classical mechanis. It turns out, however, that this requirement is strictly valid only for electrons. Much heavier nuclei can frequently be properly described in terms of classical mechanics. This has significant practical consequences: vibrational spectra (Appendix 5) and the dynamics of molecules can be successfully described in terms of Newton's mechanics (or in terms of its version originating from the 19th century, Lagrange and Hamilton mechanics) (Appendix 6).

In order to be able to analyze the vibrational motion of a polyatomic molecule, it is necessary to perform the Wilson's matrix analysis. The formulation of characteristic equations (Appendix 5) requires the evaluation of the elements of the potential (b_{ij}) and kinetic energy (a_{ij}) matrices. From a formal point of view the problems solved in Appendices 4 and 5 are identical. The characteristic values represent the energies of the individual normal modes of vibration (λ) and the associated characteristic vectors offer information on the nature of the individual vibration modes (the so-called harmonic approximation). The vibrational energies permit as to evaluate the zero-point energies of the molecules under study, to obtain their vibrational partition functions, and to decide on the nature of the localized stationary points on the potential energy surface of the molecule under study. (As mentioned before, the partition function is related to the Boltzmann distribution function. The total partition function for a system can be approximated as a product of its partial partition functions, which are related to rotational, vibrational, electronic, and translational energy).

As already mentioned, classical mechanics represents a valuable tool for investigating molecular dynamics (Appendix 6). Clearly, when using this tool, one does not obtain information on quantum effects like quantum mechanical tunneling. More specific information will be presented in section 3.5.

3.3 Intermolecular Interactions of the van der Waals Type

3.3.1 The Physical Nature of van der Waals Interactions

Let us deal with two mutually interacting species and let us call them subsystems. A chemical bond is formed between the subsystems when their electron clouds overlap. In general, however, overlap is not necessary in molecular interactions (molecules interact even at very large distance where overlap is zero) and the reason for mutual attraction must be sought in the electrical properties of the subsystems.

The nonuniform distribution of charge throughout the molecule gives rise to an electric multipole; this may be a dipole, quadrupole, octapole, etc. Interaction of these multipoles leads to the most prominent energy contribution, the electrostatic (cou-

lombic) contribution. An induction contribution appears between molecules with permanent multipoles and nonpolar molecules. In this case the attraction originates from the electrostatic interaction between permanent and induced multipoles. And, finally, also the dispersion interaction taking place between nonpolar molecules originates from the electrostatic interaction between a time-variable multipole and an induced multipole.

In addition to the attractive contributions considered, there must be another force preventing subsystems from approaching one another too closely. This force, termed exchange-repulsion, originates similarly as chemical covalent forces, from the overlap of electron clouds.

3.3.2 Classification of van der Waals Clusters

Van der Waals clusters are those in which the individual subsystems are held together by forces other than chemical covalent forces. This category of systems includes various types of clusters named after the dominant attractive contributions: ionic, electrostatic, London, charge-transfer and hydrogen-bonding. Examples of these systems are given in Table 3.

Table 3. Examples of various types of van der Waals clusters.

Type	Examples
Ionic	$Li^+ \cdots N_2$, $F^- \cdots H_2O$, $O_2 \cdots O_2^-$, phosphate anion of DNA$\cdots Mg^{2+}$
Electrostatic	$(LiF)_2$, interaction between amino acids
London	$(Ar)_2$, $(benzene)_2$, interaction between aliphatic hydrocarbons
Charge-transfer	benzene$\cdots$TCNE,* protonated amine$\cdots$phenyl group of phenylalanin
Hydrogen-bonded	$H_2O \cdots HOH$, guanine$\cdots$cytosine

* Tetracyanoethylene.

3.3.3 Calculation of the Interaction Energy

Two theoretical methods can be used for calculating the interaction energy (ΔE). The supermolecular variation method which determines ΔE as the difference between the energy of the cluster (supermolecule) and the energies of the isolated systems. The perturbational method gives ΔE directly as the sum of physically distinct contributions such as electrostatic, induction, dispersion, and exchange-repulsion. Both the methods have advantages and drawbacks. The supermolecular approach is theoretically able to provide an interaction energy at any level of accuracy, providing that a sufficiently high percentage of the correlation energy is covered. The advantage of this approach is that it is simple and straightforward and any quantum chemical code can be used. The advantage of the perturbational method is that ΔE is obtained directly and not as the energy difference. It is, however, supermolecular approach which nowadays is used almost exclusively. The perturbation approach, in the form of symmetry-adapted per-

turbation theory [18], is used for highly accurate calculations providing benchmarks for supermolecular calculations.

In the following parts various theoretical supermolecular procedures will be used for evaluation of interaction energy and other properties of molecular clusters.

3.3.3.1 Nonempirical *ab initio* Variational Method

The interaction energy of the complex is evaluated as a sum of the SCF (self consistent field) interaction energy (ΔE^{SCF}) and the correlation interaction energy (ΔE^{COR})

$$\Delta E = \Delta E^{SCF} + \Delta E^{COR} \tag{1}$$

The former contribution covers electrostatic, induction and exchange-repulsion terms. The dispersion energy, which plays important role especially in case of interaction of very large (biological) systems, is not included in the ΔE^{SCF} term. To include this contribution, the supermolecular treatment must include energy terms originating in the correlation of electron movements. ΔE^{COR} in Eq. (1) covers not only the dispersion energy but also other less important contributions.

Correlation interaction energy is always important and cannot be neglected; evaluation of this term is much more tedious and time-demanding than the evaluation of ΔE^{SCF}. Among suitable methods, the Møller-Plesset (MP) perturbational theory is now in common use. The second-order MP theory (MP2), which is easily applicable even to large clusters (having up to 100 atoms), gives surprisingly good estimates of ΔE^{COR}. MP2 covers contributions from the double-electron excitations in the second perturbative order. The good performance of MP2 is, however, due to the partial compensation of higher-order contributions.

More accurate values of ΔE^{COR} result if the MP theory is performed through the 4th order. The double electron excitations are described in the 2nd, 3rd, and 4th order. In the 4th order also single, triple, and quadruple electron excitations are covered; among them, triple excitations play a dominant role. Unfortunately, their evaluation is extremely time consuming, much more than that of the other contributions at the 4th perturbative order. The next step in the accuracy of ΔE^{COR} results from the use of the coupled-cluster (CC) method.

Basis set

Choice of the basis set is very important and the quality of the basis set needed depends on the nature of the cluster, specifically on the role of the correlation interaction energy. If the stabilization energy is properly described by the Hartree-Fock interaction energy (true for ionic, and H-bonded clusters), then relatively small basis sets give accurate values of the stabilization energy. On the other hand, the basis sets for London clusters, where all the stabilization comes from the correlation interaction energy, must be considerably larger.

Calculating the correlation energy even at the lowest, perturbational level (MP2), brings a limitation to the size of the cluster, and clusters having more than about 200

atoms are prohibitively large. The SCF interaction energy itself can, however, be evaluated for much larger clusters.

3.3.3.2 Density Functional Theory

In density functional theory (DFT) the exact Hartree-Fock exchange is replaced by a more general expression – the exchange-correlation functional [19]. The DFT energy thus includes terms accounting for both the exchange and correlation energies. Let us recall that the HF theory covers only the exchange energy and the additional evaluation of the correlation energy is tedious and time consuming.

Encouraging results were obtained with DFT for isolated molecules by incorporating nonlocal, density-gradient terms in the exchange and correlation functionals [20]. The use of DFT for molecular clusters is, however, limited for the following reasons [21]: i) the dispersion energy is not included and consequently no minimum is found for London clusters; ii) the stabilization energy of charge-transfer clusters is strongly (about ten times) overestimated; iii) in the case of H-bonded clusters, the DFT may incorrectly predict the structure of the global minimum. It can thus be concluded that the use of the DFT method in the realm of vdW and biological clusters in rather limited and cannot be recommended.

3.3.3.3 Semiempirical Methods

The use of semiempirical methods of quantum chemistry for vdW clusters cannot be recommended. None of the methods including those recently developed (AM1 and PM3) led to reliable results for various types of vdW clusters. The most complicated task is imposed by London clusters; semiempirical methods are not able properly to evaluate the dispersion energy. On the other hand, semiempirical methods can be used under some conditions for H-bonded clusters. It is, however, necessary first to confirm nonempirically the validity of semiempirical results for several H-bonded clusters of the type considered (e.g., H-bonded DNA base-pairs). Without nonempirical verification, the semiempirical methods cannot be used, even for the H-bonded clusters.

3.3.3.4 Empirical Procedures

Molecular mechanics methods, also called semiempirical force field or empirical potentials differ basically from the nonempirical or semiempirical methods of quantum chemistry because they are not based on solving Schrödinger equation. These methods treat molecules as systems composed of atoms held together by bonds, and deal with the contributions to a molecule's electronic energy from bond stretching, bond bending, van der Waals attraction and repulsion between nonbonded atoms, electrostatic interaction due to polar bonds, and energy changes accompanying internal rotation about single bonds. Empirical procedures were developed and parametrized for specific, rather narrow classes of systems. The methods are mostly successful within these classes of system and theoretical results agree with the experimental features of systems in the ground electronic state.

Widely used potentials include AMBER [22] and CHARMM [23], both parametrized for proteins and nucleic acids, and MM2 and MM3 [24], designed for hydrocarbons. The BIOSYM force fields calledCVFF and CFF91 [25], both parametrized for peptides and proteins, represent the most recent development in the field.

Parametrization of these force fields was based on various experimentaldata like geometry, conformation, various heats (formation, vaporization, sublimation, solvation, adsorption), and coefficients (virial, viscosity, transport). The basic idea of use and application of empirical procedures is based on believe that parameters of force fields are transferable, i.e., transferable from the systems used for parametrization to the system under study (not included in the parametrization). Evidently, the narrower is a set (in the sence of structural types) of systems used for parametrization, the larger chance to obtain reliable results for system studied (belonging to this "narrow" set of systems). An alternative to empirical potentials parametrized on the basis of experimental quantities are potentials derived from theoretical quantum chemical calculations. The main advantage of the latter procedure is that it can be applied to any type of interacting systems.

3.3.4 How to Obtain a Consistent Set of Various Calculated Properties for van der Waals Clusters

The main advantage of theoretical study of a molecular cluster is the fact that a nearly complete set of consistent cluster properties can be generated. The quality of the evaluated properties depends on the theoretical level used; some properties require higher-level treatment. The following steps must be performed:

3.3.4.1 Potential Energy Surface (P.E.S.)

The aim is not only to localize all the minima and saddle points of the P.E.S. but also to correctly describe regions far from the minima. Clearly, several dozen points are required even for a small molecular cluster. Detailed investigation of the P.E.S. is very tedious but without its knowledge one cannot evaluate a (nearly) complete consistent set of various properties of the cluster under study. One-electron properties are evaluated for global and local minima. The classical search for all the minima on the P.E.S. (i.e., based on chemical intuition) is limited to small clusters having not more than 10–15 atoms. For larger clusters, some objective method must be used; the quenching method [26] can be recommended.

3.3.4.2 Stabilization Energy

The stabilization energy for the global minimum or for global and nearest local minima should be investigated more carefully. The stabilization energy is more sensitive to the quality of the theoretical description than the other properties of a cluster.

3.3.4.3 Empirical Potential

The most important property of a cluster is the respective intermolecular potential. The importance of theoretical determination of the intermolecular potential is increased by the fact that its evaluation using various experimental characteristics is tedious and not sufficiently accurate. The main advantage of the theoretical procedure is the fact that it can be applied to any type of molecular cluster. The analytical form of the potential should be sufficiently flexible and should contain a maneagable number of adjustable parameters.

3.3.4.4 Vibration Frequencies

Let us begin with a statement saying that a harmonic oscillator is described by a parabolic potential energy, while an anharmonic oscillator is characterized by a nonparabolic dependence. The use of the harmonic approximation for molecular clusters – and especially for floppy clusters – is limited; therefore, for biological systems it represent a poor approximation. The vibrational energy levels of the cluster should be obtained by solving the vibrational Schrödinger equation, which requires knowledge of the intermolecular potential. In this way, anharmonic frequencies are generated. The calculation of intermolecular vibrational frequencies is topical for two reasons: a) vibrational frequencies (in contrast, e.g., to the structure or stabilization energy) are observable; b) vibrational frequencies directly probe the quality of the intermolecular potential.

3.3.4.5 Computer Experiments

The structure of higher clusters at temperature T can be determined by molecular dynamics (MD); here again, the knowledge of the respective intermolecular potential is essential. Furthermore, when performing MD calculations the thermodynamic characteristics of cluster formation can also be obtained. This step is extremely important; it must be kept in mind that the global minimum localized on the P.E.S. at 0 K can be less favorable at higher temperatures. This is due to the different role which entropy plays for various structures of the cluster under study.

3.4 Processes Involving Hydrophobic Effects

The processes named in the title of this section represent a rather narrow but extremely important group of processes taking place in condensed media. In the realm of chemistry it is frequently possible to deal with a process under study either in the gas phase or in a solution. Only rarely does the chosen solvent mimic the conditions experienced by the reactants in a more or less ideal gas phase. In the majority of instances, the influence of the solvent is significant and causes changes in the equilibria and rate constants and modifies or changes the reaction mechanism. In some cases, the solvent-induced rate constant change amounts to several orders of magnitude. The subject we are dealing with is of a different nature, being closely connected with heterogeneous equilibria: two-phase liquid systems – immiscible liquids and the behavior of low-solubility

species in water. The former systems frequently include small molecules, while the latter are mostly represented by large molecules or by macromolecules playing a role in biological systems. All these phenomena belong to a group of hydrophobic effects; their basic features are mentioned in section 3.1. For convenience, references on selected recent theoretical and experimental works dealing with various features of hydrophobic interactions are mentioned: interpretation of hydrophobic effects and calculations of these interaction [27–30], relation to adsorption [31], low solubility of hydrocarbon [32, 33], methane solvation and association [34, 35], entropy role [36, 37], protein structure and stabilization [38–42], interactions between biomolecules [43–45], and to host-guest interactions [46–48].

We do not recommend using the term *hydrophobic bond*. We believe that a bond can be considered to be formed when the approach of the subsystems is connected with a significant energy decrease. This is not the case of the class of processes under discussion. They are connected with a positive, zero, or small negative enthalpy change. The entropy change for the overall interaction including the solvent molecules is large and positive, while the total Gibbs energy change is negative. Clearly, the isolated association of partners is connected with an entropy decrease. Obviously, the above-mentioned entropy increase is due to structural reorganization of the solvent, usually water. Very numerous processes playing a fundamental role in biological systems belong to this class: protein folding, formation of micelles, enzyme-substrate interaction, antigen-antibody interaction and a vast number of related processes.

Hydrophobic effects play an essential role in both physical (e.g., transport) and chemical processes associated with biotransformations. They are specially important for the initial step (approach of partners and their partial desolvation) and the final step (separation of the transformed subsystems).

Discussion of the nature of the hydrophobic effect and its relevance to the above-cited processes is proceeding and probably an ultimate decision will not be reached rapidly. The essential difficulty is that the hydrophobic interaction (i.e., solvent-induced interaction between two or a set of nonpolar molecules, small or large) is not at present amenable to experiment because of the extremely low solubility of these nonpolar solutes in water. The situation with hydrophobic hydration seems to be more favorable but, also not clear enough at the present.

The results of the hydrophobic hydration of small molecules in water, as treated by the efficient integral equation theory of Pratt and Chanlder [49] are in fair agreement with experiments [50]. The predictive efficiency of the Pratt-Chandler theory for solvent-induced solute-solute interaction is good.

In the realm of hydrophobic interactions, MD computer experiments provide great assistance in two respects. First, they offer valuable information on specific, new systems, and, second, they provide useful data for testing theoretical models. The first computer simulations of the hydrophobic interaction were not reliable enough. An MC scheme introduced by Pangali et al. [51] led to the potential of mean force for two Lennard-Jones spheres in water. It evolved that they passed between two minima, one with closely associated spheres and one with solvent-separated spheres; the latter was more populated. MD experiments supported this finding.

The importance of solvent-separated minima was demonstrated by several authors with systems such as a pair of methane molecules, or a pair of rare-gas atoms in water.

Papers were published where solute-dimer dissociation is more pronounced that dimer formation. However, Wallquist [52] demonstrated that, in a system of 18 methane-like species plus 107 H_2O molecules, hydrophobic association takes place. Faced with the contradictory finding on the association and dissociation of nonpolar solutes, Wallquist concluded that it is the many-body part of the potential of mean force which is responsible for the association tendency between nonpolar solutes. It was Smith et al. [53] and Skipper [54] who showed that there is an attractive hydrophobic interaction between two methane molecules in water, which is entropy-driven. From the point of view of this contradictory feature of all the mentioned investigations, a study on association and dissociation of nonpolar and polar van der Waals pair in water should be mentioned [55]. A series of MD simulation (298 K, 1 atm) was performed for pairs of van der Waals spheres with radii of 200, 250, and 300 pm dissolved in 214 water molecules. These spheres were nonpolar or polar. In the latter case, they bear partial charges of the same size and opposite sign ($+0.1$ and -0.1; -0.3 and -0.3). For molecules of various sizes a smooth shift from associative to dissociative interaction was found with charging of the van der Waals pair. "Snapshots" from the course of the MD run indicate the presence of both tight and solvent-separated nonpolar van der Waals pairs. Finally, evidence was obtained that the hydrophobic interaction is an entropy-driven process [56].

3.5 Specific Illustrations

In the following section we would like to demonstrate the ability of theoretical treatment to generate an almost complete set of various physical properties of a molecular cluster. The first system studied in this way in our laboratory was the benzene$\cdots$Ar$_n$ cluster. Seemingly such a system is abstract and not relevant to biological reality. The opposite is true, however. The forces responsible for the very existence of the benzene$\cdots$Ar$_n$ clusters as well as for their structure and dynamic properties, are the same as those in case of biomacromolecules (proteins, DNA) or clusters thereof. In all these instances (benzene$\cdots$Ar$_n$, DNA, etc.) entropy contribution plays a very significant role. The other reason for selecting the benzene$\cdots$Ar$_n$ cluster was the fact that for it exist accurate experimental characteristics (structure, geometry, stabilization energy, vibrational frequencies, population of higher clusters); such characteristics, allowing us to test the quality of theoretical procedure, do not exist for larger clusters. In the present time we investigate in our laboratory clusters of benzene molecules and extensive clusters of DNA bases.

3.5.1 *Ab initio* Evaluation of a Consistent Set of Various Properties of the Benzene$\cdots$Ar$_n$ Cluster

3.5.1.1 Potential Energy Surface

Five different structures of the benzene$\cdots$Ar complex (cf. Fig. 3) were studied at the *ab initio* HF level with inclusion of correlation energy [57]. The respective stabilization energies and optimal structures are summarized in Table 4. Clearly, the C_{6v} structure A
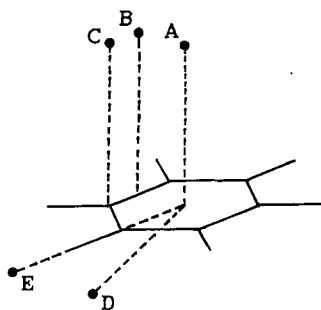
Figure 3. Structures of various benzene···Ar complexes; the individual structural types are labelled by A, B, C, D, E (cf. Table 4).

is the moststable; displacing Ar from the C_6 axis (structures B, C) leads to destabilization. Passing from sandwich structures to a planar structure results in an important stabilization energy decrease. The intermolecular distance found for structure A (3.534 Å) agrees well with the corresponding experimental value (3.584 î).

Table 4. Optimal interaction energies and distances for five different structures of the benzene···Ar cluster.

Structure[a]	A	B	C	D	E
R (Å)	3.5	3.7	3.7	5.2	6.0
$-\Delta E$ (cm^{-1})	351	261	245	147	80

[a] Cf. Figure 3

3.5.1.2 More Accurate Calculations for the Global Minimum

The C_{6v} structure of the benzene Ar···complex was studied athigher theoretical levels [58]. It was found that the MP2 calculation with small basis sets gives a very good estimate of the stabilization energy; this is, however, due to compensation of errors. Only very large basis sets in combination with the CC method give satisfactory results, which converge to the experimental value.

3.5.1.3 Preparation of the Empirical Potential

Two types of empirical potential were fitted [59] to the benzene···Ar *ab initio* P.E.S. The first potential, the global potential (Eq. 2) is generally applicable while the second one, basically the Morse-type potential (Eq. 3) is, due to its complexity, limited to the evaluation of the vibrational spectrum

$$
V = a_1 \left\{ \sum_{i=1}^{6}(C_1/r_{H_i})^N + \sum_{i=1}^{6}(C_2/r_{C_i})^N \right\} +
$$
$$
+ a_2 C_3 \left\{ -a_3 \sum_{i=1}^{6} 1/r_{H_i}^M (1 - C_4/r_{H_i}) - a_4 \sum_{i=1}^{6}(1/r_{c_i})^M \right\} \tag{2}
$$

The individual terms have the following meanings: r_{H_i} and r_{C_i} are the H···Ar and C···Ar distances, a_1, a_2, a_3 and a_4 are constants and C_1, C_2, C_3, C_4, N and M were fitted to the *ab initio* P.E.S. (for details see [59]).

$$V = k_{xx}(d_x^2 + d_y^2) + k_{xxz}w(d_x^2 + d_y^2) + k_{zz}w^2 - D_e \tag{3}$$

Here D_e is the dissociation energy of the complex, and

$$w = 1 - \exp(- ad_z) \tag{4}$$

k_{xx}, k_{zz} and k_{xxz} are parameters.

The Morse-type potential was fitted only to the sandwich structures of the cluster; for details, see [59].

3.5.1.4 Vibrational Frequencies

Intermolecular frequencies were obtained [59] numerically by solving the vibrational Schrödinger equation utilizing empirical potentials obtained in the preceeding step. Both available experimental studies yield vibrational bands at about 40 and 31 cm^{-1} and assign them to intermolecular stretching and the first overtone of the intermolecular bending vibration. The third band at about 64 cm^{-1} was assigned either to the third overtone of the intermolecular bending or the first overtone of the intermolecular stretching vibration. The theoretical Morse stretching (39 cm^{-1}) agrees nicely with the experimental findings. The other theoretical Morse frequencies, bending (29 cm^{-1}), first overtone of bending (57 cm^{-1}) and combination modes (62 and 63 cm^{-1}) differ from the experimental assignment. On the basis of our [59] and van der Avoird's [60] theoretical studies, a new assignment of the experimental peaks was made, which fully agrees with that suggested theoretically, i.e., the bands at 31 and 63 cm^{-1} correspond to the fundamental of bending and a combination mode.

3.5.1.5 Molecular Dynamics Simulations

The number of structures of higher benzene$\cdots$Ar$_n$ clusters increases rapidly with increasing n and equals about 300 for $n = 7$; localization of these minima is very tedious. It must be mentioned here that the use of "chemical intuition" for these purposes is rather limited and some more objective method, like the quenching method, should be utilized. Experimental techniques have become very sophisticated over the past few years and allow us to detect various structures of the cluster. It is, however, not easy to specify the particular structure and the theory should assist in this direction. Because experiments are carried out at non-zero temperature it is essential to include the entropy term. The only feasible way of doing this, i.e., to determine the relative ΔG term for various structures of the cluster, is to perform molecular dynamics (MD) simulations.

As an example, the MD study of the isomers of the benzene$\cdots$Ar$_2$ cluster [61] will be described. Two isomers of the cluster exist. The global minimum corresponds to the (1/1) isomer (having the argon atoms on the opposite sides of benzene; D$_{6h}$), while the (2/0) isomer (both argons are localized on one side of benzene) has higher energy. For benzene$\cdots$Ar, the "global" potential described above was utilized, while the Ar$\cdots$Ar interaction was described by the empirical 6–12 Lennard-Jones potential. To find the relative abundances of the two isomers, very long MD runs (100–400 ns) must be performed. At low temperatures (below 27 K) the population of the (1/1) isomer is

100 %. At higher temperatures, the relative population of the (2/0) isomer becomes higher than that of the (1/1) isomer (64 % : 36 %) and in the temperature interval studied (27–37 K) it is almost temperature-independent. Temperatures above 37 K could not be reached because the clusters dissociate. The preference of the energetically less favorable (2/0) structure is clearly a consequence of the entropy term. A similar study was performed in our laboratory for higher benzene$\cdots$Ar$_n$ clusters ($n = 3$–7).

3.5.2 Monte Carlo Free Energy Perturbation Calculation: Solvation Free Energy of Methanol and Ethane

The difference in free energy of two similar systems can be evaluated using the free energy calculation method [15]. One of the first application concerns the evaluation of difference of solvation free energies of methanol and ethane [16]. The free energy difference for mutation of methanol into ethane in water solution was calculated to be 6.8 ± 0.2 kcal mol^{-1}, while the respective experimental value is 6.9 kcal mol^{-1}. Agreement is evidently excellent.

It is worthwhile to mention that the Monte Carlo free energy perturbation method was recently incorporated into commonly used codes (AMBER, DISCOVER, GROMOS, CHARMM) and it is possible to use it for calculation of relative free energy of various processes, e.g., solvation, conformational change, or molecular association.

References

[1] Hobza, P., and Zahradník, R., *Intermolecular Complexes*. Elsevier: Amsterdam 1988

[2] Atkins, P. W., *Physical Chemistry*, 5th edn. Oxford University Press: Oxford 1994

[3] Zahradník, R., and Polák, R., *Elements of Quantum Chemistry*. Plenum Press: New York and London 1980

[4] Hobza, P., and Zahradník, R., *Chem. Rev.* **88**, 871–897 (1988); Hobza, P., and Zahradnáük, R., *Int. J. Quantum Chem.* **42**, 58–590 (1992)

[5] Tanford, C., *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*. Wiley: New York 1980

[6] Ben-Naim, A., *Hydrophobic Interactions*. Plenum: New York 1980

[7] Kauzmann, W., *Adv. Protein Chem.* **14**, 1–63 (1959)

[8] Némethy, G., and Scheraga, H. A., *J. Phys. Chem.* **66**, 1773–1789 (1962)

[9] A thematic issue on hydrophobic interactions, Faraday Symp. Chem. Soc. **17** (1992)

[10] Wilson Jr., E. B., Decius, J. C., and Cross, P. C., *Molecular Vibrations*. McGraw-Hill: New York 1965

[11] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., *J. Chem. Phys.* **21**, 1087–1092 (1953)

[12] Alder, B. J., and Wainwright, T. E., *J. Chem. Phys.* **31**, 459–466 (1960)

[13] Zwanzig, R. W., *J. Chem. Phys.* **22**, 1420–1426 (1954)

[14] Mezei, M., and Beveridge, D. L., *Ann. N. Y. Acad. Sci.* **482**, 1–23 (1986)

[15] Kollman, P., *Chem. Rev.* **93**, 2395–2417 (1993)

[16] Jorgensen, W., and Ravimohan, C., *J. Chem. Phys.* **83**, 3050–3054 (1985)

[17] Čársky, P., and Urban, M., *Ab Initio Calculations. Methods and Applications in Chemistry*. Springer Verlag: Berlin 1980

[18] Jeziorski, B., Moszynski, R., and Szalewicz, K., *Chem. Rev.* **94**, 1887–1930 (1994)

[19] Jones, R. O., and Gunnarson, O., *Rev. Mod. Phys.* **61**, 689–746 (1989)

[20] Andzelm, J., and Wimmer, E., *J. Chem. Phys.* **96,** 1280–1304 (1992); Johnson, B. G., Gill, P. M. W., and Pople, J. A., *J. Chem. Phys.* **98,** 5612–5626(1993)

[21] Hobza, P., Šponer, J., and Reschel, T., *J. Comp. Chem.* **16,** 1315–1325 (1995)

[22] Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A., *J. Comp. Chem.* **7,** 230–252 (1986)

[23] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M., *J. Comp. Chem.* **4,** 187–217 (1983)

[24] Allinger, N. L., *J. Am. Chem. Soc.* **99,** 8127–8134 (1977); Allinger, N. L., Yuh, Y. H., and Lii, J.-H., *J. Am. Chem. Soc.* **111,** 8551–8566 (1989)

[25] Discover, Versions 2.9.5 & 94.0, May 1994, Biosym. Technologies, San Diego, CA, USA

[26] Stillinger, F. H., and Weber, T. A., *Phys. Rev. A* **25,** 978–989 (1982)

[27] Muller, N., *Acc. Chem. Res.* **23,** 23–28 (1990)

[28] Cramer, C. J., and Truhlar, D. G., *Science* **256,** 213–217 (1992)

[29] Bairamov, Sh. K., Volkenstein, M. V., Golovanov, I. B., Sobolev, V. M., and Tsygankova, I. G., *Mol. Biology* **25,** 111–119 (1991)

[30] Hermann, R. B., *J. Comp. Chem.* **14,** 741–750 (1993)

[31] Maity, N., and Payne, G. F., *Langmuir* **7,** 1247–1254 (1991)

[32] Ruelle, P., Buchmann, M., Nam-Tran, H., and Kesselring, U. W., *Int. J. Pharmaceutics* **87,** 47–57 (1992)

[33] Ruelle, P., Buchmann, M., and Kesselring, U. W., *J. Pharm. Sci.* **83,** 396–443 (1994)

[34] van Belle, D., and Wodak, S. J., *J. Am. Chem. Soc.* **115,** 647–652 (1993)

[35] Smith, D. E., and Haymet, A. D. J., *J.Chem. Phys.* **98,** 6445–6454 (1993)

[36] Rashin, A. A., and Bukatin, M. A., *J. Phys. Chem.* **95,** 2942–2944 (1991)

[37] Searle, M. S., and Williams, D. H., *J. Am. Chem. Soc.* **114,** 10690–10697 (1992)

[38] Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W., *Science* **225,** 178–183 (1992)

[39] van den Burg, B., Dijkstra, B. W., Vriend, G., van der Vinne, B., Venema, G., and Eijsink, V. G. H., *Eur. J. Biochem.* **220,** 981–985 (1994)

[40] Pace, C. N., *J. Mol. Biol.* **226,** 29–35 (1992)

[41] Yang, A.-S., Sharp, K., and Honig, B., *FASEB J.* **6,** A344 (1992) (Meeting abstract)

[42] Fu, L., Laynez, J., Saiz, J., and Freire, E., *Biophys. J.* **66,** A179 (1994) (Meeting abstract)

[43] Ding, W.-D., and Ellestad, G. A., *J. Am. Chem. Soc.* **113,** 6617–6620 (1991)

[44] Sigurskjold, B. W., and Bundle, D. R., *J. Biol. Chem.* **267,** 8371–8376 (1992)

[45] Paliwal, S., Geib, S., and Wilcox, C. S., *J. Am. Chem. Soc.* **116,** 4497–4498 (1994)

[46] Bayly, C. I., and Kollman, P. A., *J. Am. Chem. Soc.* **116,** 697–703 (1994)

[47] Yatsimirsky, A. K., and Eliseev, A. V., *J. Chem. Soc. Perkin Trans.* **part 2,** 1769–1772 (1991)

[48] Thomas, B. E. IV, and Kollman, P. A., *J. Am. Chem. Soc.* **116,** 3449–3452 (1994)

[49] Chandler, D., *Faraday Disc. Chem. Soc.* **66,** 184–190 (1978)

[50] Pratt, L. R., and Chandler, D., *J. Chem. Phys.* **67,** 3683–3704 (1977)

[51] Pangali, C., Rao, M., and Berne, B. J., *J. Chem. Phys.* **71,** 2975–2990 (1979)

[52] Wallqvist, A., *J. Phys. Chem.* **95,** 8921–8927 (1991)

[53] Smith, D. E., Zhang, L., and Haymet, A. D., *J. Am. Chem. Soc.* **114,** 5875–5876 (1992)

[54] Skipper, N. T., *Chem. Phys. Lett.* **207,** 424–429 (1993)

[55] Jungwirth, P., and Zahradník, R., *J. Phys. Chem.* **98,** 1328–1332 (1994)

[56] Jungwirth, P., and Zahradník, R., *Chem. Phys. Lett.* **217,** 319–324 (1994)

[57] Hobza, P., Selzle, H. L., and Schlag, E. W., *J. Chem. Phys.* **95,** 391–394 (1991)

[58] Hobza, P., Selzle, H. L., and Schlag, E. W., *Chem. Rev.* **94,** 1767–1785 (1994)

[59] Bludský, O., Špirko, V., Hrouda, V., and Hobza, P., *Chem. Phys. Lett.* **196,** 410–416 (1992)

[60] Van der Avoird, A., *J. Chem. Phys.* **98,** 5327–5336 (1993)

[61] Vacek, J., Konvička, K., and Hobza, P., *Chem. Phys. Lett* **220,** 85–92 (1993)

Appendices

Appendix 1. Classical Reversible Thermodynamics

First Law

U

$H = U + PV$ $\qquad\qquad\qquad$ $\Delta H = \Delta U + P\Delta V$

$$\left[\text{ideal gas: } C_v = \left(\frac{\partial U}{\partial T}\right)_v ; \; C_p = \left(\frac{\partial H}{\partial T}\right)_p ; \; C_p - C_v = R\right]$$

Second Law

$$S = \frac{q_{rev}}{T} \qquad\qquad\qquad\qquad \Delta S = \frac{\Delta q_{rev}}{T}$$

$A = U - TS$ $\qquad\qquad\qquad\qquad$ $\Delta A = \Delta U - T\Delta S$

$G = H - TS$ $\qquad\qquad\qquad\qquad$ $\Delta G = \Delta H - T\Delta S$

Third Law

$$\lim_{T \to 0} S = 0 \text{ (pure, crystalline substance)}$$

$$\Delta G_T^0 = -RT \ln K$$

Classical thermodynamics: A, phenomenological science completely independent of concepts of the structure of the matter. U, internal energy; H, enthalpy; P, pressure; V, volume; C_v and C_p, heat capacity at constant volume and pressure, respectively; S entropy; q_{rev}, heat accompanying a reversible process, A, Helmholtz energy; G, Gibbs energy; ΔG°, standard change of Gibbs energy; R, gas constant; T, absolute temperature; K, thermodynamic equilibrium constant of the process.

Appendix 2. Statistical Mechanics

1. Rigid rotor – harmonic oscillator approximation

The number of molecules n_i having energy ε_i:

$$n_i = \frac{N g_i e^{-\varepsilon_i/kT}}{Q}$$

$$Q = \sum_i g_i e^{-\varepsilon_i/kT} \quad \text{... partition function}$$

(N is Avogadro's number, k is the Boltzmann constant, T is the absolute temperature, g_i is the degeneracy of the ε_i level; for other symbols, see Appendix 1)

$$U^0 - U_0^0 = RT^2 \frac{d \ln Q}{dT}$$

$$\frac{H^{\circ} - U_0^{\circ}}{T} = RT \frac{d \ln Q}{dT} + R$$

$$C_v^{\circ} = \frac{d}{dT}\left[RT^2\left(\frac{d \ln Q}{dT}\right)\right]$$

$$S^{\circ} = RT \frac{d \ln Q}{dT} + R \ln Q - R \ln N + R$$

$$\frac{G^{\circ} - U_0^{\circ}}{T} = - R \ln \frac{Q}{N}$$

2. Evaluation of Partition Function: Ideal gas

$$Q = Q_t\, Q_r\, Q_v\, Q_e$$

$$[\varepsilon = \varepsilon_t + \varepsilon_r + \varepsilon_v + \varepsilon_e]$$

$$Q_t = \frac{(2\pi mkT)^{3/2}}{h^3} \frac{RT}{P}$$

$$Q_r = (\pi I_A I_B I_C)^{1/2} \sigma^{-1} \left(\frac{8\pi^2 kT}{h^2}\right)^{3/2}$$

$$Q_r = \prod_i \frac{1}{1 - e^{-h\, vi/kT}}$$

The definition equation of Q (see Appendix 1) is used directly for the evaluation of the electronic partition function, Q_e.

Q_i is partition function associated with the i-th type of motion, i.e., the translational, rotational, vibrational, and electronic motion. The ε_i's are the corresponding energies. I_N is a component of the total moment of inertia, h is Planck's constant, v_i is the frequency of the *i*-th vibrational mode.

Appendix 3. Statistical Mechanics: Liquids and Solutions

Change in the Helmholtz energy $\Delta A_{i\to j}$, when the system passes from state i to state j:

1. Thermodynamic perturbation theory (PT)

$$\Delta A_{i\to j} = A_j - A_i = -kT\left\{\exp\left(-\frac{U(\lambda_j) - U(\lambda_i)}{kT}\right)\right\}$$

($<...>$ denotes ensemble average)

2. Thermodynamic integration method (TI)

$$\Delta A_{i\to j} = \int_{\lambda_i}^{\lambda_j} \frac{\partial A(\lambda)}{\partial \lambda} d\lambda = \int_{\lambda_i}^{\lambda_j}\left\{\frac{\partial U(\lambda)}{\partial \lambda}\right\}_\lambda d\lambda$$

The coupling parameter λ (0, 1) stands for a true or hypothetical reaction coordinate.

Appendix 4. Molecular Quantum Mechanics (Time-independent nonrelativistic Schrödinger Equation).

$$\hat{H}\psi = E\psi$$
$$\psi = 1/\sqrt{n!} \mid \ldots\varphi_i\overline{\varphi_i} \ldots \mid$$

where

$$\phi_i = \sum_{\mu=1}^{n} c_{i\mu}\chi_\mu$$

Application of the variation principle:

$$\sum_{\mu=1}^{n} c_{i\mu}(H_{\mu\nu} - E_iS_{\mu\nu}) = 0 \longrightarrow c_{i\mu}, \text{ i.e., } \mathring{A}_i$$

$$\downarrow$$

$$\det \mid H_{\mu\nu} - E_iS_{\mu\nu} \mid = 0$$

$$\Downarrow$$

$$E_i$$

$\hat{H}$ is the n-electron Hamiltonian (operator making total energy calculations possible), ψ is the n-electron wavefunction, E is the total energy of the system. The wave function ψ has the form of a Slater determinant or of a linear combination of Slater determinants constructed from molecular orbitals (ϕ_i), which are expressed in the LCAO form (χ_μ is an atomic orbital, AO). $H_{\mu\nu}$ are the matrix elements of the Hamiltonian, $S_{\mu\nu}$ are the matrix elements of the overlap matrix. The linear equations for $c_{i\mu}$ (i.e., for the wavefunctions) have nontrivial solutions only for those values of E_i which satisfy the secular equation $\det \mid H_{\mu\nu} - E_iS_{\mu\nu} \mid = 0$.

Appendix 5. Vibrational problem (theory of small vibrations).

Wilson's matrix analysis:

$$\sum_{i=1}^{3N} A_j(b_{ij} - \lambda_j a_{ij}) = 0 \rightarrow \text{vectors } A_j \text{ describe the motion associated with the individual vibrational modes}$$

$$\downarrow$$

$$\det \mid b_{ij} - \lambda_j a_{ij} \mid = 0$$

$$\Downarrow$$

vibrational energy λ_j

b_{ij} and a_{ij} are the elements of the potential and kinetic energy matrices, λ_j's are the energies of the vibrational modes, which are described by vectors. A_j. The diagonal elements of the potential energy matrix are the force constants, the nondiagonal elements are the interaction constants.

Appendix 6. Classical mechanics.

Newton's 2nd principle:

$$m\ddot{x} = F$$

Integration yields $x = f(t)$

m is the mass of a mass point moving along the x-coordinate, F is the force, $\dot{x} = dx/dt$ is the velocity, $\ddot{x} = d^2x/dt^2 = d\dot{x}/dt$ is the acceleration.

4 Intramolecular Interactions Encoded in Lipophilicity: Their Nature and Significance

Bernard Testa, Pierre-Alain Carrupt, Patrick Gaillard and Ruey-Shiuan Tsai

Abbreviations

CsA	Cyclosporin A
DBE	Dibutyl ether
HPLC	High performance liquid chromatography
MLP	Molecular Lipophilicity Potential
RP-HPLC	Reversed-phase HPLC
2D	Two-dimensional
3D	Three-dimensional

Symbols

k_w	Capacity factor in RH-HPLC (k_{calc}, k_{exp} are its calculated and experimental values, respectively)
P	Partition coefficient
V	Molecular volume
α	H-bond donor acidity
β	H-bond acceptor basicity
Λ	Polarity parameter
π^*	Dipolarity/polarizability

4.1 Introduction: The Concept of Molecular Structure

Molecular structure is conveniently approached by considering several conceptual levels of description [1], as presented in Table 1.

4.1.1 The Elementary and Geometric Levels of Description

The description starts at an *elementary level* (the one-dimensional structure), where molecules are represented by their chemical formula. The structural information contained in the chemical formula is very limited (e.g., no information on two- and three-dimensional structure), and the single structural attribute one can accurately derive from it is the molecular weight.

The description continues with levels of progressively increasing complexity and richness. At the *geometric levels* of two-dimensional (2D) and three-dimensional (3D) description, molecules are still considered as abstract entities, i.e., geometrical objects

Table 1. A multi-level description of molecular structure and properties [1]

Conceptual levels of structural description	Corresponding structural attributes and molecular properties
A) The elementary level	● Molecular weight
B) The geometric levels ● 2D Structure ● 3D Structure	● Atom connectivity; Z/E configuration ● Relative configuration; Absolute configuration
C) The stereoelectronic levels ● Bulk ● Stereodynamic structure ● Stereoelectronic structure	Attributes and properties of isolated molecules: ● Molar volume, surface area ● Flexibility, conformation; Prototropic equilibria ● Ionization, electron distribution, polarizability; Molecular electrostatic field
D) The level of intermolecular interactions	● Medium-influenced properties of level C ● Emergent properties: melting point, boiling point solvation and hydration chromatographic properties partitioning (lipophilicity) colligative properties
E) Interactions with a biological environment	● Biological properties

consisting of atoms represented by their symbols and of bonds represented by lines. This is by far the most common way of representation, and it contains *explicit* information on such structural attributes as the geometry of the molecule and the adjacency or union of atoms, as well as much *implicit* chemical information which, however, becomes explicit only at higher levels of description. The two-dimensional description considers how atoms are connected, i.e., it defines the connectivity of atoms in the molecule (presence and nature of chemical bonds). It also explicates configuration (*cis* or *trans, Z* or *E*) in case of geometrical isomerism at double bonds. The three-dimensional description views molecules as rigid geometrical objects in space and explicates not only the nature and connectivity of atoms (2D-structure), but also the overall configuration of the molecule in the case of diastereomerism or enantiomerism.

4.1.2 The Stereoelectronic Levels of Description

It is at the next, the *stereoelectronic levels* of description that molecules gain flesh and reality. Here, they are no longer viewed as geometrical abstractions, but as objects with a volume and a shape. At this level, the attributes/properties of bulk, surface and volume are modeled with van der Waals radii. When the temporal dimension is added

to the picture, the molecules are no longer treated as purely spatial, but as spatio-temporal entities. In other words, it is the full *stereodynamic structure* of molecules that is described at this new level, to which correspond structural attributes or properties such as flexibility, conformational behavior and prototropic behavior. These depend heavily on the valence electrons, but it is at the next higher level of description that the latter enter explicitly into the picture. The stereoelectronic attributes are expressed as various measurable or computable electronic properties characteristic of the molecules themselves (e.g., electronic distribution), or affecting the space surrounding them (e.g., electrostatic field) [2].

4.1.3 Social Molecules

At the levels discussed above, the molecules are described in isolation, although this is only partly correct since probes are necessary to define such attributes/properties as molecular volumes and molecular electrostatic potentials. This description "in the vacuum" has consequences as far as the corresponding attributes and properties are concerned. Indeed, attributes and properties corresponding to the elementary and geometric levels of description are environment-invariant, whereas electronic attributes/properties are influenced by the molecular environment. It is only at the next level, namely the *level of intermolecular interactions*, that the complex interplay between a molecule and its environment (e.g., solvent, bulk liquid or crystal) is explicitly considered. At this level of highly complex description, two classes of molecular properties are encountered, many of which have biological importance. First, environment-influenced electronic attributes/properties express the fact that electrons, which are responsible for most intermolecular interactions, cannot remain unaffected by these interactions. These mutual influences are an essential factor accounting for many observable properties and they must be recognized for what they are.

Social molecules (i.e., molecules described at the level of intermolecular interactions) also display *emergent* properties, in other words properties not encountered at the previous levels of description. This is the case of such physicochemical properties as melting point, solvation, behavior in chromatographic systems, and the biologically essential property of lipophilicity. Colligative properties (i.e., concentration-dependent properties) also emerge at this level.

Lipophilicity is a popular and – as far as structure-activity relationships are concerned – remarkably successful property. Reasons for this are to be found in the richness and diversity of the structural information expressed in lipophilicity, which makes it dependent on all structural attributes pertaining to the levels A to C in Table 1. Indeed, lipophilicity results from a vast array of intermolecular interactions ranging from hydrophobic and van der Waals forces to ion-dipole interactions and hydrogen bonds, the latter contributing as significantly to lipophilicity as they do to biochemical recognition [3]. This property and its significance constitute the topic of the present review.

Biological properties (i.e., biological responses) do not belong to a description of chemical structure *stricto sensu*. Nevertheless, "interactions with a biological environment" have been added to Table 1 to indicate that a continuum exists between this level of properties and the previous one, namely that of intermolecular interactions, as

documented by the telling example of membrane/water partition coefficients which can be viewed as both a physicochemical and a biological property.

4.2 Intermolecular Forces Encoded in Lipophilicity

This section compares intermolecular forces expressed in lipophilicity with those forces that underlie molecular recognition in all pharmacological and biological processes. As will be shown, the overlap is broad, although not complete, between these two sets of intermolecular forces.

4.2.1 Recognition Forces in Molecular Pharmacology and Biology

The left-hand side of Fig. 1 lists all major recognition forces of significance in molecular pharmacology and biology. Most of them are classified as electrostatic, although some are more electrodynamic than electrostatic, i.e., charge transfer interactions, ion-induced dipole interactions, induction forces, and dispersion forces.

This classification is an extrathermodynamic one since it voluntarily ignores the division between the enthalpic and entropic components of the binding free energy. Such a neglect is tolerable in the present context, for two reasons. First, many if not all of the forces listed in Fig. 1 contain both components, and their interpretation in terms of enthalpy and entropy is a different issue altogether. And second, lipophilicity is also an extrathermodynamic parameter which can be interpreted as such.

Recognition Forces	**Lipophilicity**
Electrostatic interactions	
Ionic bonds	
Aryl-aryl charge transfer interactions	
Ion - dipole (permanent, induced) bonds	
Reinforced H-bonds	
Normal H-bonds	
Van der Waals forces — **Orientation forces** (permanent dipole - permanent dipole); **Induction forces** (permanent dipole - induced dipole)	***Polarity***
Van der Waals forces — **Dispersion forces** (instantaneous dipole - induced dipole)	
Hydrophobic interactions	***Hydrophobicity***

Figure 1. A comparison between recognition forces in molecular pharmacology and biology (left panel) and intermolecular forces encoded in lipophilicity (right panel).

4.2.2 Factorization of Molecular Lipophilicity

As a ratio of two concentrations at saturation, the partition coefficient (log P)is the net result of all intermolecular forces between a solute and the two phases between which it partitions. When a given type of interaction elicited bythe solute, say H-bond donation, is of equal energy in the two solvents, the two interactions will compensate each other and log P will contain no information about them.

One highly informative interpretation of lipophilicity is based on its factorization of log P into the so-called solvatochromic parameters [4]. The major such parameters are:

- π^*, a measure of the solute's dipolarity/polarizability and thus of its capacity to elicit orientation and induction forces;
- α and β, the solute's H-bond donor acidity and H-bond acceptor basicity, respectively.
- In addition to π^*, α and β, analyses of this type require a parameter to assess the solute's capacity to elicit nonpolar interactions (see section 4.2.4). A steric parameter such as the molar or molecular volume (V) is able to account satisfactorily for such interactions.

Thus, the octanol/water and the heptane/water partition coefficient can be expressed as [4]:

$$\log P_{octanol} = 5.83(\pm 0.53)V/100 - 0.74(\pm 0.31)\pi^* - 3.51(\pm 0.38)\beta$$
$$- 0.15(\pm 0.23)\alpha - 0.02(\pm 0.34)$$

$$n = 78, r^2 = 0.922, s = 0.296, F = 248 \tag{1}$$

$$\log P_{heptane} = 6.78(\pm 0.69) V/100 - 1.02(\pm 0.39) \pi^* - 5.35(\pm 0.50) \beta$$
$$- 3.54(\pm 0.30)\alpha - 0.06(\pm 0.43)$$

$$n = 75, r^2 = 0.955, s = 0.360, F = 438 \tag{2}$$

As a result of equations of this type, it is now common to factorize lipophilicity into two sets of terms, namely polar terms negatively related to lipophilicity (section 4.2.3) and nonpolar terms positively related to lipophilicity (section 4.2.4).

A complementary approach towards factorizing lipophilicity has been proposed by us a few years ago [5]. This approach is based a) on the fact that n-alkanes are completely apolar and non-polarizable, and b) on the well-documented linear relationship that exists within n-alkanes between partition coefficients and molecular or molar volumes. The latest version of this relation between log P and molecular volume (V) is [6]:

$$\log P_{octanol} = 0.0309(\pm 0.0014) V + 0.346(\pm 0.199)$$
$$n = 14, r^2 = 0.997, s = 0.145, F = 3619 \tag{3}$$

In a plot of V versus log P, the observations fall either on the straight line (i.e., n-alkanes) or below it (all other solutes). To the best of our knowledge, no solute has yet

been found which would lie above the line defined by *n*-alkanes. In other words, solutes either are completely apolar (i.e., *n*-alkanes), or are polar to a slight or large degree (all other solutes). The information of relevance in the present context is the vertical distance between a solute and the line for *n*-alkanes, a distance that has been taken as a measure of the global polarity of a given solute. This parameter of polarity has been designated as Λ (capital lambda = inverted V); for a given solute in a given solvent system (e.g., octanol/water or alkane/water), it is defined as the difference between its measured lipophilicity and that intrapolated for a hypothetical *n*-alkane of identical volume.

Using Eq. (3) and the Λ parameter, lipophilicity can thus be factorized into a polar and a nonpolar term, as exemplified by Eq. (4) for the octanol/water system:

$$\log P_{octanol} = 0.039\, V - \Lambda + 0.346 \tag{4}$$

4.2.3 Polar Interactions Encoded in Lipophilicity

The parameter Λ, being a global measure of a solute's polarity, should by definition contain the same information as, e.g., Eqs. (1) or (2), except for the volume term. This has been confirmed for octanol/water (Eq. (5)) and alkane/water (Eq. (6)) partition coefficients [5]:

$$\Lambda_{octanol} = -0.636(\pm 0.124)\pi^* - 3.90(\pm 0.20)\,\beta - 0.186(\pm 0.103)$$
$$n = 168,\ r^2 = 0.918,\ s = 0.25 \tag{5}$$

$$\Lambda_{alkane} = 1.37(\pm 0.30)\,\pi^* - 6.19(\pm 0.48)\,\beta - 3.42(\pm 0.35)\,\alpha - 0.626(\pm 0.234)$$
$$n = 104,\ r^2 = 0.944,\ s = 0.46 \tag{6}$$

These equations make explicit the individual polar terms which are negatively related to lipophilicty and account for most polar interactions between a solute and the two solvent phases. In the case of the octanol/water system (Eq. (5)), the main contributor to a solute's polarity is thus its H-bond acceptor basicity (β), and to a lesser extent its H-bond donor acidity (α) and dipolarity/polarizability (π*). By contrast, the polar interactions expressed in alkane/water partition coefficients (Eq. (6)) are the H-bond donor acidity (*α*) and the H-bond acceptor basicity (*β*), and to a lesser extent the dipolarity/polarizability (π*).

One very important point must be made here, namely that the above analyses (i.e., Eqs. (1), (2), (5), and (6)) are valid only for large and well-distributed series of solutes. For smaller and biased series, it is not unfrequent to observe that the Λ parameter will contain a different balance of polar forces, e.g., $\Lambda_{octanol}$ may depend more on α than on β. This has the major consequence that the factorization of Λ into individual components may be misleading and should be verified for each series. If this is not possible, interpretation should remain cautious and avoid any conclusion about the nature of the major polar force(s) involved. This caution is all the more reasonable considering the often neglected overlaps that exist between π*, *α* and *β*.

At this stage, we can offer the preliminary conclusion that lipophilicity encodes such polar interactions as H-bonds, orientation forces and induction forces, the relative

contributions of which may be difficult or impossible to assess in any given series of solutes. In addition, ionic solutes generate intermolecular forces not included in the π^*, α and β parameters, namely ion-dipole interactions and reinforced H-bonds.

4.2.4 Nonpolar Interactions Encoded in Lipophilicity

The steric term used in factorizing lipophilicity, and which Eqs. (1) and (2) tell us correlates positively with it, is best approached by what it is not. Thus, the volume parameter can be defined as a descriptor of the solute's capacity to enter nonpolar interactions with the aqueous and organic phases, i.e., hydrophobic interactions and dispersion forces. Whether cavity formation also plays a role is debatable and will not be discussed here.

To simplify the vocabulary, it is convenient to equate with **hydrophobicity** the nonpolar interactions encoded by the steric term. In this nomenclature, hydrophobicity is not synonymous with lipophilicity, but a mere component of it.

4.2.5 Recognition Forces Encoded in Lipophilicity

At this point, a general and qualitative expression of Eq. (4) can be written, namely [3]:

Lipophilicity = Hydrophobicity − Polarity (7)

Such an expression allows us to compare section 4.2.1 with sections 4.2.2–4.2.4. This is done in Fig. 1, where the polar component of lipophilicity is seen to correspond to ion-dipole bonds, hydrogen bonds, orientation forces and induction forces, while the hydrophobic component corresponds to dispersion forces and hydrophobic interactions.

Only a limited number of recognition forces cannot find expression in lipophilicity as conventionally measured, namely ionic bonds, charge transfer interactions and aryl/aryl stacking interactions. The latter two would require an aromatic solvent, e.g., benzene or nitrobenzene, to be used in partitioning experiments. As for ionic interactions, they might perhaps be approachable in HPLC using an ionic stationary phase, but the problem of counterions and their influence is far from solved.

4.3 Intramolecular Interactions Affecting Lipophilicity

Functional groups in solute molecules interact with each other in a number of ways depending on their own electronic and steric properties, on the number and nature of interconnecting bonds, and on intramolecular distances. Schematically, a number of dichotomic distinctions can be made, e.g., electronic versus steric effects, or through-bond versus through-space interactions. However, such distinctions may be misleading since they tend to neglect overlaps and intermediate cases. The headings and subheadings in this section present one possible classification in which intramolecular interac-

tions are considered first for themselves (sections 4.3.1–4.3.3), and then as influenced by isomerism and other aspects of molecular polymorphism (section 4.4).

4.3.1 Electronic Conjugations

4.3.1.1 In Aromatic Systems

Substituents in aromatic rings may influence each other in a number of ways depending on their chemical nature, mutual position, and the presence of other substituents. For example, methyl groups have relatively little electronic interactions with the aromatic ring and with each other. Their incremental contribution to the lipophilicity of, e.g., xylenes, is additive as expected and independent from their relative position (*ortho* versus *meta* versus *para*). Such groups are "well-behaved" in lipophilic fragmental systems [7–10].

In contrast, many groups interact strongly by resonance with the aromatic ring, and these groups must necessarily interact with each other. Such interactions are particularly marked in *ortho* and *para* isomers. Substituents in *ortho* position may display a further, through-space level of interaction, namely internal H-bonds (which increase lipophilicity) or steric hindrance with out-of-plane rotation (which decreases lipophilicity). These electronic and steric interactions may strongly perturb the lipophilic increment of some substituents, rendering difficult the calculation of reliable log P values by fragmental systems. The use of correction factors may improve results in some cases, but with serious limitations due to large differences in substituent characteristics and intensity of interaction.

As an example, we present here a systematic study in which the lipophilicity of 75 disubstituted benzene derivatives (i.e., 25 triplets of *ortho, meta* and *para* isomers) was measured as capacity factors (log k_w) in reversed-phase HPLC (RP-HPLC) [11]. The substituents, e.g., CH_3, Cl, OH, NH_2, OCH_3, COOH, $COOCH_3$, $CONH_2$, SO_2CH_3 and SO_2NH_2, were combined pairwise in a variety of possibilities. Substituent constants were first determined from the corresponding monosubstituted benzenes, and used to predict the capacity factors of the 75 disubstituted benzenes. The differences between predicted and measured log k_w values ranged from negligible (e.g., dimethylbenzenes) to very high (e.g., nitrophenols). In about two thirds of cases, such interactions resulted in an increased lipophilicity compared with the expected value, while in about one-third of cases the measured capacity factors were lower than expected.

The measured and predicted capacity factors could be correlated by multiple linear regression (Eq. (8)):

$$\log k_{exp} = 0.943(\pm 0.043) \log k_{calc} + 0.790(\pm 0.168) \varrho\sigma + 0.255$$
$$(\pm 0.208) \varrho\sigma_o + 0.297(\pm 0.050) I_o + 0.170(\pm 0.080)$$
$$n = 75, r^2 = 0.976, s = 0.136, F = 520 \tag{8}$$

where log k_{exp} and log k_{calc} are the experimental and calculated log k_w values, respectively, $\varrho\sigma$ is a parameter expressing the mutual electronic influence of two substitutents in *meta* or *para* position, $\varrho\sigma_o$ is the same parameter for *ortho* disubstitution, and I_o is an indicator of *ortho* effects taking the value of + 1 or + 2 in case of weak or strong in-

ternal H-bonds, and -1 or -2 in case of out-of-plane rotation producing loss of resonance.

Interestingly, Eq. (8) was challenged with a test set of 11 tri- and tetrasubstituted benzenes, with a good correlation between predicted and measured $\log k_w$ values ($r^2 = 0.976$).

4.3.1.2 Across Aliphatic Segments

Interactions of functional groups separated by aliphatic segments can be caused by a variety of effects, e.g., H-bonds between donor and acceptor (section 4.3.2.2), or hydrophobic interactions between two apolar moieties (section 4.3.3.1). In many cases, however, through-space interactions may be present, either between polar groups (section 4.3.2.1) or as result of internal electrostatic bonds (section 4.3.2.2 and 4.3.4.2). Indeed, electronic interactions that occur across aliphatic segments without involving a through-space/conformational component have seldom been reported in structure-lipophilicity relationship studies. In other words, the unambiguous characterization of hyperconjugation as a factor influencing lipophilicity is insufficiently documented in the literature.

An example is provided by ω-functionalized alkylbenzenes and alkylpyridines (Fig. 2), where some partly understood effects were seen [12, 13]. In phenylalkanols and phenylalkylamines (Fig. 2), the lower homologs (benzyl alcohol and benzylamine) were more lipophilic by 0.11 log P unit than predicted from the sum of their fragmental constants. The higher homologs ($n = 2$–4) were slightly less lipophilic by 0.10. This suggests a modest influence on lipophilicity caused by hyperconjugation between the phenyl ring and the functional group, but across one carbon atom only.

As compared with phenylalkanols and phenylalkylamines, various pyridylakanols, pyridylalkylamines and pyridalalkanamides (where $n = 1$–5) showed large deviations from calculated log P values. The fact that these deviations were comparable in each triplet of α-, β- and γ-regioisomers excludes internal H-bond formation between the pyridyl nitrogen and the terminal functionality as the cause of such deviations. As for the phenyl analogs, but in a more marked way, the pyridyl analogs with $n = 1$ were more lipophilic than calculated (by an average of 0.46, 0.08 and 0.27 for the alcohols, amines and amides, respectively). This again could be due to some hyperconjugative effects. However, the clear difference between the OH, NH_2 and $CONH_2$ groups would tend to implicate additionally a through-space proximity effect between polar groups, i.e, between the pyridine ring and the terminal group. Such effects are discussed in section 4.3.2.1.

All higher homologs ($n = 4$ and 5) were markedly less lipophilic than calculated, but no explanation could be offered for such intramolecular effects.
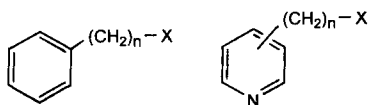
n = 1–5

X = OH, NH_2, $CONH_2$

Figure 2. Chemical structure of ω-functionalized alkyl-benzenes and alkylpyridines used as model compounds to assess the influence of intramolecular electronic interactions across aliphatic segments [11, 12].

4.3.2 Interactions Involving Polar Groups

Polar groups decrease lipophilicity by characteristic increments. When two or more such groups are present, the solute is often found to be more lipophilic than calculated by the simple addition of increments, implying that two or more polar groups may interact to prevent each other from expressing its full polarity. This phenomenon is very well known and amply documented in the literature [7–9]. There is thus little need for us to go into too many details in describing such intramolecular effects, and we shall simply restrict ourselves to discriminating between the various mechanisms by which polar groups may interact intramolecularly to increase lipophilicity. However, because such mechanisms act seldom alone, their discrimination is not always straightforward.

4.3.2.1 Proximity Effects Between Two Neutral Polar Groups

The fragmental system of Rekker afforded the first incremental method of calculation derived from and applicable to aliphatic moieties and molecules. In this system, a limited number of correction factors are necessary to take into account intramolecular interactions such as electronic conjugation and proximity effects. These occur between polar groups defined as being electronegative functionalities. In the 1979 version of the system [7], a correction factor of $+0.84$ must be added to the sum of fragmental values when such groups are separated by one sp^3-carbon. The correction factor decreases to $+0.56$ when the two groups are separated by two carbons, and becomes statistically nonsignificant for three carbons.

The point of relevance in Rekker's fragmental system is that the correction for proximity is independent of the chemical nature of the polar groups, which may be halogens, H-bond acceptors, or H-bond donors/acceptors. This implies that in Rekker's fragmental system internal H-bond formation is not regarded as playing an explicit role. A simple molecular explanation for such proximity effects is to consider each polar group as surrounded by a hydration sphere, proximity bringing these spheres to overlap and loose influence. The Molecular Lipophilicity Potential (MLP), which is to be discussed in chapter 12, allows a simpler picture to emerge. In this approach, the polarity on the solvent-accessible molecular surface is computed for its intensity and extension. The former is seen to increase, and the latter to decrease, when two polar groups become proximal. However, the deep reason for this proximity effect remains to be understood.

While empirically very useful, Rekker's assumption is an oversimplification for two major reasons. First, there is a problem of intensity of effects since it is difficult to understand how different polar groups could contribute identically to proximity effects. The second reason is related to the first, namely that the formation of internal H-bonds should be given explicit consideration to obtain more precise log P estimates.

It is a feature of the fragmental system of Hansch and Leo [8] that it acknowledges the formation of internal H-bonds and classifies polar groups into two classes, namely H-bond-forming (designated as H) and non-H-bond-forming (i.e., halogens, designated as S). As a result, three types of proximity effects are recognized, namely those resulting from S/S, H/S and H/H interactions.

Multiple halogenation on the same or adjacent carbon atoms, a typical example of S/S interaction, results in higher log P values than predicted by simple additivity. This is assumed to be due to the localized dipoles being partly shielded from water by the neighbouring halogens. In the fragmental system of Hansch and Leo (version 1979), the correction factors for N gemical halogens are $+ 0.30 N$, $+ 0.53 N$ and $+ 0.72 N$ for $N = 2$, 3 or 4, respectively [8]. In the case of vicinal halogens, the correction factors are $+ 0.28 (N - 1)$. These numbers are average values that neglect other structural factors such as conformational behavior and electronic distribution. Nevertheless, they are clearly significant and testify of the significant of polar proximity effects not involving H-bonds.

In the fragmental system of Hansch and Leo, distinct correction factors are used to account for proximity effects between H-bond-forming groups. These factors, which are always positive, depend on the nature of the polar groups and decrease with increasing number of intervening carbon atoms (1, 2, or 3). Like in Rekker's system, these factors do not incorporate internal H-bonds, which are factorized separately as discussed below. Hence, the system of Hansch and Leo also recognizes the decrease in hydrophilicity due to the mere proximity of polar groups.

4.3.2.2 Internal H-Bonds

The most important intramolecular electrostatic interactions affecting lipophilicity are ionic bonds (see section 4.3.2.3), H-bonds discussed here, and perhaps also dipole-dipole interactions. As stated above, the system of Hansch and Leo allows the average influence of H-bonds on log P to be estimated [8]. Thus, intramolecular H-bonds involving oxygen or nitrogen receive a correction factor of $+ 1.0$ and $+ 0.60$, respectively. This is a rather marked effect, but the difficulty when calculating a log P value is to decide whether such a bond exists or not in a given solute. Indeed, a correction factor for H-bond can be introduced *a priori* (subject to experimental verification) based on knowledge or expectation, or *a posteriori* if a discrepancy is found between measured and calculated log P.

Water being a dipole with strong H-bond donor and acceptor properties, it will interact electrostatically with polar groups in a solute and prevent it from forming internal electrostatic bonds. Water-saturated octanol (approximately four molecules of octanol for one of water) is a relatively polar solvent with a H-bond acceptor basicity as good as that of water, and a H-bond donor acidity markedly smaller than that of water. As a result, the tendency of solutes to form internal H-bonds is usually comparable in octanol and in water. Thus, octanol is by far not the best solvent to observe the formation of internal H-bonds, and more generally hydrophilic folding (see section 4.3.2.4). In contrast, apolar solvents such as alkanes or poorly polar solvents such as dibutyl ether (which dissolves about 0.1 % water) will strongly favor internal H-bonds and more generally hydrophilic collapse.

Thus, a thermodynamic study of the partitioning of isomeric and homologous pyridylalkanamides (see Fig. 2) afforded some insight into the underlying mechanisms [14]. In dibutyl ether/water, a single mechanism prevailed for the partitioning of all solutes, and the solutes able to form internal H-bonds (i.e., the 2-pyridyl derivatives)
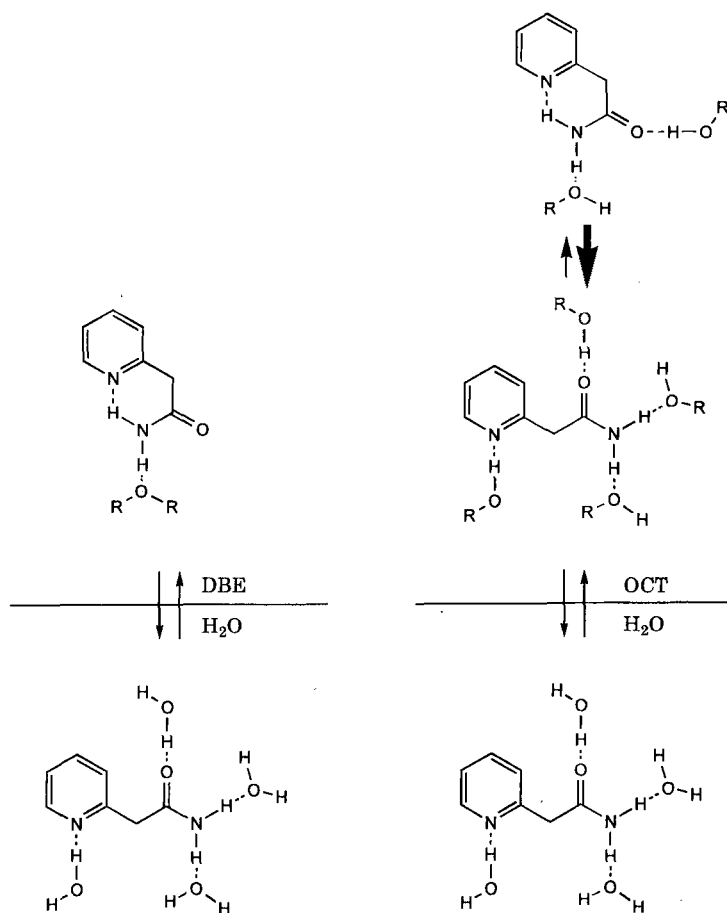
Figure 3. Mechanistic difference in the partitioning of 2-pyridylalkanamides in the systems dibu-
tyl ether/water (DBE/H_2O) and *n*-octanol/water (OCT/H_2O). In DBE, the solutes appear to exist
excusively in an internally H-bonded conformation that masks in part the hydrophobic segment
of the side-chain. In octanol, the solutes are in equilibrium between a small population of inter-
nally H-bonded conformers and a predominant population of extended conformers [14].

were more lipophilic than expected. This indicated that the 2-pyridylalkanamides ex-
isted exclusively as internally H-bonded conformers in dibutyl ether. In contrast, all
evidence from octanol/water partition coefficients indicated that in this solvent the 2-
pyridylalkanamides existed in a conformational equilibrium with only a small popula-
tion of internally H-bonded conformers (Fig. 3) [14].
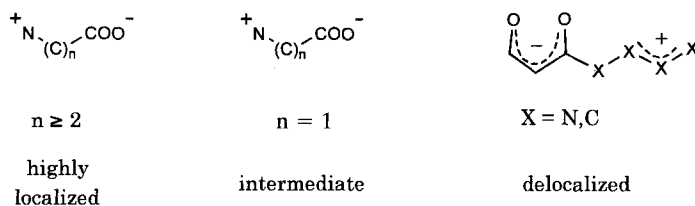
Figure 4. A schematic representation of zwitterions ranging from the highly localized (e.g. β- and γ-amino acids) to the delocalized (e.g. guanidinium/enolates).

4.3.2.3 The Case of Zwitterions

Zwitterions represent a particular and as yet insufficiently explored type of solutes. *A priori*, their intramolecular and intermolecular interactions differ from those of other solutes, although one must be aware that the differences between highly polarized nonionic solutes and zwitterions may be more quantitative than qualitative, especially for delocalized zwitterions. Indeed, we have come to feel that there is a need to distinguish between two major types of zwitterions, namely the more usual ammonium-carboxylates, and the more delocalized guanidinium or amidinium enolates. Figure 4 presents a continuum between β- and γ-amino acids and some highly delocalized structures as for example found in oxicams and other nonsteroidal antiinflammatory agents [15].

In Rekker's fragmental system, the aliphatic and aromatic carboxylate fragments (COO^-) receive an incremental value of -5.00 and -4.13, respectively [7]. The case of the ammonium group is less clear, but a value of -3.73 has been proposed for the aliphatic-NH_3^+ fragment [16]. By simply adding fragments, one would arrive at a predicted log P value for zwitterionic glycine (^-OOC-CH_2-NH_3^+) of -8.21, when the actual value is -3.00 [17]. The difference between the predicted and experimental values clearly demonstrates that the two charged groups in α-amino acids interact strongly, and gives a fair estimate of the importance of this interaction. The nature of this interaction is certainly a dual one, involving partial neutralization *via* delocalization across the sp^3-carbon (through-bond interaction), plus an internal ionic bond (through-space interaction) which further contributes to partial neutralization.
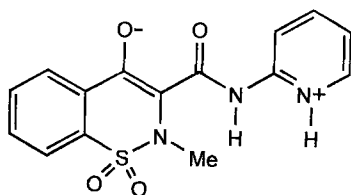
Figure 5. The structure of zwitterionic piroxicam as existing in the pH range 2–5.

In zwitterions such as piroxicam in the pH range 2–5 (Fig. 5), the two charges are formally separated by 5 atoms, but molecularorbital calculations reveal that due to marked delocalization the effective distance between the centroid of positive and negative regions becomes less than 5 î. This appears to result in a marked partitioning of the zwitterion into octanol [15].

4.3.2.4 Hydrophilic Collapse

Hydrophilic collapse is defined as a conformational change by which a solute maximizes the number and strength of internal electrostatic bonds (mainly H-bonds) and thus partly masks some of its polar groups from the solvent. The drive for hydrophilic collapse comes from a nonpolar solvent, the solute hiding its polar groups away from this nonpolar solvent in order to become less polar and resemble that solvent. As such, hydrophilic collapse is the opposite of hydrophobic collapse discussed in section 4.3.3.3, and alone or together with the latter may account for a chameleonic behavior (see section 4.4.4).

An example of hydrophilic collapse is offered by cyclosporin A (CsA), an immunosuppressive cyclic undecapeptide widely used in clinical organ transplantation. In water, CsA exists as a mixture of conformers characterized by H-bonding groups (in Abu^2, Val^5, Ala^7 and Ala^8) pointing away from the ring, i.e., towards the solvent [18]. This is in fact the active conformation of CsA as bound to cyclophilin. In apolar solvents, the conformational state of CsA is very different, being characterized by four major internal H-bonds (Abu^2-to-Val^5, Val^5-to-Abu^2, Ala^7-to-$MeVal^{11}$, and Ala^8-to-$MeLeu^6$). Thus, CsA in apolar solvents turns a number of its polar groups towards the interior of the ring. The driving force for the creation of this polar interior is the formation of the intramolecular H-bonds.

Table 2. Partition coefficients (log P) and H-bonding capacity [$\Delta(\log P_{octanol-heptane})$] of model -solutes and cyclosporin A [18]

Solute	log $P_{octanol}$	log $P_{heptane}$	$\Delta\log P$
Not forming internal H-bonds			
phenol	1.46	− 0.82	2.22
p-nitrophenol	1.77	− 2.11	3.88
benzamide	0.65	− 2.45	3.10
p-fluorobenzamide	0.96	− 2.34	3.25
acetanilide	1.16	− 1.54	2.70
p-fluoroacetanilide	1.47	− 1.57	3.04
cyclo(Phe-Phe)	1.59	< − 3.0	> 4.5
cyclo(Trp-Tyr)	1.05	< − 3.0	> 4.1
Forming internal H-bonds			
o-nitrophenol	1.68	1.04	0.64
o-fluorobenzamide	0.64	− 1.47	2.11
o-fluoroacetanilide	0.96	− 0.69	1.65.
CyclosporinA	2.92	1.40	1.52

The conformational behavior of CsA is reflected in its partitioning. As shown in Table 2, solutes unable to form internal H-bonds have $\Delta\log P$ values (i.e., $\log P_{octanol}$ minus $\log P_{heptane}$, a measure of the H-bonding capacity) of > 2, whereas for those able to form internal H-bonds the value is < 2 [18]. These numbers reflect the fact, already discussed in section 4.3.2.2, that hydrophilic collapse and particularly the formation of internal H-bonds are favored in apolar solvents (e.g., alkanes) significantly more than in solvents of low polarity (e.g., octanol).

4.3.2.5 Proximity Effects Between Polar and Nonpolar Groups

In sections 4.3.2.1 and 4.3.2.3, we have seen how the proximity of two polar groups in a solute decreases its expected hydrophilicity, i.e., produces a higher than expected lipophilicity. The interpenetration of hydration spheres, or a decrease of the polar molecular surface, are two complementary models to explain the many observations of this type and visualize their mechanism.

Interestingly, the same pictorial models allow another important phenomenon to be understood, namely the decrease in their hydrophobic increment experienced of apolar moieties in the proximity of highly polar groups. A highly illustrative example is provided by amino acids, where the carboxylate and ammonium groups decrease the hydrophobicity of neighboring CH_2 and CH_3 units in a distance-dependent manner [17]. When examining the $\log D$ of α-amino acids (Fig. 6) determined at isoelectric pH (i.e., the $\log P$ of the zwitterions), it was found that the hydrophobic increment of the CH_2 groups did not increase as predicted. Indeed, the increments in the series $R = H$ (glycine), $R = CH_3$ (alanine), $R = CH_2CH_3$ (α-aminobutyric acid), $R = CH_2CH_2CH_3$ (norvaline) and $R = CH_2CH_2CH_2CH_3$ (norleucine) were 0.23, 0.24, 0.42, and 0.57. Thus, only the fourth CH_2 unit could express a full hydrophobicity (0.57), suggesting that the first three CH_2 units are partly masked from the solvent by the polar groups [17].

A comparable observation was made with zwitterionic ω-amino acids (Fig. 6), with the first six homologs ($n = 1$ to 6) having $\log P$ values between -3.0 and -3.1. Only the seventh CH_2 group did contribute to an increased lipophilicity ($\log P$ for $n = 7$: -2.55) [17]. Here, the results indicate that the first three CH_2 groups attached to the carboxylate or the ammonium group are masked from the solvent.

$$H_3N^+\diagdown_{\underset{\underset{R}{|}}{CH}}\diagup^{COO^-} \qquad H_3N^+\diagdown_{(CH_2)_n}\diagup^{COO^-}$$
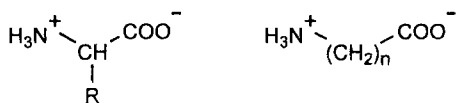
Figure 6. Amino acids used to demonstrate the "masking" of hydrophobic groups by polar groups.

4.3.3 Steric/Hydrophobic Effects

4.3.3.1 Shielding of Polar Groups

Alkyl or aryl moieties may form intramolecular hydrophobic interactions if this is compatible with their relative position and the compound's flexibility. As a rule, such internal hydrophobic interactions are characteristic of folded conformers, and render the solute less lipophilic than predicted by partly masking the hydrophobic moieties from the solvent [19].

For example, several compounds in a series of xanthine derivatives revealed a smaller than expected lipophilicity (by about 1.5 log P units) [20]. Inspection of structures showed that the outliers had a bulky N3-substituent (usually an isobutyl group), and a very large and hydrophobic C8-substituent, usually a (4-benzhydrylpyperazin-1-yl)ethyl group. Conformational analysis revealed stable conformers having the N3- and C8-substituents in hydrophobic contact (Fig. 7). In contrast, compounds unable to form such a hydrophobic interaction displayed a predictable lipophilicity, for example an analog with a hydroxylated side-chain (i.e., no internal hydrophobic bond possible). This led to the paradoxical observation that the introduction of a hydroxyl group in a congener with an internal hydrophobic bond *increased* lipophilicity by c. 1.2 log P units.
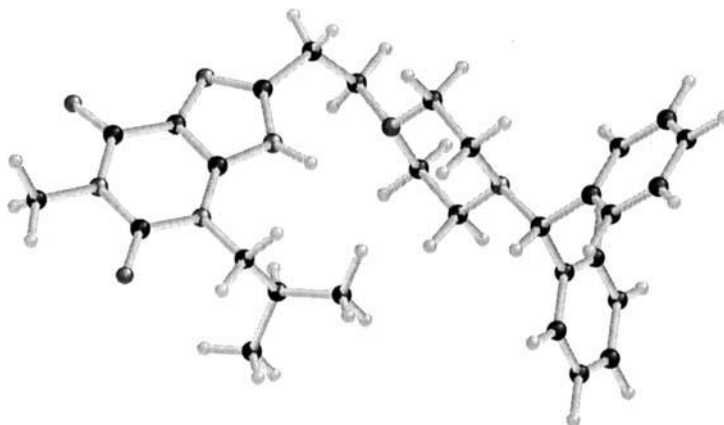
Figure 7. Low-energy conformer of a xanthine derivative showing the hydrophobic fit between the N3-isobutyl and (4-benzhydrylpyperazin-1-yl)propyl groups. Such a compound is less lipophilic than expected by c. 1.5 log P units [20].

4.3.3.3 Hydrophobic Collapse

The concept of hydrophobic collapse derives from the effect described in the previous section but conveys the idea of a phenomenon of particular magnitude [21, 22]. In other words, hydrophobic collapse as generally understood should be restricted to solutes of comparatively large molecular weight (several hundred or more) and containing a number of hydrophobic moieties able to come close together to create a hydrophobic core, for example a variety of synthetic peptides [21, 22] (see section 4.4.4).

A representative example fulfilling these conditions is offered by the antitumor drug taxol and its semisynthetic analog taxotere [23]. In aqueous solution these compounds exist in a conformation characterized by hydrophobic clustering of the 2-benzoyl, 3'-phenyl and 4-acetyl groups.

The phenomena of hydrophobic collapse, hydrophilic collapse and chameleonic behavior are discussed globally in section 4.4.4.

4.4 Structural Factors Influencing Intramolecular Interactions

The previous sections have presented a number of possibilities by which various moieties in a molecule may interact to influence the partitioning behavior of the solute. In a schematic manner, such interactions depend on the following factors:

Table 3. The influence of geometric factors and molecular states on intramolecular interactions expressed in lipophilicity

Interactions	Structural factors				
	Regio-isom-erism	Tautom-erism	Diastereo-isomerism	Conforma-tional isomerism	Ionization
Electronic conjugations					
● In aromatic systems	+	+			+
● Across aliphatic segments	+				+
Interactions involving polar groups					
● Proximity effects between two neutral polar groups	+	+	+	+	
● Internal H-bonds	+	+	+	+	+
● The case of zwitterions	+		+	+	+
● Hydrophilic collapse	+	+	+	+	+
● Proximity effects between polar and nonpolar groups	+	+	+	+	+
Steric/hydrophobic interactions					
● Shielding of polar groups	+	+	+	+	+
● Hydrophobic interactions	+		+	+	
● Hydrophobic collapse	+		+	+	

- the chemical and physico-chemical nature of the moieties, e.g., their high or low polarity;
- their distance from each other;
- the nature and number of interconnecting atoms.

In other words, a number of structural factors (geometric factors and molecular states) will influence intramolecular interactions and hence solubility and partitioning. Various possibilities are schematized in Table 3, some of which will be briefly discussed below.

4.4.1 Positional Isomerism

Positional isomerism is a geometric factor of obvious significance in lipophilicity. In fact, it may be convenient to distinguish between:

- **Regioisomerism,** which relates positional isomers whose interconversion is a high-energy process.
- **Tautomerism,** which involves the low-energy migration of a proton from one heteroatom to another.

The example discussed in section 4.3.1.1 is a fit illustration of the importance of positional isomerism in the lipophilicity of bisubstituted benzenes. Similarly, the proximity effects discussed in sections 4.3.1.2 and 4.3.2.1 are obviously critically dependent on the number of interconnecting atoms. The same applies to the possibility of forming internal H-bonds or even hydrophilic collapse (sections 4.3.2.2 and 4.3.2.4), to partial charge neutralizations (section 4.3.2.3), to various other proximity effect (sections 4.3.2.5 and 4.3.3.1), and to the possibility of forming hydrophobic interactions or even hydrophobic collapse (sections 4.3.3.2 and 4.3.3.3).

There are also a few examples in the literature where tautomerism was studied *per se* as a factor influencing lipophilicity. A case in point is that of proxibarbal, an antimigraine drug that exists in ring-chain tautomeric equilibrium with the two diastereomers of valofan (Fig. 8). Because the interconversion is rather rapid ($t_{1/2} \approx 150$ min at pH 7.4 and 20 °C), direct measurement of partition coefficients is not feasible. In con-

cis-valofan   proxibarbal   *trans*-valofan
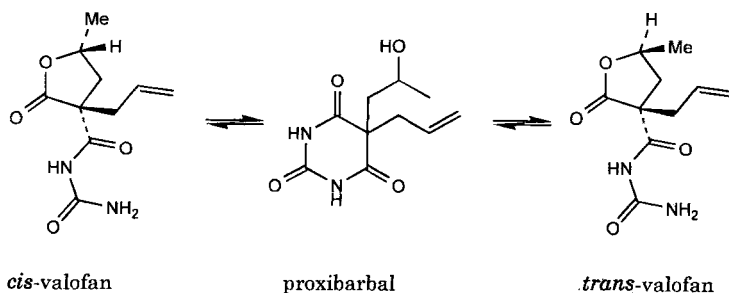
Figure 8. The equilibrium between proxibarbal and its ring-chain tautomers *cis*- and *trans*-valofan. Only relative configurations are implied.

trast, a kinetic study allowed rate constants of interconversion and transfer to be determined simultaneously, affording the following log P_{oct} values: proxibarbal -0.05, *cis*-valofan 0.28, and *trans*-valofan 0.41 [24]. The differences between tautomers are not very large (0.33 and 0.46), but they are most probably due to differences in the H-bonding capacity of proxibarbal and valofan. A modest difference between diastereomers is also noted, a factor discussed in the following section.

4.4.2 Stereoisomerism

Stereoisomerism is another geometric factor of obvious significance in lipophilicity. Again, it is convenient to distinguish between:

- **Diastereomerism,** which relates diastereomers whose interconversion is a high-energy process.
- **Conformational isomerism,** which involves the low-energy interconversion of stereo-isomers.

Systematic investigations on the compared lipophilicity of diastereomers are few. In one study, 36 pairs of diastereomers were compared [25]. The differences in log P values ranged from 0.0 to 1.0. While no quantitative interpretation proved possible, the data showed that the water-accessible surface area, and not the H-bonding capacity, was the major structural determinant in the differences of lipophilicity between relatively rigid diastereomers containing one or two polar groups. This was interpreted as a consequence of the perturbation of hydrophobic hydration exerted by the polar groups at an *endo* or a *syn* position, leading to a decrease in hydrophobicity.

Several examples discussed in section 4.3 aptly illustrate the significance of conformational factors on lipophilicity. This is particularly true for the possibility of forming internal H-bonds or even hydrophilic collapse (sections 4.3.1.1, 4.3.2.2 and 4.3.2.4). Proximity effects between polar and nonpolar groups (sections 4.3.2.5 and 4.3.3.1), as well as the possibility of forming hydrophobic bonds or even hydrophobic collapse (sections 4.3.3.2 and 4.3.3.3), are also discussed above.

Other studies can be found in the literature whose specific aim was to examine relations between lipophilicity and conformational behavior [26]. For example, a series of *N*-hydroxyureas (R-NH-CO-NR'-OH) showed complex tautomeric and conformational behavior depending on the nature of R and R' (H, *n*-alkyl or branched alkyl) [27, 28]. The major factor ultimately influencing lipophilicity was the possibility of forming or not an internal H-bond between the carbonyl oxygen and the hydroxyl proton. Solutes able for structural reasons to form this bond proved more lipophilic than those unable to form it.

4.4.3 Ionization

The possibility for a solute to exist in neutral or charged states will obviously have a major impact on its partitioning behavior. First, the solute will exhibit pH-dependent partitioning, making it indispensable to distinguish between its partition coefficients

(solvent-dependent) and its distribution coefficients (pH- and solvent-dependent) [3] (see also Chapter 7). Second – and more relevant to the present context – the fact that a polar group exists in a neutral or charged state may dramatically alter the intramolecular interactions involving this group. Thus, ionization will affect electronic conjugation (section 4.3.1), proximity effects between polar groups (section 4.3.2.1), internal H-bonds (section 4.3.2.2), internal ionic bonds and other ionic interactions (section 4.3.2.3), hydrophilic collapse (section 4.3.2.4), and the shielding of nonpolar groups by polar groups (section 4.3.2.5). In addition and indirectly, the various steric/hydrophobic effects (section 4.4.3) will also be affected.

4.4.4 Molecular Size and Chameleonic Behavior

In a most stimulating account, Jiang has commented on aggregation and self-coiling in organic molecules, stressing their major significance in the functioning of biomolecules and biomacromolecules [29]. The point to be made here is that the phenomenon of self-coiling is a capital one not only for endogenous compounds, but also for drugs and other xenobiotics and their metabolites. For a variety of (presumably historical) reasons, medicinal chemists refer to hydrophobic collapse rather than self-coiling, the term hydrophilic collapse being a more recent acquisition (section 4.3.2.4).

Self-coiling, be it due to hydrophobic or to hydrophilic collapse, requires a certain number of structural conditions to be fulfilled, namely **functionalities, flexibility** and **size**. In other words, the compound must a) contain the necessary functional groups, b) be flexible enough for these functional groups to interact *via* electrostatic and/or hydrophobic forces, and c) a large enough for collapse to occur at all (see section 4.3.3.3).

As a result of hydrophobic and/or hydrophilic collapse, a solute may become more polar in polar solvents and more lipophilic in lipidic solvents. In effect, such a solute to some extent adapts its lipophilicity to that of the medium, thereby behaving analogously to a **chameleon**, which changes color to resemble that of the environment.

An example of chameleonic behaviour can be found with the two major metabolites of morphine, namely its 3-*O*-glucuronide and its 6-*O*-glucuronide. In a RP-HPLC system, these two conjugates displayed a much higher than expected lipophilicity which could explain some of their rather unusual pharmacokinetic properties. Conformational analysis and computation of the MLP (see chapter 12) suggested that these compounds can indeed adopt two low-energy conformations, namely a population of folded conformers with partly masked polar groups and increased hydrophobic surface, and a population of extended conformers with maximally exposed polar groups and minimized hydrophobic surface [30, 31]. These two metabolites are exemplary in that they do fulfil the three conditions listed above for chameleonic behavior to be displayed at all. However, it must be noted that morphine *O*-glucuronides do not appear to display a higher than expected lipophilicity in octanol/water systems, in agreement with the arguments discussed in section 4.3.2.2.
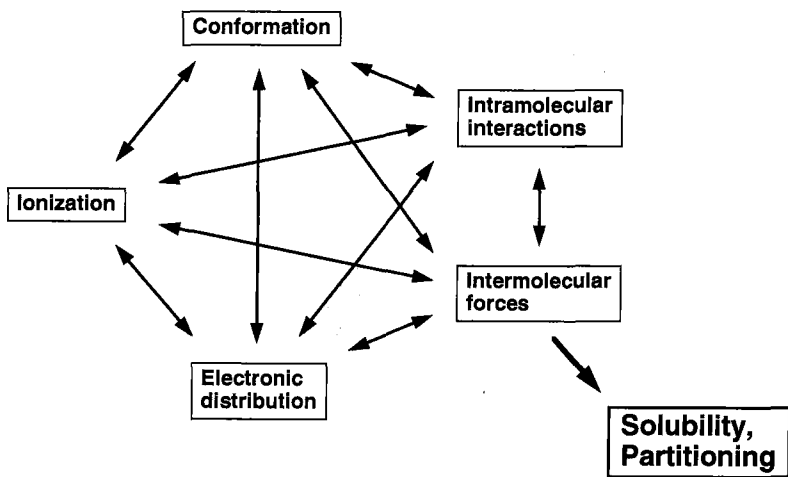
Figure 9. Interrelated factors influencing intramolecular interactions, intermolecular forces and partitioning.

4.5 Outlook: Molecular Polymorphism in Drug Design

A discussion of intermolecular forces encoded in lipophilicity (section 4.2), intramolecular interactions influencing lipophilicity (section 4.3), and structural factors influencing intramolecular interactions (section 4.4) shows very clearly their interconnectedness and interdependence. A schematic representation of these relation is proposed in Fig. 9, which can be simplified as follows:

Table 4. Major intramolecular and intermolecular processes underlying molecular polymorphism

Intramolecular processes (formally)	Intermolecular processes
• Valence isomerization	• Prototropic equilibria (acid-base behavior)
• Tautomerism	• Hydration and solvation
• Inversion of configuration	• Self-association
• Conformational behavior	• Binding to macromolecules and consequences thereof
• Internal hydrophobic bonds and hydrophobic collapse	
• Internal electrostatic bonds and hydrophilic collapse	

- solubility and partitioning properties are first a function of intermolecular forces;
- intermolecular forces and intramolecular interactions influence each other in an intimate and undissociable way;
- in addition,intermolecular forces and intramolecular interactions depend on, and influence in return, such molecular properties as conformation, electronic distribution and ionization.

Figure 9 brings us back to the multilevel description of molecular structure and properties discussed in section 4.1. It emphasizes the holistic character of molecular structure and properties, which should most fruitfully be approached in the global perspective of molecular polymorphism (Table 4). Only in this way can we hope to progress in our conceptual understanding of molecular structure and in describing it with improved depth and sophistication.

Acknowledgements

References

[1] Testa, B., and Kier, L. B., *Med. Res. Rev.* **11**, 35–48 (1991)
[2] Carrupt, P. A., El Tayar, N., Karlén, A., and Testa, B., *Methods Enzymol.* **203**, 638–677 (1991)
[3] Van de Waterbeemd, H., and Testa, B., The parametrization of lipophilicity and other structural properties in drug design. In: *Advances in Drug Research*, Vol **16**. Testa B. (Ed.). Academic Press: London; 85–225 (1986)
[4] El Tayar, N., Tsai, R. S., Testa, B., Carrupt, P. A., and Leo, A., *J. Pharm. Sci.* **80**, 590–598 (1991)
[5] El Tayar, N., Testa, B., and Carrupt, P. A., *J. Phys. Chem.* **96**, 1455–1459 (1992)
[6] Vallat, P., Gaillard, P., Carrupt, P. A., Tsai, R. S. and Testa, B., *Helv. Chim. Acta* **78**, 471–485 (1995)
[7] Rekker, R. F., and De Kort, H. M., *Eur. J. Med. Chem.* **14**, 479–488 (1979)
[8] Hansch, C., and Leo, A., *Substituent Constants for Correlation Analysis in Chemistry and Biology.* Wiley: New York 1979
[9] Leo, A. J., *Chem. Rev.* **93**, 1281–1306 (1993)
[10] Mayer, J. M., van de Waterbeemd, H., and Testa, B., *Eur. J. Med. Chem.* **17**, 17–25 (1982)
[11] Tsantili-Kakoulidou, A., El Tayar, N., van de Waterbeemd, H. and Testa, B., *J. Chromatogr.* **389**, 33–45 (1987)
[12] Mayer, J. M., Testa, B., van de Waterbeemd, H., and Bornand-Crausaz, A., *Eur. J. Med. Chem.* **17**, 453–459 (1982)
[13] Mayer, J. M., Testa, B., van der Waterbeemd, H., and Bornand-Crausaz, A., *Eur. J. Med. Chem.* **17**, 461–466 (1982)
[14] Repond, C., Mayer, J. M., van de Waterbeemd, H., Testa, B., and Linert, W., *Int. J. Pharmaceut.* **38**, 47–57 (1987)
[15] Takács-Novák, K., Kökési, J., Podányi, B., Noszál, B., Tsai, R. S., Lisa, G., Carrupt, P. A., and Testa, B., *Helv. Chim. Acta* **78**, 553–562 (1995)

[16] Testa, B., and Murset-Rossetti, L., *Helv. Chim. Acta* **61,** 2530–2537 (1978)
[17] Tsai, R. S., Testa, B., El Tayar, N., and Carrupt, P. A., *J. Chem. Soc. Perkins Trans.* **II,** 1797–1802 (1991)
[18] El Tayar, N., Mark, A. E., Vallat, P., Brunne, R. M., Testa, B., and van Gunsteren, W. F., *J. Med. Chem.* **36,** 3757–3764 (1993)
[19] Hansch, C., and Anderson, S. M., *J. Org. Chem.* **32,** 2583–2586 (1967)
[20] Walther, B., Carrupt, P. A., El Tayar, N., and Testa, B., *Helv. Chim. Acta* **72,** 507–517 (1989)
[21] Wiley, R. A., and Rich, D. H., *Med. Res. Rev.* **3,** 327–384 (1993)
[22] Rich, D. H. Effect of hydrophobic collapse on enzyme-inhibitor interactions. Implications for the design of peptidomimetics. In: *Perspectives in Medicinal Chemistry.* Testa, B., Kyburz, E., Fuhrer, W., and Giger, R. (Eds.). VCH: Weinheim; 15–25 (1993)
[23] Vander Velde, D. G., Georg, G. I., Grunewald, G. L., Gunn, C. W. and Mitscher, L. A., *J. Am. Chem. Soc.* **115,** 11650–11651 (1993)
[24] Wittekind, H. H., Testa, B., and Estreicher, J., *Helv. Chim. Acta* **71,** 1228–1234 (1988)
[25] Tsai, R. S., Carrupt, P. A., Testa, B., El Tayar, N., Grunewald, G. L. and Casy, A. F., *J. Chem. Res. (S)* 298–299; *(M)* 1901–1920 (1993)
[26] Hopfinger, A. J., and Battershell, R. D., *J. Med. Chem.* **19,** 569–573 (1976)
[27] Parker, G. R., Lemke, T. L., and Moore, E. C., *J. Med. Chem.* **20,** 1221–1225 (1977)
[28] Parker, G. R., *J. Pharm. Sci.* **67,** 513–516 (1978)
[29] Jiang, X. K., *Acc. Chem. Res.* **21,** 361–367 (1988)
[30] Carrupt, P. A., Testa, B., Bechalany, A., El Tayar, N., Descas, P., and Perrissoud, D., *J. Med. Chem.* **34,** 1272–1275 (1991)
[31] Gaillard, P., Carrupt, P. A., and Testa, B., *Bioorg. Med. Chem. Lett* **4,** 737–742 (1994)

5 Lipophilicity Measurement by Reversed-Phase High Performance Liquid Chromatography (RP-HPLC)

Han van de Waterbeemd, Manfred Kansy, Björn Wagner and Holger Fischer

Abbreviations

HIC	Hydrophobic interaction chromatography
IAM	Immobilized artificial membrane chromatography
LSER	Linear solvation energy relationships
MLC	Micellar liquid chromatography
ODP	Octadecylpolyvinyl copolymer
ODS	Octadecylsilane
OS	Octylsilane
QSAR	Quantitative structure-activity relationships
QSRR	Quantitative structure-retention relationships
RP-HPLC	Reversed-phase high-performace liquid chromatography
RP-IPC	Reversed-phase ion-pair chromatography
RPLC	Reversed-phase liquid chromatography
SPC	Structure-property correlations

Symbols

α	Hydrogen bond donor acidity
β	Hydrogen bond acceptor basicity
CLOGP	Calculated 1-octanol/water partition coefficient
$\mathcal{D}$	Function of the dielectric constant of the mobile phase
δ	Hildebrand parameter
$E_T(30)$	Empirical measure of solvent polarity
γ	Surface tension of the mobile phase
φ	Mobile phase composition
HB	Hydrogen bond descriptor
$\log k$ ($\log k'$)	Logarithm of the isocratic capacity factor
$\log k_W$	Logarithm of the polycratic (extrapolated) capacity factor
$\log P_{mw}$	Logarithm of micelle/water partition coefficients
$\log P_{oct}$	Logarithm of 1-octanol/water partition coefficient
π^*	Solvatochromic polarity descriptor
S	Slope of the extrapolation of isocratic capacity factors
t_0	Retention time of an unretained compound
t_r	Retention time of the solute
$V_2/100$	Scaled volume of a solute
Z	Retention parameter calculated from the slope of log k vs log molar concentration of organic modifier

5.1 Historical

The classical studies by Overton and Meyer, and later Fujita and Hansch, demonstrated that partition coefficients (log P values) can be correlated with many biological phenomena. It is beyond doubts that log P are among the most important descriptors for transport processes of a biologically active compound in the body.

However, log P measurements are time-consuming and are limited to a certain, range, e.g., $-3 < \log P_{oct} < 3$. Beyond these limits, the log P values assessed by the shake flask method become unreliable. Alternatives, particularly chromatographic methods, have therefore been evaluated and used successfully to assess lipophilicity of organic compounds. These include thin-layer chromatography (TLC, see Chapter 8), centrifugal partition chromatography (CPC, see Chapter 6), and reversed-phase high-performance liquid chromatography (RP-HPLC or RPLC). The first applications of lipophilicity measurements by HPLC go back to the 1970s [1–4]. A review of this early work can be found in reference [5]; more recent work is reviewed in reference [6].

It is not the intention of the authors of the present chapter to be comprehensive, since the literature on this topic literally exploded over the last 20 years. The most important subjects, critical issues, and selected recent findings will be presented here. Particularly of interest are the various packing materials, since identification of lipophilicity scales correlated with biologically important transport and distribution processes, such as gastrointestinal absorption, skin penetration, blood-brain-barrier uptake, plasma protein binding is of great interest to drug discovery. The interest of certain solid phases is their close resemblance to octanol/water partitioning, which is still considered as the standard system. Particularly some new materials, such as immobilized artificial membranes (IAM) may have a good future, since they mimic transport through biological barriers.

Besides the well-known techniques based on RP-HPLC or RPLC (or RPC), other related chromatographic methods useful for lipophilicity measurements will be briefly mentioned, such as ion pair chromatography (IPC), micelle liquid chromatography (MLC) and hydrophobic interaction chromatography (HIC).

5.2 Principle of Lipophilicity Measurements by RPLC

5.2.1 Description of the Method

Lipophilicity of organic compounds can be measured with any standard HPLC equipment. Usually UV detection can be used, but any other detection method is appropriate (e.g., refractive index or electrochemical). The choice of the stationary and mobile phase are discussed in section 5.3 and 5.4. The lipophilicity index measured by RPLC is derived from the capacity factor k (or sometimes written as k') given by

$$k = (t_r - t_0)t_0 \tag{1}$$

where t_r and t_0 are the retention times of the solute and of an unretained compound. This latter can be the solvent front, or an inorganic salt such as sodium nitrate or potas-
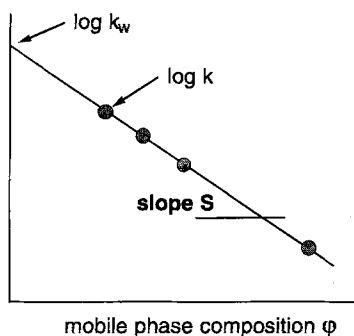
Figure 1. Schematic representation of the measurement of a lipophilicity index by RPLC. Isocratic (log k) capacity factors are obtained from the retention time measured at a selected mobile phase composition; polycratic (log k_w) capacity factors are calculated from an extrapolation procedure.

sium bichromate. The logarithmic form, log k, can be used as a lipophilicity index. Instead of the isocratic log k values (obtained at a single selected mobile phase concentration), frequently polycratic log k_w values are used, which are obtained by extrapolation of isocratic log k values against the mobile phase composition [7–9] (Fig. 1). This procedure should not be confused with the use of a mobile phase gradient, which is a typical HPLC separation technique. The mobile phase in RP-HPLC in the present context consists of a mixture of water and an organic compound, called organic modifier, typically methanol, acetonitrile or tetrahydrofuran. The most common procedure is the linear extrapolation with organic modifier content φ:

$$\log k = S\varphi + \log k_w \tag{2}$$

The slope S has been studied in more detail over the past few years and encodes to a certain extent hydrogen-bonding capability (see section 5.5.2).

It is often observed that the relationships between log k and φ are not linear, e.g., with protonated bases [9] or when using organic modifiers such as tetrahydrofuran and acetonitrile [10, 11]. In such cases two alternatives have been used. Either the linear extrapolation is performed on part of the points, i.e., those with lowest organic modifier concentration, or a quadratic fit is used [12]:

$$\log k = A\varphi^2 + B\varphi + \log k_w \tag{3}$$

A further alternative is to obtain log k_w values from extrapolation from $E_\tau(30)$ plots [13]. The $E_\tau(30)$ scale is an empirical measure of solvent polarity, and is based on charge transfer absorption measurements of 2,6-diphenyl-4-(2,4,6-triphenyl)-n-pyridino)phenolate. Finally, the solvophobic theory of Horvath should be mentioned [14, 15]. This theory was developed to account for the curvature in log k − φ relationships. In its simplified form this equation is:

$$\log k = A + B\mathcal{D} + C\gamma \tag{4}$$

where A, B and C are regression coefficients. A is a function of the mobile phase composition, $\mathcal{D}$ is related to the dielectric constant of the mobile phase, and γ is the surface

tension of the solvent [14, 15]. Since these latter two approaches are rarely used and offer no real advantages, we do not discuss the details.

A new chromatographic lipophilicity index φ_0 was recently proposed [16]. It is defined at the concentration of organic component in the mobile phase, required for log $k = 0$, and varies therefore from 0 to 100 %. No advantages in terms of lipophilicity scale are apparent.

5.2.2 Log k or log k_w

As seen in Fig. 2, the slope of the extrapolation of log k versus mobile phase composition φ may vary considerably (see also section 5.5). A proper expression of the differences in lipophilicity is found only at 100 % water composition. Therefore log k_w values should be used as RPLC measured lipophilicity parameter [7–9, 17]. However, this view is not shared by others [6, 14, 18, 19]. Yamagami and colleagues argue that log k_w values can predict log P values only for nonhydrogen-bonding or weak hydrogen-accepting substituents, but not for strong hydrogen-accepting substituents. However, such conclusion depends strongly on the type of stationary phase used. Particularly, interaction with nonprotected silanophilic groups may blur the picture (see below).

The intersection point of log k/φ lines is for closely related compounds often well defined, and in the range of $\varphi = 70$–100 % methanol. However, at present no full theoretical basis for such behavior can be offered. Discussions can be found in [11, 20].
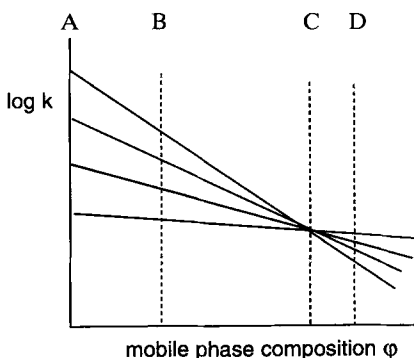
Figure 2. The choice of the RPLC lipophilicity index. Only log k_w values have a broad range in lipophilicity values (compare mobile phase composition A, representing log k_w values, and B). At certain mobile phase composition several compounds may have similar (at the isoelution point C) or even reversed lipophilicity values (for D) as compared with log k_w values.

5.3 Stationary Phases (Column Packings)

5.3.1 Overview

HPLC stationary phase materials have been reviewed and compared on several occasions, e.g., in [21, 22].

In order to obtain log k_W values closely resembling log P_{oct} values, it has been suggested to coat the solid phase with 1-octanol [3, 4, 23]. However, this appeared of limited practical use due to a rapid bleeding of the column. A more successful approach is used at Zeneca [24, 25]. 1-Octanol adsorbed onto a hyflo-supercel diatomaceous earth solid support is more stable. When such a system is calibrated with compounds of known log P_{oct} it can be used to obtain satisfactorily log P values for new compounds (J. J. Morris, personal communication).

Very popular and common are lipophilicity measurements made on octadecylsilane packings. However, it has been observed that these types of stationary phase packing are not ideal, since they contain a rather high proportion of free acidic silanol groups ($pK_a = 6.8 \pm 0.2$). This leads to difficulties with strong hydrogen-bonding compounds [25]. For this reason a masking agent should be added to the mobile phase. More recently alternatives have become available on the market, namely, polymeric packings and end-capped silica materials (see Table 1). These latter columns are treated with a secondary silanization using short alkyl groups or zwitterionic fragments. Nevertheless, even these new packings must be cafully evaluated, e.g., by solvatochromic analysis (see below).

Table 1. Selection of column packings used for lipophilicity measurements

Packing material	Abbreviation	Reference
Traditional phases (alkyl-bonded silica)		
Octadecylsilane	ODS (RP-18)	[7–9]
Octyl-silane	OS (RP-8)	[29]
Glyceryl-coated controlled-pore glass	Gly-CPG	[30]
Suppressed silanophilic interactions (polymer-coated silica and polymer based)		
Deactivated (end-capped) octyl-silane	DOS	[31]
Octadecyl polyvinylalcohol copolymer	ODP	[29, 31, 32]
Deltabond C_8	DB	[29]
Styrene-divinylbenzene copolymer	PS-DVB	[29, 33]
C_{18}-derivatized PS-DVB		[6]
End-capped RP-18	ABZ	[26–28]
Polyethylene	PE	[34]
Octadecyl-bonded alumina	ODA	[35]
Immobilized artificial membranes (IAM)		
Dipalmitoyl phosphatidylcholine or diacylphosphatidylcholine coated silica	IAM	[36–40]

The ODP column has the advantage that one can measure $\log k_w$ values directly, i.e., at 100 % water as mobile phase; furthermore, no masking agent is required. A further advantage of polymer-based columns is that they can be used in the full pH range 1–14. The disadvantages of the PLRP-S (poly(styrene-divinylbenzene)) phase are of physical nature, since it undergoes considerable shrinking and welling. C_{18}-derivatized PS-DVB columns can be used for alkane-water partition coefficient determinations [6]. End-capped RP-18 columns with maximally suppressed silanophilic interactions are of high interest for reliable lipophilicity measurements [26]. Recently, a Supelcosil ABZ-LC column was tested, which is end-capped by a small zwitterionic fragment [27, 28]. Log k_w values measured on this column give good correlations with $\log P_{oct}$ values (see below). It was also found that an ABZ column can be mimicked by adding a small amount of 1-octanol or LiCl to a normal ODS column [28]. Nevertheless, some presently unexplained artefacts may occur with an ABZ column. For instance, anions have a longer than expected retention on the ABZ packing [26].

5.3.2 New HPLC Packing Materials for Lipophilicity Measurements

Lipophilicity values are important parameters to decribe transport processes. Great interest exists in the pharmaceutical industry in the prediction of the transport across membranes such as the gastrointestinal and blood-brain barrier, and the skin. Immobilized artificial membranes (IAM) are solid-phase membrane mimetics, which are presently studied for this purpose [36–40]. It has been reported that $\log k_{IAM}$ values encode different information than $\log P_{oct}$ values [38].

5.3.3 Column Length

Very lipophilic compounds such as retinoids or lipid soluble vitamin derivatives will have long retention times on normal type columns (15–25 cm). In such cases shorter columns (2–10 cm) may be used. Correlations between lipophilicity scales and column length have been reported [23].

5.4 Mobile Phases

5.4.1 Selection of Organic Modifier

On most columns pure water cannot be used as mobile phase. Therefore, mainly binary mixtures of water with an organic modifier are used. Particularly small organic molecules such as methanol, acetonitrile and tetrahydrofuran are well suited, since they do not disturb the water "structure" too much. Since plots of $\log k$ versus mobile phase composition are often not linear using THF and MeCN, MeOH is recommended as standard organic modifier [7–9, 10]. Water molecules have a stronger hydrogen-donating ability than methanol, while they have a slightly stronger hydrogen-accepting ability than acetonitrile but slightly weaker than that of methanol [41].

A new approach is the use of micelles as organic component. Micelle/water partition coefficients (P_{mw}) measured with a HPLC system were recently proposed as a new hydrophobic index [42]. Possible advantages need to be evaluated.

5.4.2 Buffer and the Effect of Ionization

5.4.2.1 Buffer

In older publications phosphate buffers have been used. However, these have the disadvantage of forming ion-pairs to positively charged basic compounds. A good solution to this problem is zwitterionic buffers, such as morpholinopropane sulfonic acid (MOPS).

5.4.2.2 Ionization Correction

For monoprotic acids and bases the following equations can be used to correct for ionization:

acids: $\log k_w = \log k_w^{app} + \log (1 + 10^{pH-pK_a})$ (5)
bases: $\log k_w = \log k_w^{app} + \log (1 + 10^{pK_a-pH})$ (6)

where $\log k_w^{app}$ is the $\log k_w$ obtained at a particular pH. Similar equations can be used for the isocratic capacity factors. Equations for bifunctional acids and bases can be found (see [43] and Chapter 7).

The relationship between the capacity factor in the neutral form (k_n), fully ionized form (k_i), and in a partially ionized form (k) for a simple acid is given by the Horvath equation [44, 45]:

$$k = (k_n + k_i (K_a/[H^+]))/(1 + K_a/[H^+])$$ (7)

Further modifications led to [45]:

$$k = 0.5 (k_n - k_i) \tanh (pK_a - pH) + 0.5 (k_n + k_i)$$ (8)

We have developed an empirical equation also based on the tanh function, namely for monoprotic solutes [27]:

acids: $\log k_n = \log k + (1 - \tanh (pK_a - pH + 1))$ (9)
bases: $\log k_n = \log k + (1 - \tanh (pH - pK_a + 1))$ (10)

5.4.3 Masking Agents

Masking agents, such as *n*-decylamine or *N,N*-dimethyloctylamine in a concentration of 0.15 % (v/v), are often used to suppress interactions with free silanol groups on the solid phase. However, when the solute is an ionizable acid, ion pairs may be built with the masking agent. This results in higher than expected lipophilicities. It has also been proposed to add 0.25 % (v/v) 1-octanol to the mobile phase to mimic the octanol hydrogen-bonding activity [46]. Addition of small amounts of diethylamine may increase the measurable lipophilicity range up to the equivalent of $\log P_{oct} = 8$ [46].

5.4.4 Ion Pairs and Ion Pair Chromatography (IPC)

A major constraint of silica-bonded phases is the limited pH range of between 3 and 8. This may be troublesome for poorly retained hydrophilic strong acids or bases with pK_a values outside this range. For this kind of compound the retention may be modified by the addition of moderately hydrophobic and oppositely charged surfactants. This technique is known as ion-pair chromatography (IPC) [47–49]. One of its shortcomings is poor reproducibility of retention. A further alternative for charged compounds is micellar liquid chromatochrography (MLC), using longer chain surfactants, which form micelles. A disadvantage here are the rather broad peaks.

5.5 Retention Mechanism

5.5.1 Solvatochromic Analysis

An extensive review of the mechanism of the retention process in RPLC has been published in 1993 in a special issue of the *Journal of Chromatography* [50].

Quantitative structure-retention relationships have given some insight in the most important properties in retention [51]. For instance, in the case of polyaromatic hydrocarbons a combination of Van der Waals volume and kinetic energy from molecular dynamics studies can be used to predict log k values [51]. Valuable insight in the descriptors relevant for retention can be obtained by socalled solvatochromic analysis. This is based on the linear solvation-energy relationships (LSER) methodology developed by Taft, Kamlet and others. This approach has been applied to chromatographic retention, and can be written as [52, 53]:

$$\log k = \log k_o + m(\delta_m^2 - \delta_s^2) V_2/100 + s(\pi_m^* - \pi_s^*)\pi_2^* \\ + a(\beta_m - \beta_s)\alpha_2 + b(\alpha_m - \alpha_s)\beta_2 \tag{11}$$

where log k_o is an independent term, m, s, a and b are the coefficients of the regression, $V_2/100$ is the scaled volume of the solute (to bring them to the same order of magnitude as the hydrogen-bonding descriptors α and β), δ is the Hildebrand solubility parameter, and π^*, α and β are the Kamlet-Taft solvatochromic polarity (π^*), hydrogen bond acidity (α) and hydrogen bond basicity (β) parameters (see Chapter 18). The subscripts s and m denotes the stationary and mobile phases, respectively, while subscripts 1 and 2 are the solvent and solute, respectively.

With a fixed mobile phase and a fixed stationary phase, Eq. (11) simplifies to:

$$\log k = \log k_o + m_1 V_2/100 + s_1\pi_2^* + a_1\alpha_2 + b_1\beta_2 \tag{12}$$

For the same solute in the same column, but with different mobile phase compositions, Eq. (11) can be rewritten as:

$$\log k = \log k_s + m_2\delta_m^2 + s_2\pi_m^* + a_2\beta_m + b_2\alpha_m \tag{13}$$

Assuming that there is a linear relationship between δ^2 and π^*, this equation can be further simplified to:

$$\log k = \log k_s + s_2\pi_{m^*} + a_2\beta_m + b_2\,\alpha_m \tag{14}$$

Solvatochromic analysis is a suitable procedure to compare stationary phases to each other [26, 29, 31, 54–51). For example, from these analyses it was inferred that the stationary phase of ODS is considerably more polar than bulk alkane solutions [57]. In some studies, it was concluded that the hydrogen-bonding parameter α is of greater importance for retention behaviour then β [29], while in others the reverse was found [52, 53]. This prompted several groups to explore to possibility to measure hydrogen bond basicity by RPLC [59].

5.5.2 Slope Analysis and Hydrogen-Bonding Capacity

Basically, the chromatographic retention mechanism consists of two components, namely: the size of the solute (reflected by its volume or surface area) and the hydrogen-bonding capacity. It has been observed that plotting of the slope S (see Eq. (16)) against log k_w unravels differences in hydrogen bonding within a series of compounds (see Fig. 3) [59]. The slope parameter S has been correlated with solvatochromic parameters of the solutes [60].

Another means of defining a slope descriptor is by plotting log k versus organic modifier concentration D [41]:

$$\log k = q - r \log D = r \log (1/D) + q \tag{15}$$

These r values are to a large extent correlated with the nonpolar surface area of the solute.

Two sets of compounds have recently been compared, namely a set of simple organic compounds and a set of drugs, on four different columns (ODS, ODP, ABZ, IAM) of different lengths [27] (see Table 2). It was observed that the slope a of the correlation

$$S = a \log k_w + b \tag{16}$$

is nearly identical on each column and virtually 1.0. For the simple compounds we confirmed the observation made by others (e.g., [59]) that a separation in H-donors and H-acceptors is observed, while this is not the case for more complicated molecules. Here, the donating and accepting properties seem to cancel each other. Furthermore,
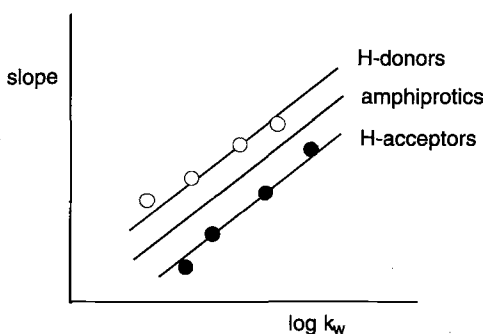
Figure 3. Example of a slope S versus log k_w plot for fictive data. Open circles are H-bond donors, and closed circles are acceptors. This generalized plot can be deduced from various experimental data sets.

Table 2. Comparison of four stationary phases for simple compounds (set 1) and more complex structures (mainly drugs, set 2)

$\log P_{oct} = a \log k_w + b$	a	b	r
Set 1 ($n = 19$)			
ABZ	1.01(± 0.08)	$-0.43(\pm 0.25)$	0.947
ODP	0.98(± 0.11)	n.s.	0.904
ODS	0.90(± 0.10)	n.s.	0.906
IAM	1.14(± 0.20)	$+0.74(\pm 0.38)$	0.804
Set 2 ($n = 16$)			
ABZ	0.84(± 0.10)	n.s.	0.915
ODP	0.69(± 0.10)	n.s.	0.873
ODS	0.75(± 0.19)	n.s.	0.733
IAM	0.88(± 0.30)	$+1.43(\pm 0.46)$	0.617

Values in parentheses are the standard deviations of the regression coefficients [27].
n.s. indicates a non-significant constant term.

these are difficult to estimate from molecular structure, since conformational effects are difficult to include reliably. Extending this study even further, it was found that 233 compounds taken [7–9, 12, 17, 46, 59] measured on different column types and lengths can be described by a single equation:

$$S = 1.04(\pm 0.06) \log k_w + 0.99(\pm 0.01)$$

$$n = 233, r = 0.984, s = 0.580, F = 6974 \tag{17}$$

5.5.3 Effect of Intercharge Distance in Zwitterions

Zwitterions are a special case in chromatography. We and others have found that the retention behaviour of zwitterions is related to the distance between the negative and positive charge [27]. Compounds with little charge separation, such as α-amino acids, have a higher lipophilicity than isomers with larger separation.

5.5.4 Effect of Conformation on Retention

Larger flexible molecules may fold on their way through the HPLC column, and thus display an unpredictable lipophilicity behavior. For example, a conformational effect can play a role when a series of homologs with varying chain length is studied. An interesting example of this conformation-dependent behavior has recently been found in a series a benzylamines [27]. Log k_w values increase from methyl to butyl, but then drop again for the pentyl homolog.

Significant conformational effects must be expected for macromolecules, such as peptides and nucleic acids. This folding process may be slow, giving rise to peak broadening.

5.5.5 Lipophilicity of Peptides and Proteins

RPLC has been used traditionally for the separation and purification of peptides and proteins. It can therefore also be employed to measure the lipophilicity of such macromolecules [61]. However, conformational effects are important and cannot easily be foreseen. This can be illustrated by considering the retention times of a peptide in which the position of only one amino acid is varied [62]. A comparison of retention behavior in reversed-phase chromatography (RPC) and hydrophobic interaction chromatography (HIC) revealed significant differences for proteins [63]. The driving force in RPC for proteins is mainly enthalpy-driven, while for HIC this is an entropic process.

5.6 Correlations of log k_w Values with log P_{oct} and Other log P Scales

Although log k or preferably log k_w values are good lipophilicity indices *per se*, very often a direct correlation or even conversion into log P_{oct} values is performed. The mean reason is that one would like to have a comparison with standard experimental lipophilicity values or to calculated log P values, which also refer to the 1-octanol/water system.
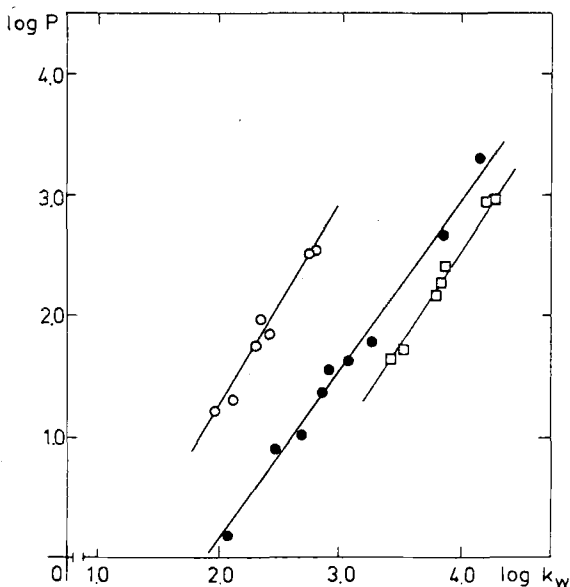
Figure 4. Typical log P_{oct} versus log k_w plot. Relationships are found for closely related compounds, namely phenylureas (●), phenoxycarboxylic acids (○) and phenoxycarboxylic acid methyl esters (□) (Reprinted with permission from [7], copyright 1983, Elsevier.)
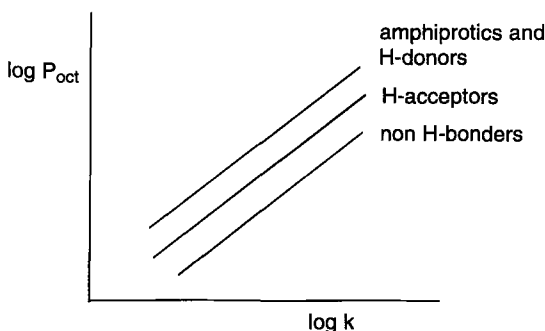
Figure 5. Generalized plot of log P_{oct} against log k_w: differences in lipophilicity scales reflect hydrogen-bonding capability in both systems.

Despite the fact that log P values are often claimed to correlate well with log k or log k_w values, this is only so for very closely related compounds (see Fig. 4). And even within an apparently structurally similar series of compounds, subtle effects such as differences in hydrogen-bonding capacity can be found (see Fig. 5). In a series of investigations Yamagami and coworkers formulated the following equations [18, 19]:

$$\log k = a \log P + h_{co}HB_{co} + h_A HB_A + h_{AM}HB_{AM} + \varrho\sigma_i + c \tag{18}$$

$$\log k_w = a \log P + \Sigma b_i HB_i + C \tag{19}$$

In the expressions the three *HB* terms are indicator variables expressing hydrogen bonding effects of substituents, defined as follows: HB_{co} = 1 for esters and amides, HB_A = 1 for H-acceptors other than COOR and CON, and HB_{AM} = 1 for H-donor substituents. This equation illustrates well that the differences between lipophilicity scales can be expressed in terms of hydrogen-bonding capacity (Eq. (19)). Nevertheless, a method based on fragment indicator values is not general enough.

A limitation of HPLC lipophilicity measurements is caused by the inhomogenous nature of the stationary phase resulting in specific interactions with certain solutes. Thus, it is expected that certain column types reflect more the partioning behavior in octanol/water than others. In a comparison of four different columns (see sections 5.5.2) it was found that the ABZ column gives the best correlations with log P_{oct}. However, for more complex structures, such as most biologically active compounds, the correlation becomes less significant (see Table 2). This limited data set is illustrative for what is often seen in practice: log k_w values for small closely related compounds correlate well with log P_{oct} values; for more structurally diverse drugs the correlation is less to nonexistent.

It has recently been found that the differences between log P values measured in two different systems gives relevant information on hydrogen-bonding capacity of the molecules, which is related to their transport properties through membranes (see Chapter 14). It would be of great interest to find similarly a $\Delta\log k$ or $\Delta\log k_w$ parameter derived from two different columns. Generally log k_w values on ODS colums corre-

asonably well with octanol/water, but poorly with alkane/water partition coefficients. Interestingly, the reverse seems to be true for C18-derivatized PS-DVB (Act-I) columns [6]. However, this observation needs further verification.

5.7 Recommendations

5.7.1 OECD/EU Guidelines

The partition coefficient P in 1-octanol/water has to be measured in all EU member countries before a chemical is marketed for the first time and is one of the chemical properties laid down by the OECD as part of the MPD (Minimum Pre-marketing set of Data) [64]. To eliminate disadvantages of the shake-flask method to estimated log P values, chromatographic methods can be used. The method is based on the putative correlation between log P and log k_w values and uses reference compounds. However, it is known now, as discussed in the present chapter, that this is not always true. For new classes of compounds there are no comparable reference structures. Therefore, this method has its limits.

5.7.2 Recommended Method

At present the best method to measure (octanol/water-like) lipophilicity values by RPLC consists of:

- use of polymer-based (ODP) or end-capped (ABZ) stationary phase;
- methanol/water as mobile phase;
- zwitterionic buffer (MOPS);
- extrapolated polycratic (log k_w) capacity factors as lipophilicity index.

References

[1] Haggerty, W. J., and Murrill, E. A., *Res. Dev.* **25,** 30 (1974)
[2] Henry, D., Block, J. H., Anderson, J. L., and Carlson, G. R., *J. Med. Chem.* **19,** 619–626 (1976)
[3] Mirrlees, M. S., Moulton, S. J., Murphy, C. T., and Taylor, P. J., *J. Med. Chem.* **19,** 615–619 (1976)

[4] Unger, S. H., Cook, J. R., and Hollenberg, J. S, *J. Pharm. Sci.* **67**, 1364–1366 (1978)

[5] Kaliszan, R., *J. Chromatogr.* **220**, 71–83 (1981)

[6] Lambert, W. J., *J. Chromatogr. A* **656**, 469–484 (1993)

[7] Braumann, Th., Weber, G., and Grimme, L. H., *J. Chromatogr.* **261**, 329–343 (1983)

[8] El Tayar, N., Van de Waterbeemd, H., and Testa, B., *J. Chromatogr.* **320**, 293–304 (1985)

[9] El Tayar, N., Van de Waterbeemd, H., and Testa, B., *J. Chromatogr.* **320**, 305–312 (1985)

[10] Reymond, D., Chung, G. N., Mayer, J. M., and Testa, B., *J. Chromatogr.* **391**, 97–109 (1987)

[11] Valko, K., Snyder, L. R., and Glajch, J. L., *J. Chromatogr. A* **656**, 501–520 (1993)

[12] Schoenmakers, P. J., Billiet, H. A. H., and De Galan, L., *J. Chromatogr.* **185**, 179–195 (1979)

[13] Hsieh, M.-M., and Dorsey, J. G., *J. Chromatogr.* **631**, 63–78 (1993)

[14] Horvath, C., Melander, W., and Molnar, I., *J. Chromatogr.* **125**, 129–156 (1976)

[15] Wells, M. J. M., and Clark, R., *J. Chromatogr.* **284**, 319–355 (1984)

[16] Valko, K., and Slégel, P., *J. Chromatogr.* **631**, 49–61 (1993)

[17] El Tayar, N., Van de Waterbeemd, H., and Testa, B., *Quant. Struct.-Act. Relat.* **4**, 69–77 (1985)

[18] Yamagami, C., Yokota, M., and Takao, N., *J. Chroamtogr. A* **662**, 49–60 (1994)

[19] Yamagami, C., Yokota, M. and Takao, N., *Chem. Pharm. Bull.* **42**, 907–912 (1994)

[20] Tan, L. C., and Carr, P. W., *J. Chromatogr. A* **656**, 521–535 (1993)

[21] Kalizan, R., *Quant. Struct.-Act. Relat.* **9**, 83–87 (1990)

[22] Rippel, G., Alattyani, E., and Szepésy, L., *J. Chromatogr. A* **668**, 301–312 (1994)

[23] Miyake, K., and Terade, H., *J. Chromatogr.* **157**, 386–390 (1978)

[24] Taylor, P. J. Hydrophobic properties of drugs. In: *Comprehensive Medicinal Chemistry*, Vol. **4**. Hansch, C., Sammes, P. G. and Taylor, J. B. (Eds.). Pergamon Press: Oxford; 241–294 (1990)

[25] El Tayar, N., Tsantili-Kakoulidou, Röthlisberger, T., Testa, B., and Gal, J., *J. Chromatogr.* **439**, 237–244 (1988)

[26] Pagliara, A., Khamis, E., Trinh, A., Carrupt, P.-A., Tsai, R.-S. and Testa, B., *J. Liq. Chromatogr.* **18**, 1721–1745 (1995)

[27] Kansy, M., Fischer, H. and Van de Waterbeemd, H., Correlations between log *P* and RP-HPLC log k_w values: effect of charge and conformation. In: *Trends in QSAR and Molecular Modeling 94*. Sanz, F. (Ed.). Prouse: Barcelona 1995

[28] Van de Waterbeemd, H., Karajiannis, H., Kansy, M., Obrecht, D., Müller, K., and Lehmann, Chr., Conformation-lipophilicity relationships of peptides and peptide mimetics. In: *Trends in QSAR and Molecular Modeling 94*. Sanz, F. (Ed.). Prouse: Barcelona 1995

[29] Altomare, C., Cellamera, S., Carotti, A., and Ferappi, M., *Quant. Struct.-Act. Relat.* **12**, 261–268 (1993)

[30] Miyake, K., and Terada, H., *J. Chromatogr.* **240**, 9–20 (1982)

[31] Vallat, P., Fan, W., El Tayar, N., Carrupt, P.-A., and Testa, B., *J. Liq. Chromatogr.* **15**, 2133–2151 (1992)

[32] Altomare, C., Tsai, R.-S., El Tayar, N., Testa, B., Carotti, A., Cellamare, S., and De Benedetti, P. G., *J. Pharm. Pharmacol.* **43**, 191–197 (1991)

[33] Miyake, K., Kitaura, F., Mizuno, N., and Terada, H., *Chem. Pharm. Bull.* **35**, 377–388 (1987)

[34] Rothemund, S., Krause, E., Ehrlich, A., Bienert, M., Glusa, E. and Verhallen, P., *J. Chromatogr. A.* **661**, 77–82 (1994)

[35] Haky, J. E. and Vemulapalli, S., *J. Liq. Chromatogr.* **13**, 3111–3131 (1990)

[36] Miyake, K., Kitaura, F., Mizuno, N. and Terada, H., *J. Chromatogr.* **389**, 47–56 (1987)

[37] Ong, S., Cai, S.-J., Bernal, C., Rhee, D., Qui, X., and Pidgeon, C., *Anal. Chem.* **66**, 782–792 (1994)

[38] Kaliszan, R., Nasal, A., and Bucinski, A., *Eur. J. Med. Chem.* **29**, 163–170 (1994)

[39] Pidgeon, C., Ong, S., Liu, H., Qui, X., Pidgeon, M., Dantzig, A. H., Munroe, J., Hornback, W. J., Kasher, J. S., Glunz, L., and Szczerba, T., *J. Med. Chem.* **38**, 590–594 (1995)

[40] Ong, S., Liu, H., Qiu, X., Bhat, G., and Pidgeon, C., *Anal. Chem.* **67**, 755–762 (1995)

[41] Kaibara, A., Hohda, C., Hirata, N., Hirose, M., and Nakagawa, T., *Chromatographia* **29**, 275–288 (1990)

[42] Tanaka, A., Nakamura, K., Nakanishi, I., and Fujiwara, H., *J. Med. Chem.* **37**, 4563–4566 (1994)

[43] Van de Waterbeemd, H., and Testa, B., *Adv. Drug Res.* **16**, 85–225 (1987)

[44] Horvath, C., Melander, W. and Molnar, I., *J. Anal. Chem.* **49**, 142–154 (1977)

[45] Hanai, T., *J. Chromatogr.* **550**, 313–324 (1991)

[46] Belsner, K., Pfeifer, M., and Wilfert, B., *J. Chromatogr.* **629**, 123–134 (1993)

[47] Rodgers, A. H., and Khaledi, M. G., *Anal. Chem.* **66**, 327–334 (1994)

[48] Zou, H., Zhang, Y., Hong, M., and Lu, P., *J. Chromatogr.* **625**, 169–175 (1992)

[49] Martorell, C., Calpena, A. C., Escribano, E., and Poblet, J. M., *J. Chromatogr. A.* **655**, 177–184 (1993)

[50] Carr, P. W., Martire, D. E., and Snyder, L. R., Special issue on the retention process in reversed-phase liquid chromatography, *J. Chromatogr.* **656A**, 1–616 (1993)

[51] Hanai, T., *J. Liq. Chromatogr.* **16**, 1453–1462 (1993)

[52] Carr, P., *Microchem. Rev.* **48**, 4–28 (1993)

[53] Cheong, W. J., and Carr, P. W., *Anal. Chem.* **61**, 1524–1529 (1989)

[54] Rosés, M., and Bosch, E., *Anal. Chim. Acta* **274**, 147–162

[55] Hellai, F., Phan-Tan-Luu, R., and Siouffi, A. M., *J. Liq. Chromatogr.* **17**, 2845–2869 (1994)

[56] Park, J. H., Chae, J. J., Nah, T. H., and Jang, M. D., *J. Chromatogr. A* **664**, 149–158 (1994)

[57] Helburn, R. S., Rufan, S. C., Pompano, J., Mitchem, D., and Patterson, W. T., *Anal. Chem.* **66**, 610–618 (1994)

[58] Tan, L. C., Carr, P. W., Fréchet, J. M. J., and Smigol, V., *Anal. Chem.* **66**, 450–457 (1994)

[59] Minick, D. J., Brent, D. A., and Frenz, J., *J. Chromatogr.* **461**, 177–191 (1989)

[60] Chen, N., Zhang, Y., and Lu, P., *J. Chromatogr.* **633**, 31–41 (1993)

[61] El Tayar, N., Karajiannis, H., and Van de Waterbeemd, H., *Amino Acids* **8**, 125–139 (1995)

[62] Büttner, K., Ostresh, J. M., and Houghten, R. A., Induced conformational effects during RPHPLC: Prediction of immunodominant T_H-cell antigenic sites. In: *Peptides. Chemistry, Structure and Biology.* Rivier, J. E., and Marshall, G. R. (Eds.). Escom: Leiden; 423–425 (1990)

[63] Lee, D. W., and Cho, B. Y., *J. Liq. Chromatogr.* **17**, 2541–2558 (1994)

[64] Harnisch, M., Möckel, H. J., and Schulze, G., *J. Chromatogr.* **282**, 315–332 (1983)

6 Centrifugal Partition Chromatography for Lipophilicity Measurements

Ruey-Shiuan Tsai, Giuseppe Lisa, Pierre-Alain Carrupt and Bernard Testa

Abbreviations

CHF	Chloroform
CPC	Centrifugal partition chromatography
DBE	Di-*n*-butyl ether
DCCC	Droplet counter-current chromatography
HDES	Hydrodynamic equilibrium systems
HSES	Hydrostatic equilibrium systems
RP-HPLC	Reversed-phase high-performance liquid chromatography
SF	Shake-flask

Symbols

D^I	Distribution coefficient measured at isoelectric point
k'	Capacity factor
P_{alk}	Partition coefficient in alkane/water systems
P_{oct}	Partition coefficient in *n*-octanol/water systems
t_R	Retention time of the solute
t_0	Retention time of nonretained solutes
U	Flow rate of the mobile phase
V_M	Mobile phase volume
V_R	Retention volume of the solute
$\Delta \log P_{oct\text{-}alk}$	$\log P_{oct}$ minus $\log P_{alk}$

6.1 Introduction: a Need for an Accurate Method for Partition Coefficient Measurements

The role of drug lipophilicity in pharmacokinetics and pharmacodynamics is both primordial and ubiquitous, as well documented in the literature [1–3]. While partition coefficients measured in *n*-octanol/water systems (P_{oct}) have encountered enormous success in the studies of drug lipophilicity and structure-activity relationships, new structural parameters arising from a combination of partition coefficients in different solvent systems such as $\Delta \log P_{oct-alk}$ ($= \log P_{oct} - \log P_{alkane}$) are emerging as structural determinants particularly in blood-brain barrier permeation [4], gastrointestinal absorption [5] and skin penetration [6] of drugs (see Chapter 14). At this point, an essential condition is to determine precisely and accurately partition coefficients in each solvent system to minimize the errors in, for example, $\Delta \log P_{oct-alk}$.

Thermodynamically, the partition coefficient is defined as a constant relating the activity of a solute in two immiscible phases at equilibrium. A number of experimental models are currently used to simulate partition processes in biological systems and to determine lipophilicity. The "shake-flask" (SF) method using water and a poorly miscible organic solvent is a technique most widely used for measuring partition coefficients [7]. However, the SF method suffers from a number of limitations such as the precision of phase volume ratio, the (im-)purity of the solvent used, the (im-)purity, volatility and adsorption of solutes, and finally the formation of microemulsions induced by vigorous mechanical agitation as previously discussed by Dearden and Bresnen [8] (see also Fig. 1). Clearly, a more accurate experimental method to determine partition coefficients is called for.

Partition, to be well distinguished from adsorption, chromatography has been explored as an alternative means for measuring lipophilicity. In particular, chromatographic retention parameters obtained by reversed-phase high-performance liquid chromatography (RP-HPLC) [9] (see Chapter 5) have become increasingly popular in replacing the octanol/water partition coefficient measured by the SF method. However, the mechanisms of retention in RP-HPLC are not truly identical, while being similar, to those of partitioning in *n*-octanol/water systems owing to the restricted mobility of the bonded alkyl chains and the presence of a solid support with a non-negligible proportion of residual silanol groups. Even for a polymeric support grafted with alkyl chains such as octadecylpolyvinyl alcohol, the solute retention behavior in some cases can be very different from the partitioning in octanol/water [T. Ter Laak, unpublished results]. While the RP-HPLC method remains useful, it is difficult – if not unlikely – to derive from it quantitative structural information such as hydrogen-bonding capacity.

In recent years, centrifugal counter-current chromatography, also known as centrifugal partition chromatography (CPC) [10], has been explored as a novel technique for
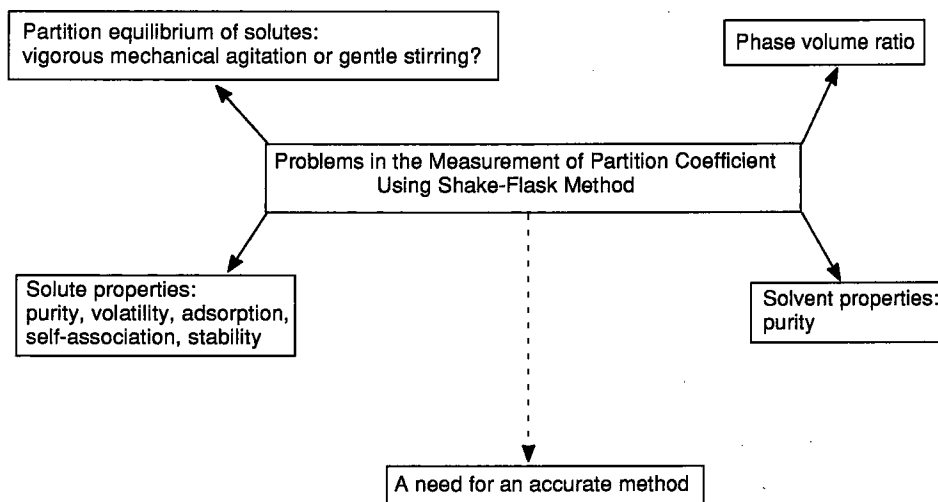
Figure 1.   Problems inherent in partition coefficient measurements using the shake-flask method.

measuring liquid–liquid partition coefficients. This is a unique form of liquid–liquid partition chromatography that is free of a solid support. Thus, the problem of adsorption is avoided and solute retention depends solely on its partition coefficient. Two poorly miscible liquids are used as the stationary and mobile phases in the chromatograph. Centrifugal forces maintain the stationary phase, while the mobile phase is pumped through the system. During the past decade, various types of CPC systems such as the flow-through multilayer coil planet centrifuge, the horizontal flow-through multilayer coil planet centrifuge, the toroidal coil planet centrifuge, and the multichannel cartridges CPC have been used to determine partition coefficients [10].

    In this chapter, we review the development and current state-of-the-art of this novel technique in lipophilicity measurements and its application in determining solute structural properties.

6.2 Historical Aspects

The first description of liquid–liquid partition chromatography dates back to the work of Martin and Synge who worked on the extraction of natural products using counter-current extraction in 1941 [11]. While the early forms of liquid–liquid partition chromatography did not appear very promising for practical purposes, Martin and Synge received the highest awards for their pioneering concept. More than 20 years later, the emergence and extensive commercialization of high-performance liquid chromatography (HPLC) have become so popular that HPLC techniques are a must for every modern analytical chemical laboratory, while liquid–liquid partition chromatography remains relatively obscure.

6.2.1 The Discovery and Development of CPC

The invention of centrifugal liquid–liquid partition chromatography is not as new as it appears. In 1966, the year in which the first description of HPLC by Horváth and Lipsky [12] was published, Ito at the National Institutes of Health, Maryland, discovered this novel form of liquid–liquid partition chromatography for the separation of cell particles [13]. He subsequently developed various types of centrifugal partition chromatography with his rich imagination, a fine sense of fluid dynamics, and a good talent in engineering. This ingenious invention was recommended for the purpose of preparative separation with much shorter separation times as compared with the previous form of liquid–liquid partition chromatography and to avoid the problems of solid-phase adsorption encountered in HPLC. As a shortcoming of CPC, its theoretical plate number is usually in the range of hundreds and much inferior to that of HPLC. The first commercialization of the prototype CPC did not begin until the 1980s, which is at least one of the factors accounting for its limited use given its attractive merits in separation technologies. To date, several commercialized chromatographs are still in a state of prototype and their many mechanical designs are to be improved.

6.2.2 From log $P_{octanol-hexane/water}$ to log P_{oct} using Multichannel Cartridge-type CPC

In view of the success of the log P_{oct} parameter in drug research, attempts to use CPC for lipophilicity measurements are understandable. Terada and coworkers [14] were the first to determine lipophilicity using the CPC technique in 1987. They used the multichannel cartridge-type CPC (model CPC-LLN, Sanki Engineering, Nagaokakyo, Japan). Perhaps due to the viscous nature of octanol and the high pressure generated in the chromatograph, no direct measurements using octanol/water systems were reported. Instead, a mixture of octanol : hexane (20 % : 80 %) was used as the mobile phase, and the calculated partition coefficients were correlated with log P_{oct} values from the literature. A log P range of –0.5 to –2 was reported. The following publication in 1988 was from Armstrong and coworkers [15], in which both octanol : hexane (40 % : 60 %) and octanol (100 %) were used as the organic phase by employing a similar apparatus (model CPC-NMF, Sanki Laboratories Inc., Sharon Hill, Pennsylvania, USA). For the measurement of compounds with log $P > 0$, they recommended use of water as the mobile phase and reported measured log P values from 0.2 to 2.3. Two points are worth mentioning in this study; first, the continuous loss of the stationary phase (i. e., "bleeding") requires the frequent control of dead time; second, the retention time of lipophilic solutes in unacceptably long (8–11 hours for compounds of log $P \sim 2$). A similar study from our laboratory [16] using the Sanki model CPC-LLN also showed that it is impossible to use octanol as the mobile phase, the pressure being too high in the chromatograph. It was thus concluded that the multichannel cartridge-type CPC is not appropriate for lipophilicity measurements.

6.2.3 From log P_{oct} to log P (solvent "quartet") using Coil Planet-type CPC

The endeavor to use CPC for lipophilicity measurements was not stopped by the difficulties encountered in the multichannel cartridge-type CPC. Our laboratory changed to the coil planet-type CPC as an alternative for this purpose. Log P_{oct} values from –1.3 to 1.3 were obtained using octanol as the mobile phase with the Ito Multilayer Coil Separator-Extractor (P. C. Inc., Potomac, Maryland, USA) [17]. Soon after that, we have further extended the log P_{oct} range from –3 to 3 using "rapid mode" operation procedures [18] (see section 6.4.3 for details).

Since partition coefficients in various solvent systems encode different structural information [19], it would be desirable to determine partition coefficients in the four model solvent systems: amphiprotic (*n*-octanol/water), protic (chloroform/water), aprotic (di-*n*-butyl ether/water) and inert (alkane/water) systems. The progress made in our laboratory in the past few years has enabled the measurements to be carried out in *n*-octanol/water, di-*n*-butyl ether/water and *n*-dodecane/water and allowed important structural properties such as hydrogen-bonding capacity to be determined from the results.

6.3 Mechanisms of Solute Partitioning in Various Types of CPC

Centrifugal counter-current chromatography is a liquid–liquid chromatographic technique resembling to some extent droplet counter-current chromatography (DCCC) (Fig. 2) [20]. In DCCC, the stationary phase is retained in a large number of vertical narrow-bore tubes, while the mobile phase, depending upon its density, is pumped through the system in an ascending or descending mode in the form of small droplets. CPC differs from DCCC in that centrifugal and/or Archimedean screw forces maintain the stationary phase, while the mobile phase is pumped through [21]. These features allow high partition efficiency and large retention capacity of the stationary phase under a high flow-rate of mobile phase.

In a recent review, Ito presented the historical background, development and mechanisms of distribution of the stationary and mobile phases in the CPC coil [21], classifying the mechanisms of CPC systems into two forms, namely hydrostatic equilibrium systems (HSES) and hydrodynamic equilibrium systems (HDES).
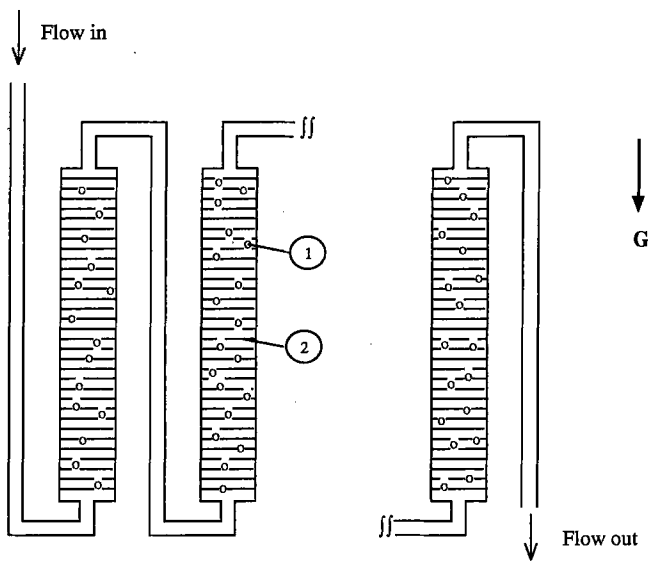
Figure 2. Schematic illustration of the principle of DCCC in the ascending mode (lighter phase as mobile phase). 1, mobile phase; 2, stationary phase.

6.3.1 Hydrostatic Equilibrium Systems

The hydrostatic equilibrium systems use stationary coils, for example PTFE tubing such as those used in the toroidal coil CPC (Fig. 3), or multichannel cartridges such as those used in the Sanki CPC (Fig. 4) [22]. The coils are so-called stationary with respect to the rotating frame towards the central axis of the centrifuge. Measurements begin with filling the coil with the stationary phase. The mobile phase is then pumped into the rotating centrifuge and starts to percolate through the stationary phase segments on one side of the coil, displacing nearly half the volume of stationary phase
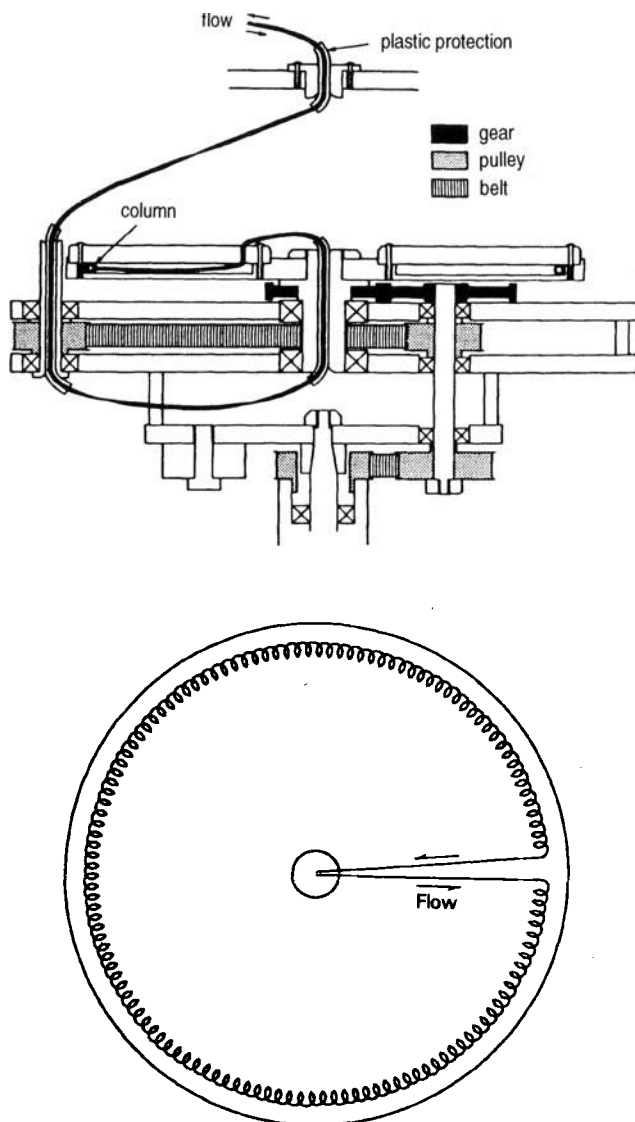
Figure 3. Schematic illustration of toroidal coil planet centrifuge.

in the coil. Hence, solutes introduced at the inlet of the coil are subjected to a continuous partitioning process between the two phases. In this system, the retention of the stationary phase and the distribution of stationary and mobile phases in the coil are governed mainly by the centrifugal force. As it appears, the pressure in the system is relatively high and the leakage of phase solutions becomes one of the major problems for the measurement. Moreover, for the measurement of very lipophilic ($\log P > 2$) or very hydrophilic compounds ($\log P < -2$), these systems appear impractical due to the low retention volume of the stationary phase (c. 50 %).
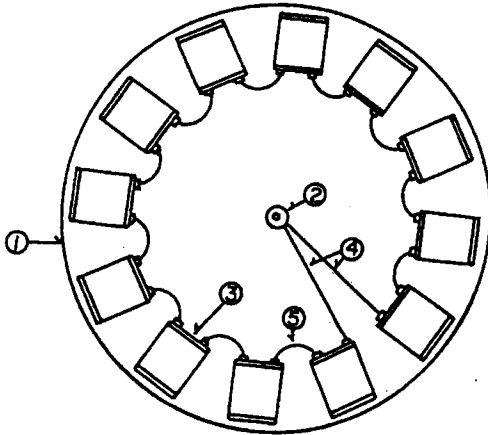
Figure 4.   Centrifuge of the multichannel cartridge CPC. 1, rotor of the centrifuge; 2, rotary seal joint; 3, column cartridge; 4, tubing connecting rotary seal joint and column cartridges; 5, tubing connecting cartridges.

6.3.2 Hydrodynamic Equilibrium Systems

In the hydrodynamic equilibrium systems such as the flow-through multilayer coil CPC (Fig. 5) or the horizontal flow-through multilayer CPC, the rotation of the coiled column around its own axis creates an Archimedean screw force which, in combination with a revolutionary centrifugal force towards the center of the centrifuge, allows a continuous mixing of the two phases while retaining a high proportion of the stationary
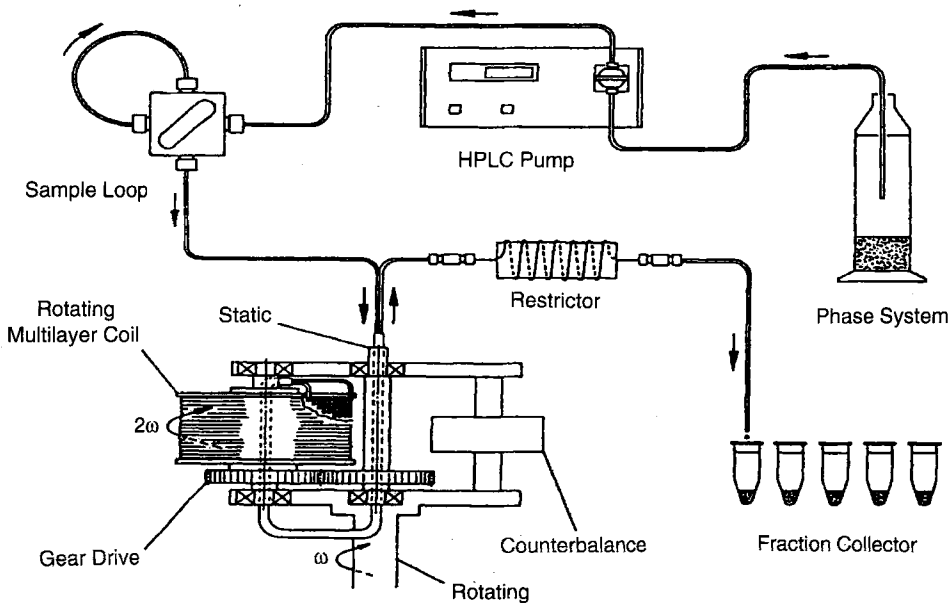
Figure 5.   Schematic illustration of the Ito multilayer coil separator–extractor.

phase. Using a stroboscope, Ito observed that under high centrifugal forces and flow rates the mixing zone, being located near the center of the centrifuge where the centrifugal force is the weakest, travels towards one end of the coil [23]. This indicates that the two phases are subjected to a partitioning process of repetitive mixing and settling at a rate of 13–16 times per second under a revolutional speed of 800–1000 r.p.m. when the mobile phase is steadily passing through the stationary phase. All these features assure a high partition efficiency and a high retention capability of the stationary phase under a high flow rate of the mobile phase in comparison with the hydrostatic equilibrium systems.

6.4 Method Development for log *P* Measurements Using CPC

6.4.1 Calculation of Partition Coefficients

Like in the other chromatographic techniques, the capacity factor (k') determined in CPC is related to the retention volume of the solute (V_R) and the mobile phase volume (V_M) by the equation:

$$k' = \frac{V_R - V_M}{V_M} \tag{1}$$

When using the aqueous solution as eluent and knowing the volume ratio of the mobile and stationary phase, the partition coefficient is calculated as:

$$P = \frac{V_R - V_M}{V_M} \times \frac{V_M}{V_t - V_M} \tag{2}$$

where V_t is the total volume of the column. Eq. (2) can be rearranged to give Eq. (3):

$$\log P = \log \frac{(t_R - t_0) \times U}{V_t - U \times t_0} \tag{3}$$

where t_R is the retention time of the solute, t_0 the dead time or retention time of non-retained solutes and U the flow rate of the mobile phase. Similarly, when using the organic phase as the mobile phase, partition coefficients are then calculated as:

$$\log P = \log \frac{V_t - U \times t_0}{(t_R - t_0) \times U} \tag{4}$$

6.4.2 Considerations about the Equipment (Mainly the Centrifuge)

Various prototypes of counter-current chromatographs have been ingeniously designed and developed by Ito. However, only four types are used to measure partition coefficients due, at least in part, to their commercial availability. These are multichannel cartridges CPC (Sanki Engineering, Kyoto, Japan), toroidal coil planet CPC, flow-through multilayer coil planet CPC (also called Ito multilayer separator-extractor, P. C.

Inc., Kim Place, Potomac, Maryland, USA) and horizontal flow-through multilayer CPC (model CCC-1000, Pharma-Tech Research, Baltimore, Maryland, USA). In contrast to the coil type column where the tubing diameter and length and the total volume can be varied, the cartridges in multichannel CPC apparatus are not readily modifiable.

For hydrodynamic reasons such as pressure build-up in the column [17], our laboratory has given the preference to the CPC systems employing hydrodynamic equilibrium mechanisms. This however does not imply that other types of CPC systems cannot be used for lipophilicity measurements as evidenced in the next section. The Ito multilayer coil separator-extractor is a simple centrifuge with a single coiled column balanced by counter-weights. This suggests the need for a readjustment of counter weights whenever the volume ratio of organic and aqueous phases is changed, since a fine balance is of particular importance in stabilizing the system. A defect in balancing may lead to a continuous "bleeding" of the stationary phase and perturbs the measurement. In addition, when filling a vertical coiled column with volatile organic solvents such as chloroform, air bubbles are likely to accumulate in the tubing and will equally perturb the measurement. In contrast, the horizontal flow-through multilayer CPC uses three coiled columns to avoid the problems of imbalance when changing the volume ratio. The problem of air bubble accumulation in the column is also easier to overcome when using horizontal coiled columns. However, this type of CPC has its own limitations: the increasing number of intercolumn connections would increase the frequency of tubing disconnection and the leakage of phase solutions. Fortunately, the design of intercolumn connections was much improved during the past few years, and tubing disconnection is no longer a major practical problem.

It is immediately apparent that the diameter and length of tubing are important parameters in influencing the partitioning equilibrium of solutes. For the purpose of measuring the partition coefficients of very lipophilic or very hydrophilic compounds, it is important to have a maximal retention of the stationary phase. In our experience, a length of 50–60 m of tubing is sufficient to ensure a partitioning equilibrium of solutes between the two phases. Increasing the internal diameter of tubing generally results in a greater retention volume of the stationary phase, but the use of tubing with an i. d. >2.6 mm does not improve further the retention. The polytetrafluoroethylene (PTFE) tubing with an i. d. of 2.6 mm and an o. d. 3.4 mm (#10, Zeus Industrial Products, Orangeburg, Southern California, USA) has thus become our optimal choice. This internal diameter together with a length of 50–60 m yields a total volume of ca. 300–320 ml.

Another factor critical to the accuracy of measurements is the quality of tubing. The chemical and mechanical properties of tubing must be carefully considered. To allow most organic solvents and solutes at all pH values, it must be chemically resistant. As the tubing is under centrifugal forces, mechanical resistance should also be taken into account. Also, the hardness of tubing would surely influence the adsorption of solutes on the tubing and hence the peak symmetry of chromatograms. The PTFE tubing fulfills satisfactorily these criteria and is economically reasonable.

6.4.3 Experimental Design

Experimental layout for the measurement of partition coefficients using CPC is similar to that of capacity factors using HPLC, except that CPC uses a centrifuge mounted with coiled columns rather than a solid phase column (Fig. 6). However, experimental design in various types of CPC can be quite different. The differences are particularly marked for HSES and HDES. Experimental designs from our laboratory with different CPC systems have been described elsewhere [10], only those with flow-through multilayer CPC being detailed here.

A Kontron model 420 HPLC pump (Kontron Instrument, Zürich, Switzerland) was used to propel the mobile phase at a flow rate of 0.5–10 ml/min and a Kontron model 432 UV-visible detector (variable wavelength) coupled with a Hewlett-Packard 3392A integrator (Hewlett-Packard, Avondale, Pennsylvania, USA) to detect the eluate. A flowmeter (Phase Separations, Queensferry, UK) is used to measure precisely the flow-rate. It must be noted that a stable flow-rate is of critical importance for the measurements.

It is important to estimate the partition coefficient of the investigated compounds in order to set up optimal experimental conditions. We have used the CLOGP algorithm of Hansch and Leo [24] to estimate partition coefficients in n-octanol/water systems. As for partition coefficients in the other systems, a good guess based on the log P_{oct} value and hydrogen-bonding capacity is often needed. For the operation procedures, we use a "normal mode" process. Namely, for compounds of log P values > 0, the organic phase is used as the mobile phase in a column with a total volume of c. 300 ml. For compounds of log P values < 0, the aqueous phase is used as mobile phase. Measurement begins by filling the columns at a flow-rate of 5 ml/min, the stationary phase being presaturated with the mobile phase. When filling the columns with volatile solvents such as n-heptane, the flow-rate should be reduced to prevent the accumulation of air bubbles in the columns. When the columns are full, the centrifuge is rotated at a speed of c. 800–1000 r.p.m. and the mobile phase is propelled into the columns. For different log P ranges, the flow-rate of mobile phase should be accordingly adjusted in
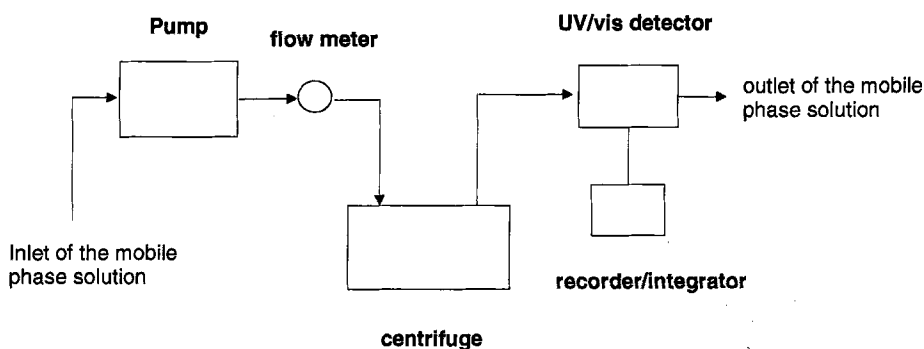
Figure 6.   Experimental layout of centrifugal partition chromatographs for partition coefficient measurements.

order to elute the solute at an appropriate retention time. Here we recommend an approximate scheme for the use of flow-rate at different log P ranges (see also Fig. 7):

For log $P > 2.5$ or log $P < -2.5$ $U = 0.5$ ml/min

For $1.5 < \log P < 2.5$ or $-2.5 < \log P < -1.5$ $U = 1$ ml/min

For $0.5 < \log P < 1.5$ or $-1.5 < \log P < -0.5$ $U = 3$ ml/min

For $0 < \log P < 0.5$ or $-0.5 < \log P < 0$ $U = 6$ ml/min

Under flow rates of 0.5, 1, 3 ad 6 ml/min ca. 88 %, 86 %, 83 % and 77 % of the stationary phase of an octanol/water system can be retained.

After the system has reached its equilibrium, i. e., no more stationary phase exudes from the columns, a Merck injector is used to inject 20 µl of samples containing 0.1–5 mM of solutes dissolved in the mobile phase. The amount of sample injected should be appropriately increased by increasing either the concentration or the injected volume when using a higher flow-rate.

A precise determination of the dead volume or retention time of non retained solutes (t_0) is of critical importance for the accurate measurement of partition coefficients, particularly for compounds with log $P > 2.5$ or < -2.5. In the past, we used either anthracene or biphenyl as the nonretained solute when the organic solvent was the eluent. However, anthracene is easily oxidized in solution, while biphenyl is more stable and has hence become our preferred standard of t_0 determination. As for nonretained hydrophilic compounds when water is the eluent, potassium dichromate is used satisfactorily in our laboratory. However, potassium dichromate appears to undergo

U (ml/min, the flow rate of mobile phase)
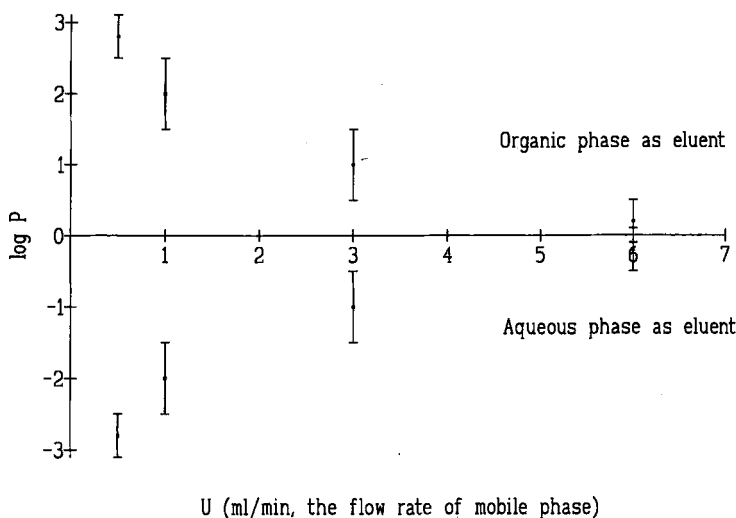
Figure 7. A proposed scheme for the operational flow-rates of the mobile phase in different log P ranges. The flow-rates of the mobile phase are selected and adjusted according to the different log P ranges of the investigated compounds.

chemical reactions and/or protonation at pH < ca. 3 and partition into the organic phase. In this case, cobalt chloride can be used to replace potassium dichromate for dead time determination.

In principle, UV-vis spectroscopy is not the only mode to detect eluates. Refractometry should allow detection of compounds lacking a chromophore. However, we have not yet investigated UV-inactive compounds for this type of measurement.

It is important to note that it is not necessary to empty and refill the columns each day provided that the same solvent system is used. The organic solvents can be reused after distillation and saturation with aqueous solution. Before changing the solvent system, it is advisable to wash the coils with methanol and dry them with a flow of air.

The above-mentioned experimental design and procedures are not applicable if a small column of say 30 ml is to be used. In such a case, the operation mode – called "reversed mode" in contrast to the "normal mode" described above – is completely different. For example, the aqueous phase is recommended as the mobile phase when measuring compounds of log $P > 0$. Since this type of experimental design would limit the measured range of log P values (ca. –2 to 2), it was not employed in our laboratory. However, this mode of operation may be used in combination with a column having a small total volume.

6.5 Validation of log P values Obtained from CPC

One of the main problems about CPC for log P measurements concerns solute partition equilibria in the chromatograph. Since a measurement usually takes from 10 minutes to 1 hour to complete, it is questionable whether the solute partition equilibrium is indeed reached in such a short period. Furthermore, different solvent systems possess rather distinct interfacial tensions ($\gamma = 8.5$ mN m^{-1} at 20 °C) in comparison with alkane/water ($\gamma = $ c. 50 mN m^{-1} at 20 °C) [25] (see Table 1). This implies that the droplet size of emulsion formed at the inner part of the column close to the central axis would be much larger in alkane/water than in n-octanol/water under the same experi-

Table 1.   Interfacial tensions (γ in mN m^{-1}) of some organic solvents against water at 20 °C [25]

Organic Solvent	γ
n-Hexadecane	54
n-Dodecane	53
n-Octane	52
Cyclohexane	51
Tetrachloromethane	45
Benzene	35
Chloroform	33
Oleic acid	16
Diethyl ether	11
n-Octanol	8.5
n-Butanol	1.6

mental conditions. Consequently, the total contact surface area between the two phases would be smaller in alkane/water than in n-octanol/water, suggesting that partition equilibria take longer to be reached in alkane/water. This is certainly an important aspect when using the shake-flask method for partition coefficient measurements. It will be clear later that the mixing efficiency in the coil planet-type centrifuge seems to be sufficient to ensure that solute partition equilibria are reached in a relatively short time.

6.5.1 Partition Coefficients in *n*-Octanol/water Systems

One way, and perhaps the only one, to validate the partition coefficients measured by the CPC method is to compare them with "good" literature values measured by the conventional SF method. With the "normal mode" operation procedures, each compound was measured at least in triplicate and three measurements usually took no more than 2 hours. It would surely take much longer with "reversed mode" operation procedures. Among the compounds examined, which included alcohols, benzenes, phenols, anilines, benzamides, acetanilides, benzoic acids, benzenesulfonamides, amino acids, nucleosides, etc., excellent correlations between the partition coefficients measured by the two methods were found [10]:

$$\log P_{oct}^{SF} = 0.99 \ (\pm 0.01) \log P_{oct}^{CPC} - 0.01 \ (\pm 0.02) \tag{5}$$

$$n = 89, r^2 = 0.99, s = 0.12$$

where n is the number of compounds including model compounds and drugs, r is the correlation coefficient, and s is the standard deviation of the regression. The values in parentheses are the 95 % confidence limit of regression coefficients. These results have demonstrated the applicability of the CPC method for measuring partition coefficients. Since the precision of the obtained log P values depends much on the difference between t_R and t_0, our experience has shown that the CPC method is limited to compounds with log P values between ca. -3 and 3 [18]. Beyond this range, the solute elutes some seconds later than the nonretained compounds, resulting in large errors in the calculation of partition coefficients. While different modes of operation such as the interchange of operation modes could enlarge the range of measurable log P values, these methods are rather tedious and likely to impose additional errors.

The octanol phase is relatively viscous and the maximal proportion of stationary phase with a flow rate of 0.5–1 m/min is usually 87–90 %. Because the mutual solubility of the two phases could be increased at an elevated temperature, risking a loss of stationary phase during operation, the temperature in the centrifuge should be suitably regulated with ventilation or refrigerating systems. It is also necessary to control the dead time after several hours of operation since a change in t_0 may occur due to loss of stationary phase or a rise in phase solution temperature.

6.5.2 Partition Coefficients in Alkane/Water Systems

The measurement of alkane/water partition coefficients has become a necessity due in part to the emerging important Δlog P parameter [19]. With the coil planet-type CPC,

n-heptane/water was first selected for the determination of log P_{alk}. Excellent correlations in log P values were again obtained from the CPC and SF methods for the model compounds studied [10]:

$$\log P_{alk}^{SF} = 0.99\ (\pm 0.01) \log P_{alk}^{CPC} - 0.04\ (\pm 0.02) \tag{6}$$

$$n = 30,\ r^2 = 0.99,\ s = 0.08$$

The maximal volume of the stationary phase with a flow rate of 0.5–1 ml/min is usually > 95 %, leading again to a measurable log P range of –3 to 3. Due to the volatile nature of the heptane phase, small air bubbles were formed in the column during operation and eluted with the mobile phase, perturbing the detection of eluates. Its high volatility also interfered with the regulation of flow-rate which appeared to vary during the measurement. These two main drawbacks led us to switch to the less volatile *n*-dodecane for log P_{alk} determination [26]. It should be noted that the partition coefficients in different alkane/water systems such as *n*-heptane/water, *n*-dodecane/water and *n*-hexadecane/water are almost identical for the same compound [27]. Their differences in structural information are indistinguishable according to solvatochromic analyses. The *n*-dodecane/water [26] has circumvented the problems of *n*-heptane/water and been satisfactorily applied to the studies of structure-activity relationships in our laboratory.

6.5.3 Partition Coefficients in di-*n*-Butyl Ether/Water and Chloroform/Water Systems

While the partition coefficients from di-*n*-butyl ether/water (DBE/water) and chloroform/water (CHF/water) are expected to encode different structural information from log P_{oct} and log P_{alk}, there is yet no substantial applications of these parameters to structure-activity relationship studies. We have, however, undertaken the partition coefficient measurements of DBE/water and CHF/water systems using the CPC method.

The difficulties with DBE/water are similar to those encountered in heptane/water, irregular flow rates and air bubbles elution being frequently observed. Despite these problems, the log P_{DBE} values of ca. 60 compounds were measured and their structural information analyzed using solvatochromic parameters [28].

The measurements of log P_{CHF} are not straightforward as expected, in particular when chloroform is used as the mobile phase. A continuous loss of stationary phase was observed in the eluent. Thus far, we have not yet ventured to solve the problem.

6.6 Application to the Determination of Solute Structural Properties

6.6.1 The Case of Zwitterionic Amino Acids

The lipophilic expression of the side-chains of amino acids is of interest from biochemical and physicochemical viewpoints. On the one hand, the folding and stability of proteins are highly related to the hydrophobic nature of amino acid residues [29]. On the other hand, intra- and intermolecular interactions of zwitterionic amino acids in solu-

tion are reflected in their solvation/hydration behavior [30] and hence lipophilicity. Since the literature log P_{oct} values for these highly hydrophilic compounds were obtained from the SF method and, in some cases, varied markedly, the CPC method appeared to us an appropriate technique to reveal the structure-lipophilicity relationships of zwitterionic amino acids.

To see the lipophilic expression of the methylene groups in the side-chains of flexible α-amino acids, the log D^I values (distribution coefficient measured at isoelectric point) of α-amino acids with an alkyl chain length of 0 (glycine) to 4 (norleucine) were determined [31]. Figure 8 shows the non-additive log D^I increments of methylene groups.
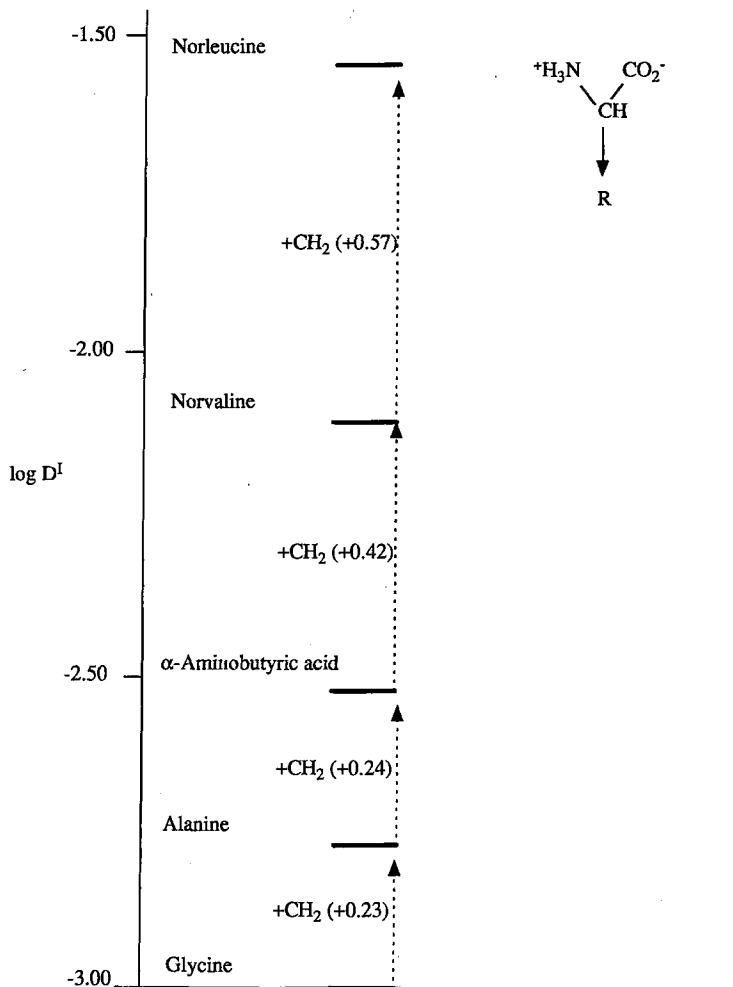
Figure 8. Nonadditive log D^I increments of methylene groups in the flexible side-chains of zwitterionic amino acids. The lipophilic expression of methylene groups near the zwitterionic dipole is partially masked. Only when the fourth methylene group is added, does it gain a normal log D^I increment of a methylene group (ca. 0.6).

Indeed, abnormal lipophilic increments of methylene groups are usually observed when they are in close proximity to polar moieties due, at least in part, to the electronic structure of methylene groups near polar entities differing from that of distal ones. However, hydration features must also be considered here since proximity effects have been postulated to be a consequence of hydration effects.

The influence of intercharge distance on lipophilicity was also examined using conformationally defined positional isomers of piperidinyl carboxylic acids, i. e., pipecolic acid, nipecotic acid, and isonipecotic acid [31]. A plot of log D^I against the intercharge distance reveals that increasing the intercharge distance decreases the log D^I values with a slope of c. 0.4 units Å^{-1}.

6.6.2 The Case of Anti-Dopaminergic 6-Methoxysalicylamides

In the course of studying the physico-chemical and structural properties of antidopaminergic 6-methoxysalicylamides [32, 33], the compound raclopride was shown by us to exist in water at pH 7.4 as a zwitterion with stereoelectronic features completely different from those of other classes of dopamine antagonists such as orthopramides. Using first-derivative UV spectroscopy, the uncharged rather than zwitterionic form of raclopride was found to partition predominantly into the octanol phase. Using the CPC method, we established its pH-lipophilicity profile in the two solvent systems octanol/water and heptane/water [33]. Furthermore, the very low $\Delta\log P_{\text{oct-alk}}$ value of raclopride (0.5) indicates the existence of strong internal hydrogen bonds (OH...O=C and CH_3O...HNC=O) in biological lipophilic media. A simplified scheme showing the interconversion, predominance, and conformational behavior of neutral and zwitterionic raclopride in a biphasic system is presented in Fig. 9. These results thus help in revealing the true pharmacophoric features of 6-methoxysalicylamides which are therefore stereoelectronically similar to those of other orthopramides.

6.7 Advantages and Limitations of the CPC Method for log *P* Measurements

The CPC method is probably the most precise and accurate method existing thus far for partition coefficient measurements. The problems inherent in the SF method caused by the adsorption, instability, impurity and volatility of solutes, and the impurity of solvents are not at all disturbing in the CPC method. When using different phase volume ratios for the same compound, the differences in the obtained log *P* values are negligible. This novel method needs relatively short times for a measurement in comparison with the shake-flask method due to its efficient mixing. Furthermore, the amounts of samples needed for the measurement are minute and comparable with the RP-HPLC method. In principle, this technique can be automated using an autosampler and computer monitoring if no mechanical problems occur during the measurement; however, we have not tried this possibility yet.
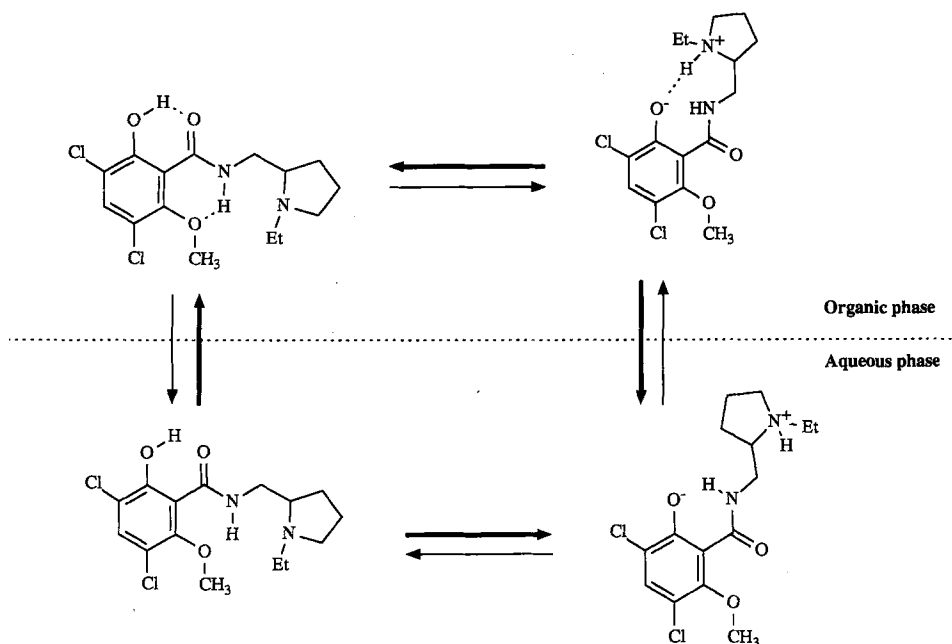
Figure 9. The interconversion, predominance, and conformational behavior of uncharged and zwitterionic raclopride in a biphasic system. In the aqueous phase at pH 7.4, raclopride exists predominantly at a zwitterionic state. The zwitterionic raclopride is, however, relatively unstable and converted to an uncharged form having two internal hydrogen bonds as suggested by its low $\Delta\log P$ value (0.5).

The CPC method has limitations of its own: the limited range of measurable $\log P$ values (−3 to 3) is a genuine one, but it concerns a small proportion of compounds of interest. While methods of different operational modes using Sanki CPC have been applied to measure highly lipophilic compounds [34], they were impracticable due to extremely long retention time (15 hours for a compound of $\log P = 3$). In comparison with the RP-HPLC method, it consumes relatively large amounts of organic solvent when using the organic phase as the mobile phase and running at a flow-rate of 3–6 ml/min. It should also be noted that mechanical expertise is highly desirable in order to maintain the functioning of the centrifuge, since the current engineering standards of the centrifuge have yet to be improved.

6.8 Concluding Remarks

The liquid–liquid chromatographic method CPC allows for the first time very accurate partition coefficients to be measured in biphasic organic/aqueous systems such as n-octanol/water and n-dodecane/water. The structural information of drug molecules such as hydrogen-bonding capacity can thus be unambiguously revealed using the $\Delta\log P$

parameter, providing better insight into the mechanisms of drug-biomembrane and drug-receptor interactions.

While the measurable log P range is limited to –3 to 3, it seems appropriate for most drug compounds by suitably modifying experimental conditions, e. g., pH. However, mechanical expertise about the centrifuge is inevitable for the introduction of this novel technique into any laboratory.

Acknowledgements

References

[1] Hansch, C., and Leo, A., *Substituent Constants for Correlation Analysis in Chemistry and Biology.* Wiley: New York 1979

[2] Rekker, R. F., *The Hydrophobic Fragmental Constant.* Elsevier: Amsterdam 1977

[3] Hansch, C., Vittoria, A., Silipo, C., and Jow, P. Y. C., *J. Med. Chem.* **18**, 546–548 (1975)

[4] Young, R. C., Mitchell, R. C., Brown, T. H., Ganellin, C. R., Griffiths, R., Jones, M., Rana, K. K., Saunders, D., Smith, I. R., Nerina, E. S., and Wilks, T. J. *J. Med. Chem.* **31**, 656–671 (1988)

[5] Burton, P. S., Conradi, R. A., Hilgers, A. R., Ho, N. F. H., and Maggiora, L. L., *J. Controlled Release* **19**, 87–98 (1992)

[6] El Tayar, N., Tsai, R.-S., Testa, B., Carrupt, P.-A., Hansch, C., and Leo, A. *J. Pharm. Sci.* **80**, 744–749 (1991)

[7] Leo, A., Hansch, C., and Elkins, D., *Chem. Rev.* **71**, 525–555 (1971)

[8] Dearden, J. C., and Bresnen, G. M., *Quant. Struc.-Act. Relat.* **7**, 133–144 (1988)

[9] Lambert, W. J., *J. Chromatogr.* **656**, 469–484 (1993)

[10] El Tayar, N., Tsai, R.-S., Vallat, P., Altomare, C., and Testa, B., *J. Chromatogr.* **556**, 181–194 (1991)

[11] Martin, A. J. P., and Synge, R. L. M., *Biochem. J.*, **35**, 1358–1368 (1941)

[12] Horváth, C., and Lipsky, S. R., *Nature* **211**, 748–749 (1966)

[13] Ito, Y., Weinstein, M. A., Aoki, I., Harada, R., Kimura, E., and Nunogaki, K., *Nature* **212**, 985–987 (1966)

[14] Terada, H., Kosuge, Y., Murayama, W., Nakaya, N., Nunogaki, Y. and Nunogaki, K.-I., *J. Chromatogr.* **400**, 343–351 (1987)

[15] Berthod, A., Han, Y. I., and Armstrong, D. W., *J. Liq. Chromatogr.* **11**, 1441–1456 (1988)

[16] Et Tayar, N., Marston, A., Bechalany, A., Hostettmann, K., and Testa, B., *J. Chromatogr.* **469**, 91–99 (1989)

[17] Vallat, P., El Tayar, N., Testa, B., Slacanin, I., Marston, A., and Hostettmann, K., *J. Chromatogr.* **504**, 411–419 (1990)

[18] Tsai, R.-S., Carrupt, P.-A., and Testa, B., Measurement of partition coefficient using centrifugal chromatography method development and application to the determination of solute properties. In: *Countercurrent Chromatography.* Conway, W. D. and Petroski, R. (Eds.). ACS Symposium Series 593, American Chemical Society: Washington, DC; 143–154 (1995)

[19] El Tayar, N., Tsai, R.-S., Testa, B., Carrupt, P. A., and Leo, A., *J. Pharm. Sci.* **80**, 590–598 (1991)

[20] Tanimura, T., Pisano, J. J., Ito, Y., and Bowman, R. L., *Science* **169**, 54–56 (1970)

[21] Ito, Y., *CRC Crit. Rev. Anal. Chem.* **17**, 65–148 (1986)

[22] Foucault, A. P., (Ed.). *Centrifugal Partition Chromatography.* Marcel Dekker: New York 1994

[23] Conway, W. D., and Ito, Y., Abstract of the 1984 Pittsburgh Conference and Exposition on Analytical Chemistry and Applied Spectroscopy, 471 (1984)

[24] Hansch, C., and Leo, A., *The Pomona College Medicinal Chemistry Project.* Version 3.55, Clarement 1993

[25] Zografi, G., and Yalkowsky, S. H., *J. Pharm. Sci.* **63**, 1533–1536 (1974)

[26] Ter Laak, T., Tsai, R.-S., Donné-Op den Kelder, G. M., Carrupt, P.-A., Testa, B., and Timmerman, H., *Eur. J. Pharm. Sci.* **2**, 373–384 (1994)

[27] Seiler, P., *Eur. J. Med. Chem.* **9**, 473–479 (1974)

[28] Fan, W., *A New Experimental Hydration Parameter: the Water-Dragging Effect.* PhD Thesis, Université de Lausanne, Lausanne 1992

[29] Rose, G. D., Gierash, L. M., and Smith, J. A., *Adv. Protein Chem.* **37**, 1–109 (1985)

[30] Abraham, D. J., and Leo, A., *Proteins: Structure, Function, and Genetics* **2**, 130–152 (1987)

[31] Tsai, R.-S., Testa, B., El Tayar, N., and Carrupt, P.-A., *J. Chem. Soc. Perkins Trans.* **II**, 1797–1802 (1991)

[32] Carrupt, P.-A., Tsai, R.-S., El Tayar, N., Testa, B., de Paulis, T., and Högberg, T., *Helv. Chim. Acta* **74**, 956–968 (1991)

[33] Tsai, R.-S., Carrupt, P.-A., Testa, B., Gaillard, P., El Tayar, N., and Högberg, T., *J. Med. Chem.* **36**, 196–204 (1993)

[34] Berthod, A., Menges, R. A., and Armstrong, D. W., *J. Liq. Chromatogr.* **16**, 2769–2785 (1992)

7 Assessment of Distribution-pH Profiles

Alex Avdeef

Abbreviations

Pg	Prostaglandin, anion
Cpz	Chlorpromazine, free base
M	Morphine
M3G	Morphine-3β-D-glucuronide
M6G	Morphine-6β-D-glucuronide
Glu	*N*-methyl-*D*-glucamine, free base
Ibu	Ibuprofen, anion
DOPC	Dioleylphosphatidylcholine
eggPC	Egg phosphatidylcholine

Symbols

D — Lipid-water distribution pH-dependent function (also called the apparent partition coefficient, PC_{app}), usually expressed in logarithmic (base 10) form as log D

P — Lipid-water pH-independent partition coefficient (sometimes called K_{ow} or PC), usually expressed as log P

K_e — Extraction constant

pK_a — Ionization constant (negative log form)

p_oK_a — Apparent ionization constant in an octanol/water titration

$p_oK_a^{SCH}$ — "Scherrer" pK_a (also $p_oK_a^{lim}$), the limiting p_oK_a in titrations with very high octanol/water volume ratios

β — Cumulative stability constant

$\bar{n}_H$ — Bjerrum "difference" function

pH — operational pH scale

p_cH — pH scale based on hydrogen ion concentration

Δlog P — Difference in log P between neutral and ionized species

log p — Micro partition constant

7.1 Introduction

The Dyrssen [1] dual-phase potentiometric titration method for measuring the oil-water partition coefficient, log P, has undergone considerable recent development [2–9]. The robust pH-metric technique can be used to determine the ionization constants (pK_a) and the partition constants (log P and log K_e, the ion-pair extraction constant [10]) of ionizable drug substances. The present review will focus on how these

equilibrium constants are used to calculate the lipophilicity profile, which is a plot of the apparent partition coefficient, that is, the log of the distribution function, log D, vs pH. The mathematical basis of the fully generalized distribution function (applicable to multiprotic substances or substances which undergo aggregation or complexation reactions), the experimental design and limitations of the Dyrssen technique, the effect of salt on lipophilicity, the relationship between the aqueous pH scale and the pH scale in the lipid phase, the distinctions between *micro-* and *macro*constant-based calculations of the lipophilicity profile will be considered. Applications to a variety of pharmaceutically interesting compounds, both lipophilic and hydrophilic, will be presented.

7.2 Partition Coefficient, log *P*, and the Dyrssen Two-Phase Titration

7.2.1 Historical Background

In a series of papers (1952–1957) on solvent extraction of metal complexes, Dyrssen and coworkers [1,11–14] and Rydberg [15] appear to be the first to have applied the dual-phase potentiometric titration to determining log *P*. Not only that, but they considered ion-pair log *P*, studied dimerization reactions of dialkylphosphates in aqueous as well as chloroform solutions, and used log *D* vs pH plots. Brändström [16] and Seiler [17] were also among the earliest users of the pH-metric log *P* technique. Brändström performed constant pH titrations (using a pH-stat) to determine log *P*, while Seiler proposed the determination of pK_a and log *P* from a single titration, where octanol is added half way through the assay, near the half-ionization point. Kaufman et al. [18] introduced the use of difference plots to the interpretation of log *P* titration data. There were other significant early studies which contributed to the development of the pH-metric technique [19–26].

7.2.2 Titrations

The pH-metric technique consists of two linked titrations. Typically, a pre-acidified solution of a weak acid is alkalimetrically titrated to some appropriately high pH; octanol (or any other useful organic partition solvent that is immiscible with water) is then added, and the dual-solvent mixture is acidimetrically titrated back to the starting pH. After each titrant addition, pH is measured. If the weak acid partitions into the octanol phase, the two assays show non overlapping titration curves, as shown in Fig. 1a for flumequine and Fig. 1b for diacetylmorphine. (The horizontal axes in Fig. 1 are in units of base equivalents, defined as the number of moles of base titrant added per mole of drug substance titrated.) The greatest divergence between the two curves occurs in the buffer region. Since the pK_a is approximately equal to the pH at the mid-buffer inflection point (equivalent to 0.5 in Fig. 1), the two-part assay yields two constants: pK_a and p$_o K_a$, where p$_o K_a$ is the *apparent* constant derived from the octanol-containing segment of data. A large difference between pK_a and p$_o K_a$ indicates a large value of log *P*.

(a) Flumequine (25°C, 0.15M NaCl)

pH

Equivalents

(b) Diacetylmorphine (25°C, 0.15M KCl)

pH

Equivalents
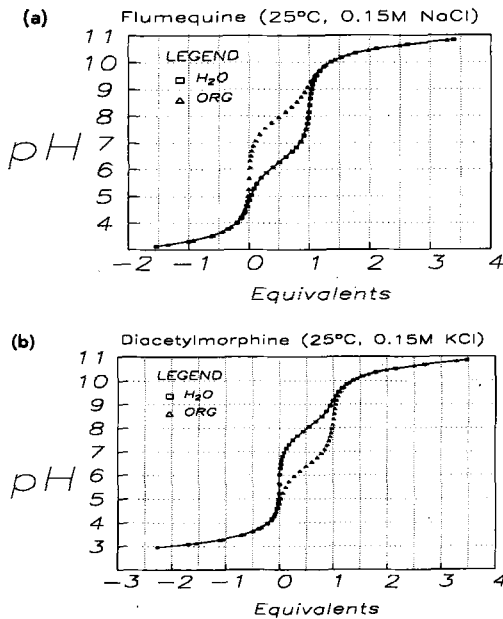
Figure 1. Water (squares) and 1 : 1 octanol/water (triangles) titration curves (0.15 M KCl) of (a) weak acid, flumequine (pK_a 6.27), and (b) weak base diacetylmorphine (pK_a 7.96). Octanol causes the curve to shift to higher pH with the weak acid and to lower pH with the weak base.

For a *monoprotic weak acid* which dissolves both in water and a water-immiscible partition solvent (e. g., octanol), the partition coefficient (which is a true equilibrium constant) is defined as

$$P_{HA} = [HA]_{ORG}/[HA] \tag{1}$$

where $[HA]$ is the weak acid aqueous concentration, moles/liter aqueous solution, and $[HA]_{ORG}$ is the concentration in the oil phase, moles/liter of organic solvent. For this simple system the relation between log P and pK_a is simply

$$P_{HA} = \frac{10^{p_oK_a - pK_a} - 1}{r} \tag{2}$$

where

$$r = \frac{\text{vol. organic phase}}{\text{vol. aqueous phase}} \tag{3}$$

If the two phases are equal in volume and the substance is considerably lipophilic, then Eq. (2) reduces to log $P_{HA} \approx (p_oK_a - pK_a)$. In Fig. 1a, the octanol volume equals the aqueous volume, so by inspection, log $P = 1.7$ for flumequine.

For a *monoprotic weak base*, B, partitioning the corresponding equation is

$$P_B = \frac{10^{-(p_oK_a - pK_a)} - 1}{r} \tag{4}$$

Similarly, for equivolume titrations of lipophilic bases, log $P_B \approx -(p_oK_a - pK_a)$. In Fig. 1b, the octanol volume equals the aqueous volume, so by inspection, log $P = 1.6$ for diacetylmorphine.

7.2.3 Bjerrum Difference Plots

The Bjerrum difference plots [3, 21, 27, 28] are probably the most important graphical tools in the initial stages of equilibrium analysis. The difference curve is a plot of $\bar{n}_H$, the average number of bound protons (that is, the hydrogen ion-binding capacity), versus p_cH ($-\log[H^+]$). Such a plot can be obtained by subtracting a titration curve containing no sample ("blank" titration) from a titration curve with sample, hence the name "difference" curve. Another way of looking at it is the following. Since one knows how much strong acid and strong base have been added to the solution at any point and since one knows how many dissociable protons the sample substance brings to the solution, one knows the *total* hydrogen ion concentration in solution, regardless of what equilibrium reactions are taking place. By measuring the pH (and after converting it into p_cH), one knows the *free* hydrogen ion concentration. The difference between the total and the free concentrations is equal to the concentration of the *bound* hydrogen ions. The latter concentration divided by that of the sample substance gives the average number of bound hydrogen ions per molecule of substance, $\bar{n}_H$.

Fig. 2 shows the difference plots for the simple monoprotic examples flumequine and diacetylmorphine. Fig. 3 shows plots for the diprotic examples salicylic acid, morphine, and nicotine. The difference plots in Fig. 2 and 3 reveal all the pK_as and p_oK_as as p_cH values at half-integral $\bar{n}_H$ positions. By mere inspection, the pK_as of salicylic acid can be estimated to be 2.9 and 13.3, while the p_oK_as are 5.1 and very slightly greater than 13.3 (as indicated in the inset in Fig. 3a). It would not have been possible to deduce the constants by simple inspection of the titration curves, pH vs base equivalents. For morphine, the approximate pK_as are 8.1 and 9.3, while the corresponding p_oK_as are 7.2 and 10.2; for nicotine, the pK_as are 3.2 and 8.1, while the p_oK_as are 3.2 and 6.8. The differences between the pK_as and p_oK_as can be used to determine log P constants, using equations slightly more complicated than Eqs. (2) and (4), as detailed elsewhere
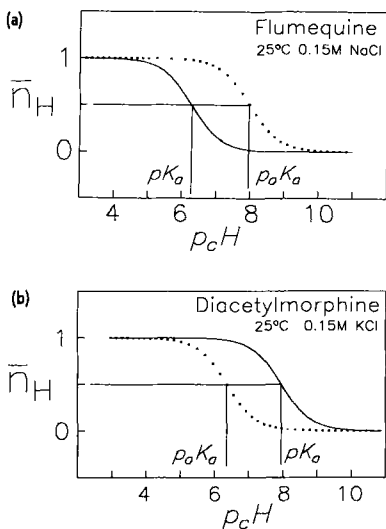
Figure 2. Water (solid curves) and 1 : 1 octanol/water (dotted curves) difference plots (0.15 M KCl) of (a) flumequine (pK_a 6.27), (b) diacetylmorphine (pK_a 7.96).
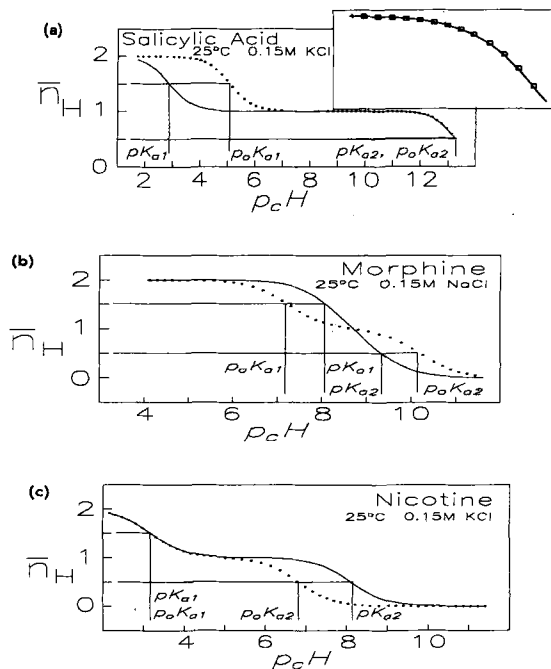
Figure 3. Water (solid curves) and 1 : 1 octanol/water (dotted curves) difference plots (0.15 M KCl) of (a) salicylic acid (pK_a 13.31 and 2.88), (b) morphine (pK_a 9.26 and 8.17), and (c) nicotine (pK_a 8.11 and 3.17).

[3]. Difference curve analysis often gives one the needed "seed" values for refinement of equilibrium constants by mass-balance-based nonlinear least squares [4, 25].

7.2.4 pH Definitions and Electrode Standardization

To establish the operational pH scale, the pH electrode is first calibrated with a single aqueous pH 7 phosphate buffer and the Nernst slope is assumed. (Most users of pH electrodes are familiar with this step.) Because the $\bar{n}_H$ calculation requires the "free" hydrogen ion concentration (as described in the preceding section) and because we employ the concentration scale for the ionization constants (next section), and additional electrode standardization step is necessary. That is where the operational scale is converted to the *concentration* scale p_cH ($= -\log[H^+]$) using the four-parameter equation [5,20],

$$pH = \alpha + S\,p_cH + j_H[H^+] + j_{OH}\,K_w/[H^+] \tag{5}$$

where K_w is the ionization constant of water. The four parameters are empirically determined by a weighted nonlinear least squares procedure using data from alkalimetric titrations of known concentrations of HCl (from pH 1.8 to 12.2) or standard buffers [20]. Typical aqueous values of the adjustable parameters at 25 °C and 0.15 M ionic strength are $\alpha = 0.08 \pm 0.01$, $S = 1.001 \pm 0.001$, $j_H = 1.0 \pm 0.2$ and $j_{OH} = -0.6 \pm 0.2$.

7.2.5 Definitions of Constants

All equilibrium constants in the present discussion are based on the concentration scale. This is a perfectly legitimate thermodynamic scale, provided the ionic strength of the solvent medium is kept constant. Most of the results reported here were determined in 0.15 M KCl or NaCl.

An example will illustrate the definitions of the equilibrium constants used here. Consider the diprotic amino acid phenylalanine (Phe); let us say X^- represents the fully deprotonated Phe species. At the physiological level of background salt (0.16 M NaCl) X partitions both as the zwitterion and the cation: log $P(XH^+)$ = −1.38 and log $P(XH_2^+)$ = −1.41; the dissociation constants are pK_{a1} = 2.20 and pK_{a2} = 9.08. Ion-pair partitioning can be characterized either by a log P constant (which is only valid at a particular level and type of background salt, hence a "conditional" constant) or by an extraction constant, log K_e. For the above example, log K_e = −0.62. The equilibrium reactions corresponding to the above constants (in the order of mention) are

$$XH^{\pm} \leftrightarrows (XH^{\pm})_{OCT}: \qquad P = [XH^{\pm}]_{OCT}/[XH^{\pm}] \qquad (6)$$

$$XH_2^{+} \leftrightarrows (XH_2^{+})_{OCT}: \qquad P = [XH_2^{+}]_{OCT}/[HX_2^{+}] \qquad (7)$$

$$XH_2^{+} \leftrightarrows XH^{\pm} + H^{+}: \qquad K_{a1} = [XH^{\pm}]\,[H^{+}]/[XH_2^{+}] \qquad (8)$$

$$XH^{\pm} \leftrightarrows X^{-} + H^{+}: \qquad K_{a2} = [X^{-}]\,[H^{+}]/[XH^{\pm}] \qquad (9)$$

$$XH_2^{+} + Cl^{-} \leftrightarrows (XH_2^{+}, Cl^{-})_{OCT}: \qquad K_e = [XH_2^{+}, Cl^{-}]_{OCT}/[XH_2^{+}][Cl^{-}] \qquad (10)$$
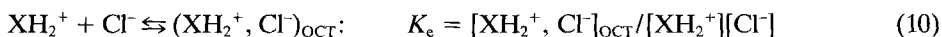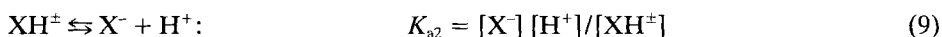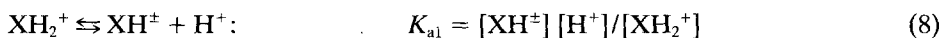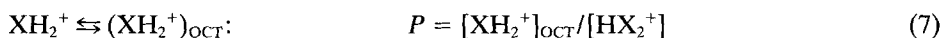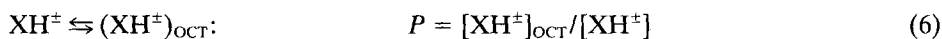
Eqs. (7) and (10) are two ways of expressing the same ion-pair partitioning. One can convert log $P(XH_2^+)$ to log K_e according to the relation.

$$\log K_e = \log P(XH_2^{+}) - \log [Cl^{-}] \qquad (11)$$

Since all components of the extraction equilibrium expression are explicitly identified in Eq. (10), the ionic-strength dependence of the extraction constant can be predicted by the Debye-Hückel theory. That is, in principle, a result determined at 0.15 M KCl background, could be "corrected" to another background salt concentration, provided the ionic strength is within the limitations of the theory (< 0.5 M for the Davies [29] variant of the Debye-Hückel expression). It is often convenient to convert constants to "zero ionic strength" in order to compare values to those reported in older literature.

The "log β" form of constants are also used in the discussion. For the above example,

$$X^{-} + H^{+} \leftrightarrows XH^{\pm}: \qquad \beta_{11} = K_{a2}^{-1} \qquad (12)$$

$$X^{-} + 2\,H^{+} \leftrightarrows XH_2^{+}: \qquad \beta_{12} = K_{a1}^{-1}K_{a2}^{-1} \qquad (13)$$

These are called stability constants, and are quite useful in many calculations involving multiprotic substances. The double-digit β subscript indices refer to the stoichiometric coefficients of the species on the right side of the equilibrium equation, in the order X, then H. For phenylalanine, log β_{11} = 9.08 and log β_{12} = 11.28 (9.08 + 2.20).

7.3 Distribution Function (D) and Lipophilicity Profile (log D vs pH)

The distribution function, D [1,30], is a useful parameter related to the partition coefficient, P, and the two parameters are unfortunately sometimes interchanged in the literature. The partition coefficient, P is a *constant* and refers to a *single* molecular species partitioning between two phases. On the other hand, the distribution coefficient, D, can be thought of as an effective (or apparent) partition coefficient which varies with pH when ionizable substances are considered. In the pH region where a substance is unionized, $D = P$, and the substance is usually maximally partitioning in the organic phase. However, as pH is changed (increased for weak acids or decreased for weak bases), the substance ionizes. This in turn causes its redistribution between the two phases, with more substance shifting to the aqueous phase. The concentration ratio P is still the same, but D decreases, reflecting that there is more of the substance in ionic form, which favors the water phase.

Let us consider the protophilic substance X (neutral or charged). The general definition of the distribution coefficient is

$$D = X_{ORG}/X$$

$$= (X_{ORG}'/X)r$$

$$= \{ ([X]_{ORG}' + [XH]_{ORG}' + [XH_2]_{ORG}' + \ldots) / ([X] + [XH] + [XH_2] + \ldots) \}/r \qquad (14)$$

where we use the notation (italic) X to represent the *total* concentration of the substance in all of its protolytic forms. The primed quantity is defined in concentration units of moles of species dissolved in the organic phase per liter of *aqueous* phase; the octanol/water volume ratio, r, takes into account the change in volume scale. Assumptions must not be made as to which species partitions predominant into the organic phase. If the fully-deprotonated base X partitions into the organic phase, we have

$$P_{10} = x_{ORG}'/x = (x_{ORG}'/x)/r \qquad (15)$$

Here, lower case (italic) x refers to the concentration of the fully deprotonated substance X (the conjugate base). The index 10 refers to stoichiometric coefficients of the partitioning species: the 1 refers to one unit of substance X and the 0 refers to the number of protons associated with X (i. e., fully-deprotonated base). This type of indexing has been used already for the β cumulative protonation constants in Eqs. (12) and (13). We will find this general notation increasingly useful as we consider increasingly complicated protonation reactions. If a j-protonated from of X partitions into the organic phase, we have

$$P_{1j} = [XH_j]_{ORG} / [XH_j]$$

$$= ([XH_j]_{ORG}' / [XH_2])/r \qquad (16)$$

We can rearrange Eqs. (15) and (16) into very useful forms.

$$x_{ORG}' = x \, r \, P_{10} \qquad (17)$$

$$[XH_j]_{ORG}' = x \, h^j \, \beta_{1j} \, r \, P_{1j} \qquad (18)$$

Here h denotes [H^+]. Substituting Eqs. (17) and (18) into the last line of Eq. (14), we can state the distribution function in terms independent of x.

$$D = \frac{P_{10} + h\,\beta_{11}P_{11} + h^2\,\beta_{12}P_{12} + \dots}{1 + h\,\beta_{11} + h^2\,\beta_{12} + \dots} \tag{19}$$

The distribution coefficient is only a function of the variable p_cH and the pK_a (expressed in β form) and log P constants. The latter constants depend on ionic strength. So, the calculation of D associated with a titration can be quite complicated, depending on the level of ionic strength adjuster, the sample concentration, and dilution effects as a result of weak titrant concentrations or extreme starting and/or ending pH values in an assay. Ion-pair partitioning is also an added complication when the background salt changes in concentration due to dilution effects. A further subtlety may arise: the lipophilicity profile constructed from measurements by the shake-flask technique is invariably expressed as a function of the operational (activity) pH scale. All calculations here are based on the p_cH concentration scale. This distinction must be kept in mind when comparing literature values.

Eq. (19) is applicable to all lipophilicity calculations. It may be helpful to identify special cases where the lipophilicity equation is considerably simplified, and made more easily comparable to forms found in the literature. Table 1 shows the simplified

Table 7.1. Lipophilicity equations

Monoprotic substance

1a. *XH partitions (weak acid)*
   (e.g., flumequine)
$$\log D = \log P_{XH} - \log(1 + 10^{-pK_a + p_cH})$$

1b. *X partitions (weak base)*
   (e.g., diacetylmorphine)
$$\log D = \log P_X - \log(1 + 10^{+pK_a - p_cH})$$

Diprotic substance

2a. *XH$_2$ partitions (weak acid)*
   (e.g., salicylic acid)
$$\log D = \log P_{XH_2} - \log(1 + 10^{-pK_{a1} + p_cH} + 10^{-pK_{a2} - pK_{a1} + 2p_cH})$$

2b. *XH partitions (ampholyte or zwitterion)*
   (e.g., morphine)
$$\log D = \log P_{XH} - \log(1 + 10^{-pK_{a2} + p_cH} + 10^{+pK_{a1} - p_cH})$$

2c. *X partitions (weak base)*
   (e.g., nicotine)
$$\log D = \log P_X - \log(1 + 10^{+pK_{a2} - p_cH} + 10^{+pK_{a2} + pK_{a1} - 2p_cH})$$

Table 7.1. Continued

Triprotic Substance

3a. *XH_3 partitions (weak acid)*
(e.g., citric acid)

$$\log D = \log P_{XH_3} - \log\left(1 + 10^{-pK_{a1}+p_cH} + 10^{-pK_{a2}-pK_{a1}+2p_cH} + 10^{-pK_{a3}-pK_{a2}-pK_{a1}+3p_cH}\right)$$

3b. *XH_2 partitions*
(e.g., terbutaline)

$$\log D = \log P_{XH_2} - \log\left(1 + 10^{+pK_{a1}-p_cH} + 10^{-pK_{a2}+p_cH} + 10^{-pK_{a3}-pK_{a2}+2p_cH}\right)$$

3c. *XH partitions*

$$\log D = \log P_{XH} - \log\left(1 + 10^{-pK_{a3}+p_cH} + 10^{+pK_{a2}-p_cH} + 10^{+pK_{a2}+pK_{a1}-2p_cH}\right)$$

3d. *X partitions (weak base)*
(e.g., diethylenetriamine)

$$\log D = \log P_X - \log\left(1 + 10^{+pK_{a3}-p_cH} + 10^{+pK_{a3}+pK_{a2}-2p_cH} + 10^{+pK_{a3}+pK_{a2}+pK_{a1}-3p_cH}\right)$$

---

equations for the cases of mono-, di- and triprotic substances. The equations in Table 1 apply only to cases where a single species partition into the lipid phase. This is seldom realized in real systems. Invariably, a monoprotic substance partitions in two forms: neutral and ion pair, often with a log P difference of 3–4 units. For example, the weak acid ibuprofen has a neutral log P_{HA} 3.97 and an ion-pair log P_A –0.05, in 0.15 M KCl (Fig. 4a). For ibuprofen, case 1a equation in Table 1 correctly represents the lipophilicity profile for pH < 7; above that pH, ion-pairing becomes substantial, but the case 1a equation has no provision to describe it. The weak base propranolol has a neutral log P_B 3.47 and an ion-pair log P_{BH} 0.88, also in 0.15 M KCl (Fig. 4b). Likewise, case 1b equation in Table 1 is only accurate for pH > 8. To describe a monoprotic substance completely, we need an expression that incorporates partitioning of the ion pair. We can proceed in the same way that led us to derive the equations in Table 1 from Eq. (19), to get.

$$\log D = \log\left(P_X + P_{XH}\, 10^{-p_cH + pK_a}\right) - \log\left(1 + 10^{-p_cH + pK_a}\right) \tag{20}$$

This equation is valid for both ibuprofen and propranolol, for the entire pH range. For a weak acid, $P_{XH} > P_X$ and the log D curve decreases with pH; for a weak base, $P_X > P_{XH}$, and the log D curve increases with pH, as the above equation describes.

Fig. 4a shows several examples of lipophilicity profiles for weak acids, with log Ps ranging from 4.0 to –0.4. Fig. 4b shows several examples of weak bases, with log Ps from 5.4 to –1.3. Eq. (20) can be used to represent any one of the curves in Fig. 4.

(a)

(b)
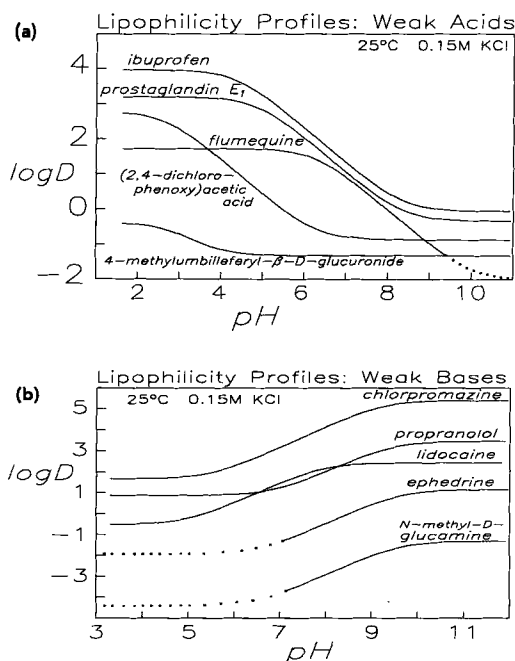
Figure 4. Examples of lipophilicity curves of (a) weak acids and (b) weak bases.

7.3.1 Experimental Evidence for Ion-Pairing; Shake-Flask vs pH-Metric

In the standard shake-flask method log D is measured at several different values of pH. Different buffers are used to control each pH used in the study. (Unfortunately, the ionic strength is not generally controlled with a background electrolyte, such as NaCl.) Usually in a comprehensive study, ten or so such measurements are made (a laborious process) and values of log D are plotted against the pH (usually the operational scale, defined by the buffers used). This is the lipophilicity profile determined directly by the shake-flask method. From this curve, in principle, one can determine log Ps and pK_as.

Consider a "generic" base as an example. Fig. 5a shows a family of plots; all six of the hypothetical substances in Fig. 5a have the same value of pK_a, 10. In the shake-flask method such plots would be used to estimate the neutral and ion-pair log Ps. One plot in Fig. 5a levels off at log D of 2 on the low-pH side; from such a plot one would visually surmise that one was dealing with a weak base possessing a log P_X 5 and log P_{XH} 2, with a Δlog P of 3, a typical difference. One could furthermore estimate the pK_a from such a plot. Note that the shapes of the whole family of curves in Fig. 5a are similar. At low and high pH, the slopes of the curves are near zero. In the intermediate pH, the slopes are approximately 1. Fig. 5b illustrates this; the pH at the mid-point between the slope 0-to-1 transition at the high-pH end corresponds to the pK_a. Furthermore, this is the point where the slope equals 0.5 (except when Δlog P is <2). The horizontal line at dlog$D$/dpH 0.5 in Fig. 5b intersects each of the six first-derivative
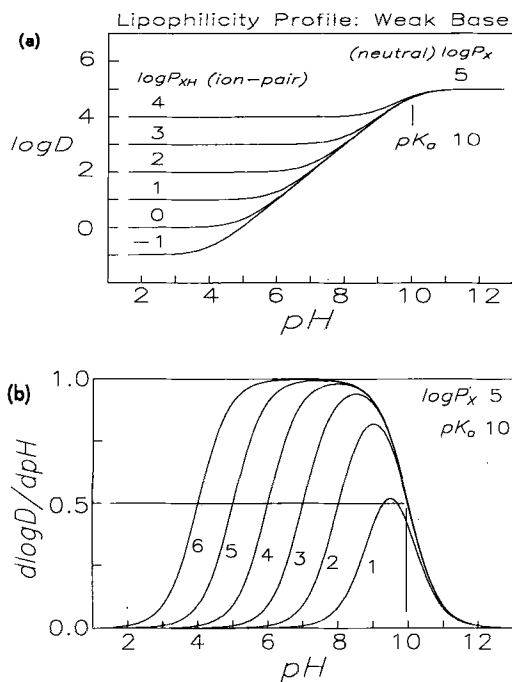
Figure 5. (a) The family of lipophilicity profiles for a series of hypothetical molecules which all have pK_a 10, log P 5 and ion-pair log P from 4 to –1. (b) The corresponding first derivative plots of the curves in Fig. 5a.

curves on the right side at pH 10, the pK_a. (This is not precisely so for the curve corresponding to log P_{XH} 4, where the intersection is at about 9.6).

Most shake-flask analyses, if they get this far, stop at this point. However, there is also an intersection of the curves at slopes 0.5 one the *left* sides in the curves in Fig. 5b, at pHs 9.2, 8, 7, 6, 5, and 4. This is also a pK_a but very unusual kind. It was first noted by Scherrer [23] in one-phase octanol titrations, and it is appropriate to call it the Scherrer-pK_a, or simply $p_oK_a^{SCH}$ (the subscript o is a reminder that the pK_a is observed in an octanol-containing medium). Scherrer made the valuable observation that $\Delta \log P = pK_a - p_oK_a^{SCH}$, thus suggesting a way to measure log P_{ion}.

The above discussion describes the ideal of shake-flask analyses: log D is measured directly, and log P, log P_{ion}, pK_a are derived if measurements are done at several judiciously selected pHs. By contrast, in the pH-metric analysis, one gets log P, log P_{ion}, pK_a from difference curve analysis (followed by mass-balance least squares refinement) as a starting position and *calculates* the log D as the final step, using using Eq. (19).

How does one get log P_{ion} by the pH-metric technique? Eqs. (2) and (4) deal with only a single species. The answer is that it is necessary to do *two* octanol-containing titrations (rather than one, as in the cases associated with Figs. 1–3), selecting a very small octanol/water volume ratio for one (< 0.02) and a very large ratio for the other (> 1). Just as Eq. (20) is a general form of case 1a and 1b equations in Table 1, Eqs. (2a) and (2b) in [5] are the generalized forms of Eqs. (2) and (4), where ion pairing is

(a)

$$\overline{n}_H$$
(bound H^+
per Ibu)

$$-log[H^+]$$

(b)

Ibuprofen Apparent pK_a in Octanol—0.15M KCl

Figure 6. (a) Difference plots for ibuprofen based on octanol/water titrations: (squares) 0.1 mL octanol : 20 ml 0.15 M KCl, (triangles) 0.25 : 20, (hexagons) 1 : 19, (diamonds) 10 : 10, (pentagons) 15 : 5. (b) Calculation illustrating the effect of ion-pair partitioning on the apparent pK_a of ibuprofen in octanol – 0.15 M KCl solutions as a function of the octanol/water volume ratios.

taken into account. Fig. 6a shows not two, but five octanol/water Bjerrum plots for ibuprofen, each at a different octanol/water volume ratio. Eq. (2) predicts that as one adds more octanol to an aqueous solution of a weak acid, the apparent pK_a, p$_o K_a$, increases. So, for 0.1 mL octanol/20 mL 0.15 M KCl, p$_o K_a$ is 6.0, compared with pK_a 4.35; for 15 mL octanol, 5 mL 0.15 M KCl, p$_o K_a$ would be expected to be 8.7. Such would be the prediction of Eq. (2). However, the observed p$_o K_a$ is about 8.1. Ion pairing causes an *attenuation* effect. Fig. 6b shows that Eq. (2) would predict the upper curve, which would correspond to p$_o K_a$ 9.35 at V_{OCT}/V_{water} of 10. The lower curve in Fig. 6b is the one actually observed. At volume ratio of 10, the true p$_o K_a$ is 8.35. In fact, if one measured it in 1000 : 1 volume ratio, it would only be 8.37. This latter value is the *Scherrer* pK_a; that is, in the limit of highest octanol : water volume ratios, p$_o K_a$ = p$_o K_a^{SCH}$. One does not need to do the one-phase octanol titration that Scherrer describes [23,31]. Our experience suggests that pH electrodes do not function well in such a solution.

In the pH-metric technique, one can calculate the lipophilicity profile from the derived ionization and partition constants, using Eq. (19). The upper-most curve in Fig. 4a shows the Scherrer pK_a at pH 8.37, where the slope of the lipophilicity plot is 0.5.

Scherrer did not extend his analysis to multiprotic substances. It was possible to do so with our approach. Fig. 7 is an illustration of the pH-metric relationships for quinine, which is a diprotic base. One distinct limiting p$_o K_a$ in Fig. 7b is 6; the other is below 3, compared with pK_as 8.55 and 4.24 in 0.15 M KCl. The topic of the limiting p$_o K_a$s will be better illustrated when we consider lipophilicity plots of diprotic substances in later sections.
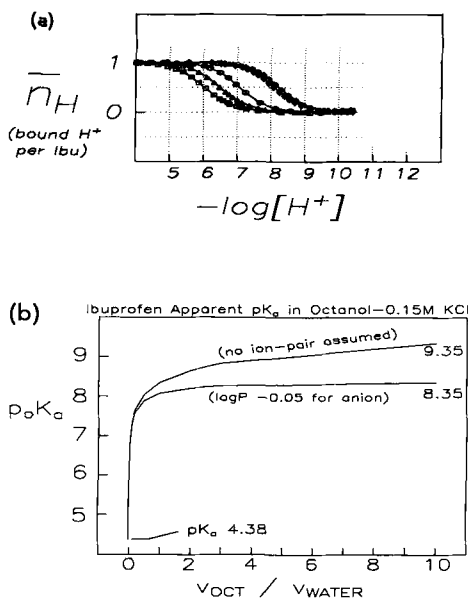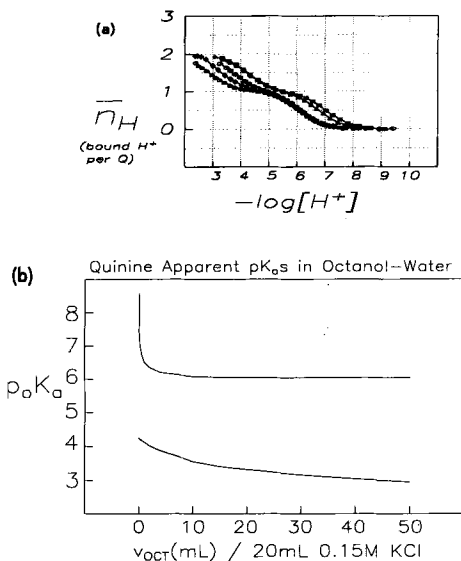
Figure 7. (a) Difference plots for quinine based on octanol/water titrations: (squares) 0.25 mL octanol : 20 mL 0.15 M KCl, (triangles) 0.5 : 20, (hexagons) 5 : 15, (diamonds) 10 : 10, (pentagons) 15 : 5. Note that no curves intersect. (b) Calculation illustrating the effect of ion-pair partitioning on the apparent pK_as of quinine in octanol – 0.15 M KCl solutions as a function of the octanol/water volume ratios.

7.3.2 Further Insights into the Scherrer pK_a

Fig. 8 is a scheme of reactions depicting the most general octanol/water partitioning of a weak acid. In the octanol phase, all ions are treated as ion pairs. It is not as likely to expect an octanol-solvated H^+, free of other ions, as it is to expect to find (H^+, Cl^-) solvated by water-saturated octanol, because the dielectric constant of the wet organic solvent is much lower than that of water. The scheme uses extraction reactions formalism. The octanol : water extraction constants log K_e are –3.12 for KCl and –1.34 for HCl, whose values were carefully determined by conductivity measurements by Westall et al. [32]. We have already defined the K_e, K_a, and P equilibrium expressions associated with the scheme except that of the *ion-pair exchange reaction* associated with the constant K_i. This reaction corresponds to proton exchange in water-saturated octanol between the weak acid HA and the weak base KCl, whose equilibrium quotient is

$$K_i = [K^+A^-]_{OCT}[H^+Cl^-]_{OCT} / [HA]_{OCT}[K^+Cl^-]_{OCT} \tag{21}$$
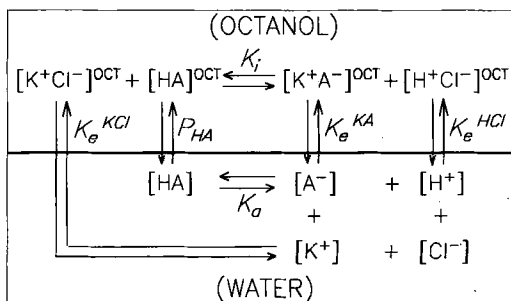
Figure 8. Scheme illustrating the octanol/water partitioning of a weak acid, along with the extraction reactions of various salts.

From the relationships in the scheme of Fig. 8, we can derive the expression

$$pK_i - pK_a = \log P_{HA} - \log K_e^{KA} + \log K_e^{KCl} - \log K_e^{HCl} \tag{22}$$

For example, ibuprofen ($\log K_e^{KA}$ 0.77) has $pK_i = 4.35 + 3.97 - 0.77 - 3.12 - (-1.34) =$ 5.79 in 0.15 M KCl. This number is not one that we can easily have a intuitive feel for. Eq. (22) is not as simple an expression as the one Scherrer derived, since it involves the partitioning of HCl and KCl, in addition to the weak acid HA and the ion-pair KA. We can relate Eq. (22) to that derived by Scherrer. Scherrer did not explicitly consider counter ions. We can safely assume that the concentration $[K^+A^-]_{OCT}$ and $[A^-]_{OCT}$ are the same; since $P_A = [A^-]_{OCT}/[A^-]$, then $K_e^{KA} = P_A/[K^+]$. Substituting this into Eq. (22) produces

$$\{\log [HA]_{OCT} - \log [K^+A^-]_{OCT} - p_cH\} - pK_a = \log P_{HA} - \log P_A \tag{23}$$

The term in braces is the Scherrer pK_a. The advantage of our derivation is that it precisely states what equilibrium corresponds to p_oK^{SCH}, it is

$$(HA)_{OCT} \leftrightharpoons (K^+A^-)_{OCT} + H^+ \tag{24}$$

Scherrer's intuitive assessment was correct: the constant refers to the *aqueous* p_cH at which point the *octanol* concentrations of HA and A$^-$ are equal. It is a mixed constant. This is delightful from an experimental point of view, since pH electrodes work in aqueous media; it has been our experience that they are extremely drifty in wet octanol.

In Scherrer's one-phase titration, he used 0.1 M NaOH as titrant, dissolved in methanol/isopropanol (1 : 4). The pH electrode was effectively working in an isopropanol/methanol medium. We know that such a solvent can shift pH readings by about 0.2–0.5 pH units [5] (K. J. Box, in preparation). So, we believe that the Scherrer pK_a is best determined not in one-phase wet octanol titrations but from the analysis of the lipophilicity profile at slope 0.5 (above discussions), determined from much better behaved dual-phase titrations. When the p_cH is at the Scherrer value, what is the pH in octanol?

7.3.3 pH Scale in Lipids?

We can answer the question from the last section if we define the pH scale in octanol to be based on the concentration of the ion pair $(H^+Cl^-)_{OCT}$. If there were no species other than KCl and HCl partitioning into the octanol phase, the two extraction constants K_e^{HCl} and K_e^{KCl} would fix the relationship between p_cH and $-\log [H^+Cl^-]_{OCT}$. A mass-balance calculation of concentrations, using Westall et al. [32] extraction constants, produces the linear relationship

$$p[H^+Cl^-]_{OCT} = p_cH + 2.18 \tag{25}$$

in the pH interval from 2 to 12 when 0.15 M KCl is considered. At very low pH, the difference is very slightly smaller, due to the common ion (Cl$^-$) effect. At very high pH, the extraction of (K$^+$OH$^-$) into octanol would have to be considered. We do not have reliable constants for that extraction process.

We did a similar calculation in the presence of 0.5 mM ibuprofen and 0.15 M KCl. The difference between the p_cH and $p[H^+Cl^-]_{OCT}$ scales was 2.20, and was also nearly constant in the pH interval from 2 to 12.

We repeated the calculation at 0.03 M KCl, still with 0.5 mM ibuprofen, and found that the difference between the two pH scales was still constant, but the value increased to 2.86, reflecting the role that salt plays in the extraction of hydrogen ion into the lipid phase. The background salt appears to play a key role in maintaining a constant difference between p_cH and $p[H^+Cl^-]_{OCT}$ values across the useful pH range. It appears that the shift factor 2.18 in Eq. (25) is approximately equal to $pK_e^{HCl} + pCl$. It seems that Eq. (25) is worth further study.

Is "what is the pH *inside* of cell membranes?" a fruitful question? Water can diffuse across the lipid layer quite rapidly in phospholipid-based membranes. Even the tight blood-brain barrier is permeable to small molecules like water. The transition between the bulk aqueous environment and the lipid of the biological membrane is not abrupt. The phosphatidylcholine zwitterion and the ester groups of the phospholipid are polar and can accommodate a quasi-aqueous environment for small ions. Hydrogen ions may be found in such a zone, but with reduced activity compared with that found in bulk aqueous solution. Water-saturated octanol has 27 mol % water, hydrogen bonded to octanol molecules in clusters. Perhaps such an environment can model some of the interfacial properties of the biological membranes.

7.3.4 Monoprotic Substance log *D*-pH Curve Shape Analysis

Fig. 5 shows that the lipophilicity curve has a predictable structure. In this section we will further examine this. Fig. 9a is that of a hypothetical weak base, with pK_a 10, log P_X 5 and log P_{XH} 0. It necessarily has $p_oK_a^{SCH}$ 5, from the preceding discussion. The curve in Fig. 9a is broken down into zones of zero slope (1 and 1′), zones of curvature (2 and 2′) and a zone of slope 1 (zone 3). Fig. 9b and 9c show the detail of zones 2 and 2′. If we believe that the precision of log D measurement is 0.01, for what value of slope is the curve 0.01 log D units away from the zero-slope line? This appears to be a slope of 0.022. A similar departure from the unit-limit slope is 1–0.022. These two slopes mark the two departures of the actual curve from the slope 1 and 0 asymptotes. Hence in this way we define the zone of curvature (zone 2 and 2′) to be between point A and B in Fig. 9c and A′ and B′ in Fig. 9b. Point C with slope 0.5 corresponds to Scherrer pK_a and point C′ (which also has slope 0.5) corresponds to the pK_a. The practical upshot of these definitions is that the zone of curvature is 3.30 pH units wide and 1.66 log D units high. This information can be put to use in the next section.

7.3.5 Application of Shape Analysis to One-Point log *D* Shake-Flask Measurement

It is a common practice to measure log D at one buffered pH (often 7.4) and *calculate* the log P of the substance using case 1a equation for weak acids and case 1b equation for weak bases (Table 1). If the pH of the buffer and the pK_a of the molecule differ by
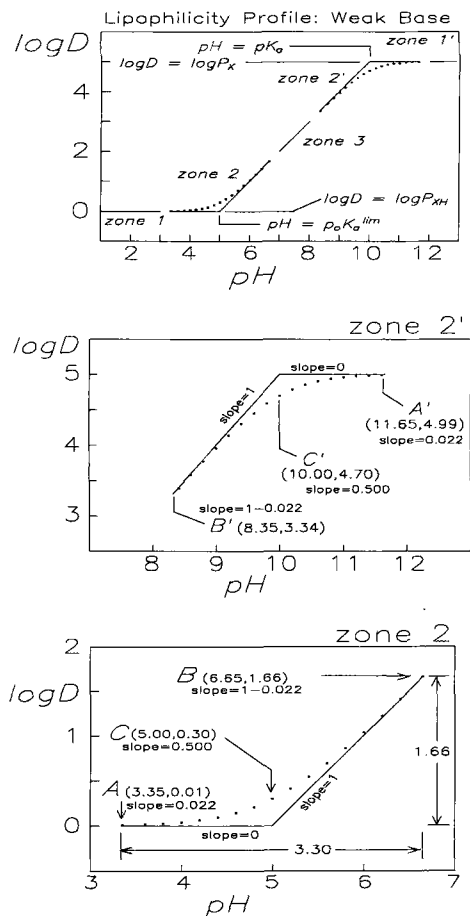
Figure 9. (a) Dissection of the lipophilicity profile for a weak base with a pK_a 10, log P 5 and log P (ion pair) 0 (the "Scherrer" pK_a, $p_oK_a^{lim}$, is 5) into zones of zero-slope (1 and 1'), zones of curvature (2 and 2') and zone of unit-slope (3). (b) Zone 2' close-up view. (c) Zone 2 close-up view.

more than 1–2 units, the calculation can be in error. This is because ion-pairing may be an *unaccounted* contribution in a single log D measurement.

Let us take aprindine as an example. The pK_a of aprindine is•9.95 and log D was measured as –0.09 at pH [33]. The calculated log P_X was reported to be 4.86. How reliable is the "correction"? Let us apply the shape analysis concepts of the last section. If the single log D is in zone 3, then indeed log P_X would have to be 4.86. What boundaries can one put on the value of log P_{XH}? We know that since zone 2 is 1.66 log D units in height, the ion-pair log P_{XH} cannot be greater than –1.75 (–0.09–1.66). This is depicted by curve 1 in Fig. 10. Curve 1' in the figure corresponds to lower values of log P_{XH}. Therefore, the difference between the neutral log P_X and ion-pair log P_{XH} is greater than 6.6. This is unreasonably high.

For simple monoprotic substances, the difference between log $P_{neutral}$ and log P_{ion} is usually 3–4 units [23, 31]. This is substantiated by our own measurements, shown in Table 2 for a variety of drug substances. Bulky, multiprotic molecules with extensive electronic delocalization can have substantially lowered differences. Small peptides
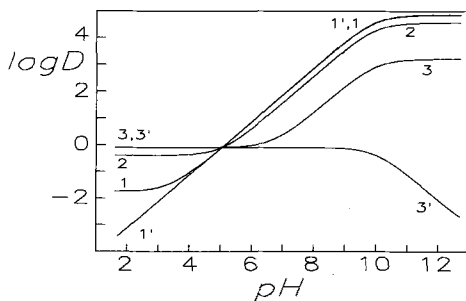
Figure 10. Several possible interpretations of the aprindine lipophilicity profile (see text).

can even have negative differences (that is, log P_{ion} > log $P_{neutral}$), as indicated in Table 2.

Let us test another hypothesis. Let us assume the single log D measurement is from point C in zone 2 (the Scherrer pK_a). If that were so, then log P_X is 4.57 (1.66 – 0.30 –0.09 + log (1 + $10^{9.95-5}$)) and log P_{XH} is –0.39 (–0.30 –0.09). Now the log P difference is 4.96; this is more reasonable. This situation is represented by curve 2 in Fig. 10.

Let us consider the third hypothesis: the log P measurement was made from the zone 1 portion of the lipophilicity curve. This immediately fixes log P_{XH} to –0.09. What can one say about log P_X? It can be no higher than 3.21 (–0.09 + 9.95 –6.65), which is a constraint imposed by curve shape and the pK_a. This maximum log P is depicted by curve 3 in Fig. 10. In principle, the value can be lower, as illustrated by the bounding curve 3' in the figure. Therefore, for this case the log P difference is no greater than 3.30.

The hypotheses testing suggests that a single log D measurement (plus "correction") is not an ideal way to determine log P. Had two points been measured, then in principle, eq. (20) could have determined both the ion-pair and neutral log Ps. Of course, many log D measurements for a given molecule would reliably determine the underlying partition coefficients.

Table 7.2. Differences between log P (neutral) and log P (ion-pair)[a]

Substance	Δlog$P$ (0.15 M KCl)	Δlog$P$ (0.15 M NaCl)	Δlog$P$ (no added salt)	Reference
Buprenorphine	**4.73**			b
Ibuprofen	4.02		4.89	c
Deprenyl	**3.85**			d
Chlorpromazine	**3.73**	**3.75**	**5.04**	c
(2,4-Dichlorphenoxy) *acetic acid*	3.65			d
pF-deprenyl	**3.64**			d
Prostaglandin E$_2$	3.44			[9]
Prostaglandin E$_1$	3.43	3.80	3.95	[9]
Papaverine	**3.17**			d
Lidocaine	**2.96**			d
Niflumic acid		2.65	3.27	[8]

Table 7.2. Continued

Substance	$\Delta\log P$ (0.15 M KCl)	$\Delta\log P$ (0.15 M NaCl)	$\Delta\log P$ (no added salt)	Reference
Quinine	**2.60**		**3.43**	c
Propranolol	**2.59**			c
Hydroxyzine	**2.56**			d
Diacetylmorphine	**2.52**			b
TRP-Phe	2.16			d
Morphine	**2.13**		2.27	b
Buprenorphine	2.13			b
6-Acetylmorphine	1.97			b
Terbutaline	**1.89, 1.97**			d
3-Aminobenzoic acid	**1.27**			d
4-Aminobenzoic acid	**1.26**			d
4-Methylumbilleferyl-β-glucuronide	**0.93**			b
Trp-Trp	0.89			d
Pyridoxine		**0.83**	0.23	c
Tryptophan	0.80			d
Niflumic acid		**0.68**	**1.23**	[8]
Phe-Phe-Phe	0.57			d
Ofloxacin		0.43	0.36	c
Morphine-3β-D-glucuronide	**0.21**			b
Trp-Phe			**0.11**	d
Phenylalanine		**0.03**	**0.00**	c
Aspartic acid			0.01	c
Tryptophan	**–0.22**			d
Phe-Phe-Phe			**–0.31**	d
Phe-Phe	**–0.58**			d
Trp-Trp	**–0.59**			d
Trp-Phe	**–0.61**			d
Aspartic acid		**–0.72**		c
Phe-Phe-Phe	**–0.80**			d

a Bold entries refer to log $P_{neutral}$ – log P_{cation}, other entries refer to log $P_{neutral}$ – log P_{anion}.
b Avdeef, A., Knaggs, R., Barrett, D., Shaw, P. N., Davis S. S. in preparation.
c Avdeef, A., Box, K. J., Takács-Novák, K., in preparation.
d Avdeef, A., Box, K. J., in preparation.

7.3.6 Effect of Salt: Monoprotic Examples

We have measured the partition behaviour of ibuprofen and chlorpromazine at 0.15 M salt and under conditions of no added background salt. Fig. 11a shows that the neutral log P_{XH} of ibuprofen remains unchanged but the ion-pair log P_X depends dramatically on the level of background salt. Similarly, chlorpromazine shows salt dependence, as illustrated in Fig. 11b. Scherrer demonstrated an effect of a similar magnitude with

(a)

IBUPROFEN   25°C; 0.014M KCl, 0.152M KCl

(b)

CHLORPROMAZINE   25°C; 0.001M, 0.151M NaCl, 0.151M KCl
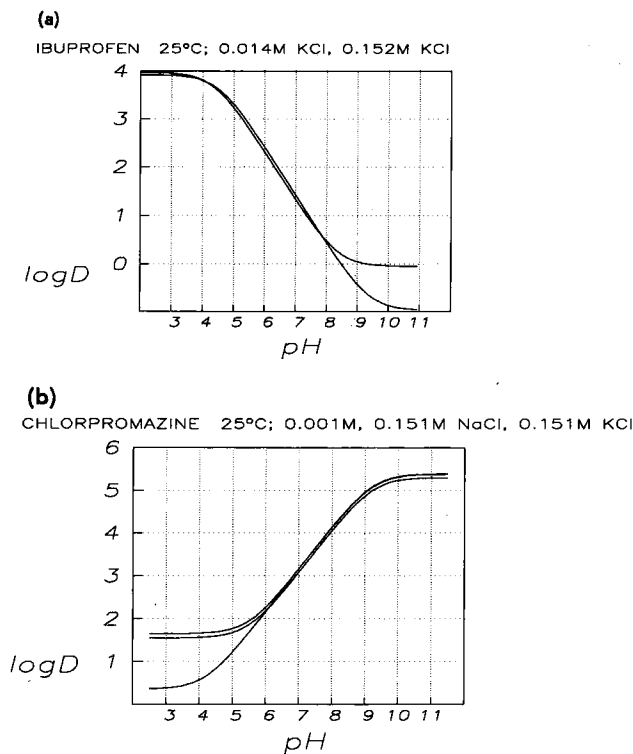
Figure 11. (a) Ibuprofen lipophilicity plot at 0.152 M ionic strength (upper curve) and 0.014 M ionic strength (lower curve). (b) Chlorpromazine lipophilicity plot at 0.15 M KCl (upper-most curve), 0.15 M NaCl (intermediate curve), and 0.001 M NaCl (lowest curve).

propranolol and a few other compounds [23]. Murthy and Zografi [34] studied the effects of many salts on the ion-pair partitioning of chlorpromazine. They established the relative lipophilicity order of the anions:

$$Br^- > CH_3CH_2CH_2SO_3^- > NO_3^- >> Cl^- > CH_3CH_2SO_3^- >> SO_4^{2-}, CH_3SO_3^-$$
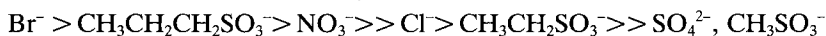
Akamatsu et al. [35] have also observed salt effects on the partitioning of hydrophilic dipeptides, with the order of effect:

$$ClO_4^- >> NO_3^- >> Cl^- > H_2PO_4^-$$
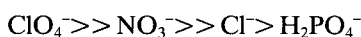
This appears to be a general phenomenon. We will consider this further when we discuss the partition behavior of phenyalanine and tryptophanylphenylalanine.

7.3.7 Debye-Hückel Corrections to Octanol/Water Partition Constants

With the ion-pair *conditional* constant (log P_{ion}) formalism, no opportunity exists to apply the Debye-Hückel theory to correct constants for changes in the ionic strength. When neutral species partition into octanol, there should be no dependence of log P on ionic strength. Largely, this is borne out by experiment, as shown, for example, for ibuprofen and chlorpromazine in Fig. 11.

In the course of a titration, ionic strength is maintained nearly constant by the background salt; still it does deviate from the average value, I_{avg}, at very low or high pH, or when the amount of sample is relatively large and when highly ionized species form. All ions contribute to the ionic strength. A scheme was described by us to take into account the effect on the equilibrium constants due do slightly changing ionic strength [3]. We can apply a similar scheme to extraction reactions.

The last squares refinement calculation [3] converts all equilibrium constants to the $\log\beta$ form; these in turn are used to calculate p_cH, the critical step in the regression analysis. For each p_cH the $\log\beta$ constants are first adjusted for ionic strength deviations from the mean value, according to the expression

$$\log \beta_j(I) = \log \beta_j(I_{avg}) + (a_j \, Q_x^2 + b_j \, Q_y^2 + c_j) \, \{F(I) - F(I_{avg})\} \tag{26}$$

$F(I)$ is the Davies [29] activity coefficient expression,

$$F(I) = -0.5 \left(\frac{\sqrt{I}}{1 + \sqrt{I}} - 0.3 \, I \right) \tag{27}$$

Q_x and Q_y are the charges of the fully deprotonated components in the reactions. (The overall charge of the ion-pair complex extracted into octanol is assumed to be zero.) Let us consider the octanol : water extraction reaction in the binary prostaglandin (Pg)-glucamine (Glu) system.

$$Pg^- + GluH^+ \rightleftharpoons (Pg^-, GluH^+)_{OCT} : K_e = [Pg^-, GluH^+]_{OCT} / [Pg^-] [GluH^+] \tag{28}$$
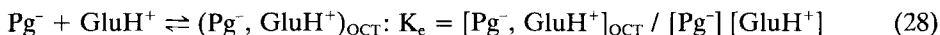
In the example, x may designate Pg ($Q_x = -1$) and y may designate Glu ($Q_y = 0$). The coefficients a_j, b_j, and c_j are the stoichiometric indices for the components Pg, Glu and H, respectively, in the jth equilibrium expression. Consider Eq. (28): $a=1$, $b=1$, and $c=1$. Thus $\log \beta_j(I) = \log \beta_j(I_{avg}) + 2 \, \{F(I) - F(I_{avg})\}$.

Likewise, if we wished to know the theoretically expected value of the extraction constant associated with Eq. (28) at zero ionic strength, we would predict that the extraction constant changes by $-2F$ (evaluated at $I=0.15$ M). The ion-pair constant would thus be expected to *increase* by 0.23.

7.3.8 Diprotic Substance log *D*-pH Curve Shapes (12 Cases)

As noted, lipophilicity equations for cases 2a–2c in Table 1 only represent one-species partitioning. If in addition to the neutral species, an ion pair also partitions, the basic Eq. (19) can be solved for the condition. Fig. 12a–12l represent the 12 resulting cases.

The diprotic molecules we considered earlier, salicylic acid, morphine, and nicotine, are examples of cases (a), (b), and (c), respectively. We will later point out other examples.
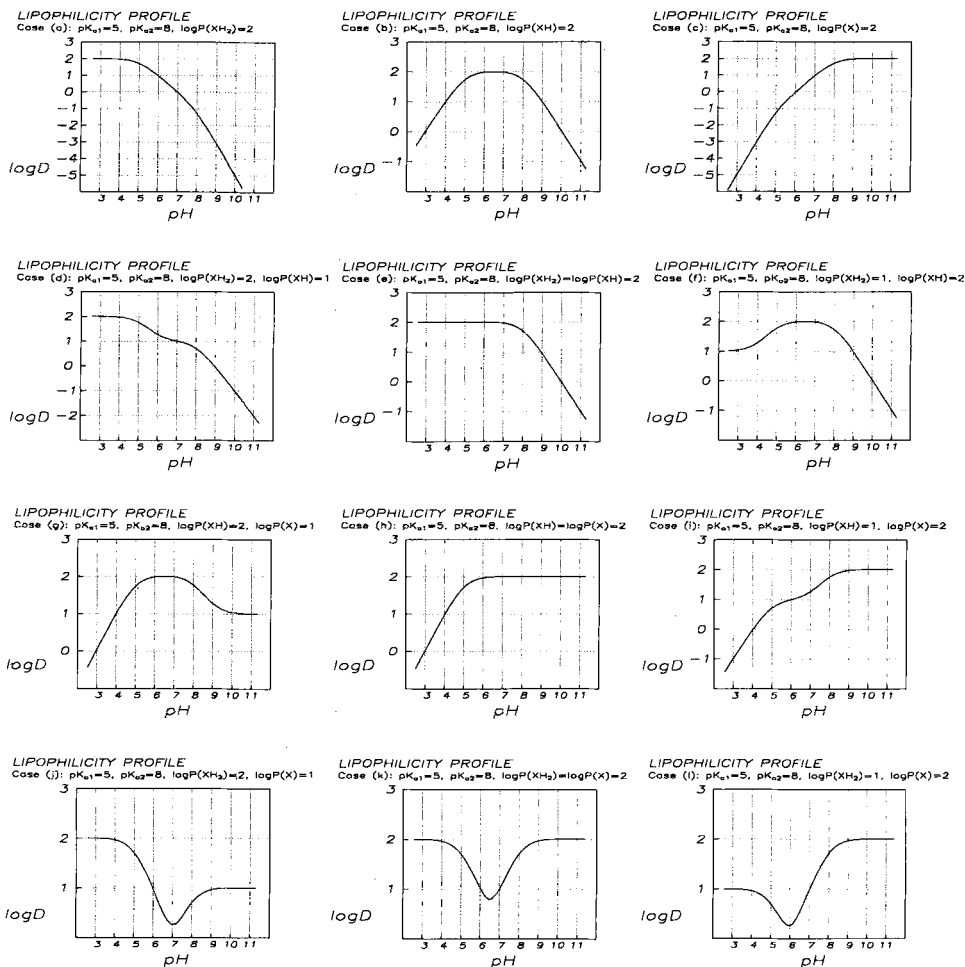
Figure 12. Twelve possible shape types of lipophilicity plots for a diprotic molecule exhibiting two-species lipid partitioning.

7.3.9 Diprotic Molecules with Two Different Ion Pair Partitionings

Niflumic acid partitions into octanol as a neutral species, an anion, and a cation [8]. Solution of Eq. (19) for three-log P case of a diprotic molecule produces the complicated equation

$$\log D = \log (A_0 + A_1 + A_2) \tag{29}$$

where

$$A_0 = \frac{P_X}{1 + 10^{pK_{a2} - pH} + 10^{pK_{a2} + pK_{a1} - 2\,pH}} \tag{30}$$

$$A_1 = \frac{P_{XH}}{1 + 10^{pH - pK_{a2}} + 10^{pK_{a1} - pH}} \tag{31}$$

$$A_2 = \frac{P_{XH_2}}{1 + 10^{pH - pK_{a1}} + 10^{2\,pH - pK_{a2} - pK_{a1}}}$$

(32)

Eq. (29) is illustrated with the examples niflumic acid in Fig. 13a, nitrazepam in Fig. 13b, buprenorphine in Fig. 13c, and quinine in Fig. 13d. Niflumic acid and quinine also illustrate the effect of salt on the partitioning of ions.
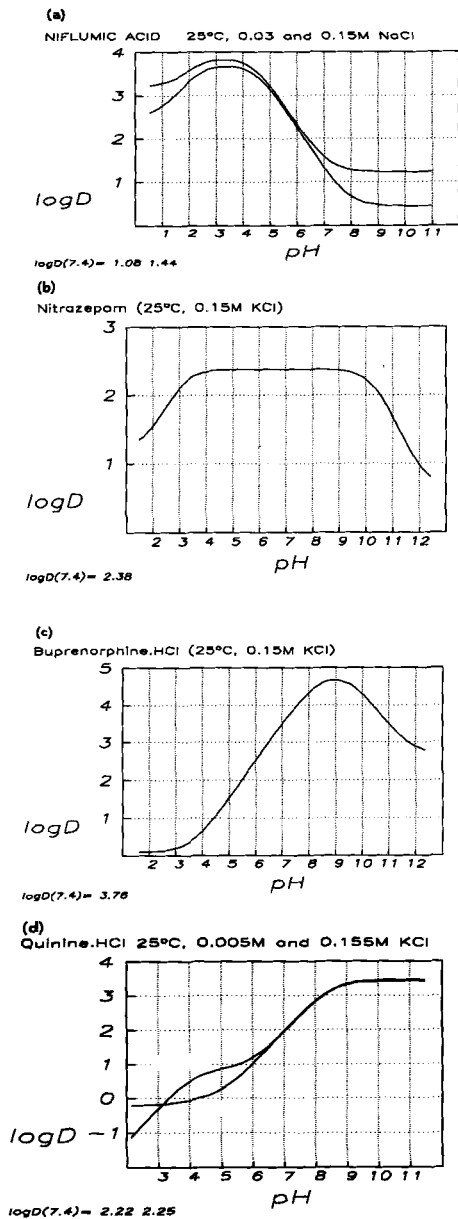
Figure 13. Lipophilicity profiles for diprotic substances illustrating the partitioning of *two* ion pairs along with the neutral species. (a) Niflumic acid lipophilicity plots at 0.15 M NaCl (upper curve) and 0.03 M NaCl (lower curve). The curve levels off below pH 2 because the cation partitions into octanol; the curve levels off for pH > 8 because the anion partitions into octanol. For 0.15 M : log $P(XH_2^+)$ 3.20, log $P(XH)$ 3.88, log $P(X)$ 1.23; for 0.03 M : 2.48, 3.71, 0.44. (b) Nitrazepam cation (log P 1.21), neutral (2.38), and anion (0.64) species octanol/water partitioning. (c) Buprenorphine cation (0.09), neutral (4.82), and anion (2.69) partitioning (d) Quinine lipophilicity profiles. At 0.155 M KCl, log $P(X)$ 3.47, log $P(XH^+)$ 0.87; at 0.005 M KCl, log $P(X)$ 3.43, log $P(XH^+)$ 0.00, log $P(XH_2^{2+})$ –0.21.

7.3.10 Macro-pK_a, Micro-pK_a, and Zwitterions

In certain types of molecule it is possible that chemically (or stereochemically) different species of the same stoichiometric composition are formed. The pH-metric titration technique cannot distinguish between such tautomeric species. In such cases the determined pK_a is a composite constant, a "macroconstant". The thermodynamic experiment is a "proton counting" technique. It cannot identify the site in the molecule the proton comes from. It can only be said that a proton emerges from somewhere in the molecule. On the other hand, "microconstants" are characteristic of individual species, of which there may be more than one with the same composition.

We will use lowercase k for microconstants and uppercase K for macroconstants. (Unfortunately the lower case k is standard notation for rate constants. In the present context there should be no confusion.) Fig. 14 shows microconstant and macroconstant schemes for a diprotic molecule undergoing tautomerization; niflumic acid would be an example of such a molecule [8]. Fig. 14 is labeled with proton *formation* constants K and k, rather than proton *dissociation* constants K_a. The two types of constants are inversely related: $\log K_1 = pK_{a2}$ and $\log K_2 = pK_{a1}$. The use of formation constants simplifies the form of the derived relationships between micro- and macroconstants.

The tautomeric constant is defined as

$$k_Z = [XH^{\pm}] / [XH^0]$$

$$= k_1^{\pm} / k_1^{0} = k_2^{0} / k_2^{\pm} \tag{33}$$

where the microconstants are defined in terms of the quotients

$$k_1^{\pm} = [XH^{\pm}] / x\, h \tag{34}$$

$$k_2^{\pm} = [XH_2^{+}] / [XH^{\pm}]h \tag{35}$$

$$k_1^{0} = [XH^0] / x\, h \tag{36}$$

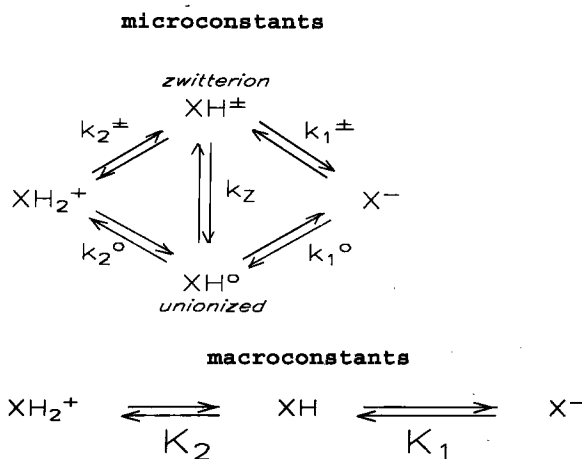$$k_2^{0} = [XH_2^{+}] / [XH^0]h \tag{37}$$

microconstants

macroconstants

Figure 14. Microconstant ionization scheme.

From the mass balance expression

$$[XH] = [XH]^{\pm} + [XH^0] \tag{38}$$

and Eqs. (34–37) one can derive the relationships between the microconstants and the corresponding macroconstants.

$$K_1 = k_1^{\pm} + k_1^0 \tag{39}$$

$$K_2^{-1} = (k_2^{\pm})^{-1} + (k_2^0)^{-1} \tag{40}$$

The stability constants, Eq. (12) and (13), are related to the above constants as $\beta_{11} = K_1$ and $\beta_{12} = K_1 K_2 = k_1^0 k_2^0 = k_1^{\pm} k_2^{\pm}$. In the next section these relationships are needed to convert Eq. (19) into a microconstant basis.

7.3.11 Relationship between Micro-log p, Macro-log P, and log D

The distribution coefficient, D, is the same whether the species are described with microconstants or macroconstants. However, *two types of log P are possible*. Eq. (19) was derived with macroconstants. The second term in the numerator and denominator of Eq. (19) can be modified to utilize microconstants, drawing on Eqs. (34–40). From the substitutions, one can derive the following useful relation between micro- and macro-log P. We will use lowercase p to denote the micro partition constant.

$$P_{11} = p_{11}^0 (k_1^0/K_1) + p_{11}^{\pm} (k_1^{\pm}/K_1) \tag{41}$$

where

$$p_{11}^0 = [XH^0]_{ORG}/[XH^0] \tag{42}$$

$$p_{11}^{\pm} = [XH^{\pm}]_{ORG}/[XH^{\pm}] \tag{43}$$

The micro log p_{11}^0 refers to the partitioning of just the unionized species between water and octanol. The micro log $p_{11}^{\pm}$ refers to just the zwitterion partitioning. Fig. 15 illustrates the distinction between the two simultaneously occurring processes. If both the unionized microspecies and the zwitterion partition into the organic phase, Eq. (19) reduces to

$$D = \frac{p_{11}^0 h\ k_1^0 + p_{11}^{\pm} h\ k_1^{\pm}}{1 + h\ k_1^0 + h\ k_1^{\pm} + h^2 k_1^0 k_2^0} \tag{44}$$

If $p_{11}^0 >> p_{11}^{\pm}$ the equation can be further rearranged to

$$\log p_{11}^0 \approx \delta + \log D + \log \left(1 + k_Z + \frac{1}{h\ k_1^0} + h\ k_2^0\right) \tag{45}$$

where $\delta = k_Z\ p_{11}^{\pm}/(2.303\ p_{11}^0)$. If the zwitterion does not partition into the organic phase ($\delta = 0$), then Eq. (45) reduces to Eq. (9) reported by Takács-Novák et al. [36].
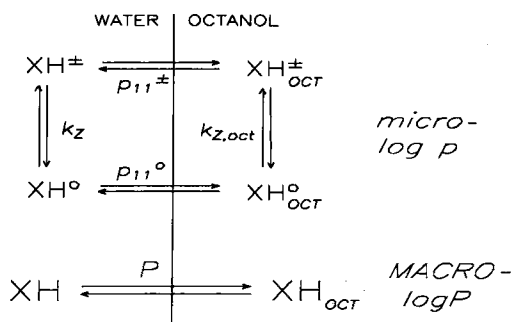
Figure 15. Microconstant log p scheme.

7.3.12 Micro-log p Application

Niflumic acid was studied both pH-metrically and spectroscopically using the shake-flask method [8]. The macroscopic log P was measured as 3.88 in 0.15 M NaCl. It was assumed that a negligible amount of zwitterion $XH^{\pm}$ partitions into octanol. Because there is an equilibrium between the two micro-forms of the XH species in the aqueous phase (tautomeric ratio, Eq. (33), was measured spectroscopically to be 17.4), the macro-log P value differs from the micro-log p value according to Eq. (41): $\log p_{XH_o} = \log P_{XH} + \log K_1 - \log k_1^0 = 5.1$. The value determined by shake-flask measurements was 4.8, indicating good agreement. The number is quite a bit higher than the macro log P.

7.3.13 Partitioning of the Amino Acids Phenylalanine and Tryptophanylphenylalanine

The pH-metric technique has a practical low log P limit of detection that is approximately –2. For hydrophilic molecules with log P of –2, the octanol-containing and the aqueous titration curves differ in the buffer region by about 0.01 pH units. This is about the level of reproducibility of a good research-grade pH electrode. One would not expect the technique to work with amino acids; most amino acids are very hydrophilic and are charged in the interval pH 2–12. Even when the molecule is uncharged in the overall sense, it still exists in the zwitterionic form, with a plus-charged center and a minus-charged center. Since we were able to determine ion-pair partitioning in many other substances, with log P often in the range between 0 and –2 and were able to demonstrate the dependence of the phenomenon on background salt concentration, we thought that it would be productive to try measuring log P of some amino acids. What we immediately saw was met with some scepticism: although we were sometimes unable to characterize the zwitterion log P, we often saw convincing indication of the monocation partitioning at low pH and the monoanion partitioning at high pH, producing an "inverted" parabola log D curve, with the *minimum* at the isoelectric point. This was different from what was observed with the shake-flask method, employing pH buffers. Our solutions contained 0.15 M NaCl or KCl, and it appeared reasonable to us that ion pair partitioning was taking place.

Fig. 16a shows the lipophilicity profile of phenylalanine in 0.027 and 0.161 M NaCl. Also shown in the plot are measured values of log D using radiolabeled phenylalanine (data kindly provided by Dr Keith Chamberlain, 1994, Rothamsted Experimental Station, UK). The agreement is superb over much of the pH range. The measured log P of the zwitterion was –1.4. We were also able to see evidence for the partitioning of the amino acid cation in low-pH solution. Fig. 16a shows slight salt dependence, most prominent at pH 2.

Fig. 16b shows the lipophilicity profile of tryptophanylphenylalanine (Trp-Phe), obtained from pH-metric data in 0.15 M KCl. Lipophilicity profiles of amino acids determined by the shake-flask method, using phosphate buffers, often show a parabolic shape with a *maximum* log D at the isoelectric point (pH 5.24). The plot in Fig. 16b is controversial because it reveals as *smaller* log$D$ at the isoelectric point than at lower pH, where the cationic species predominates. It appears that the ion-pair partitioning is more prevalent than zwitterion partitioning.

Precedence for the "inverted" behavior exists in the literature, though it appears not to have received the attention we think it deserves. Fig. 16c displays the lipophilicity
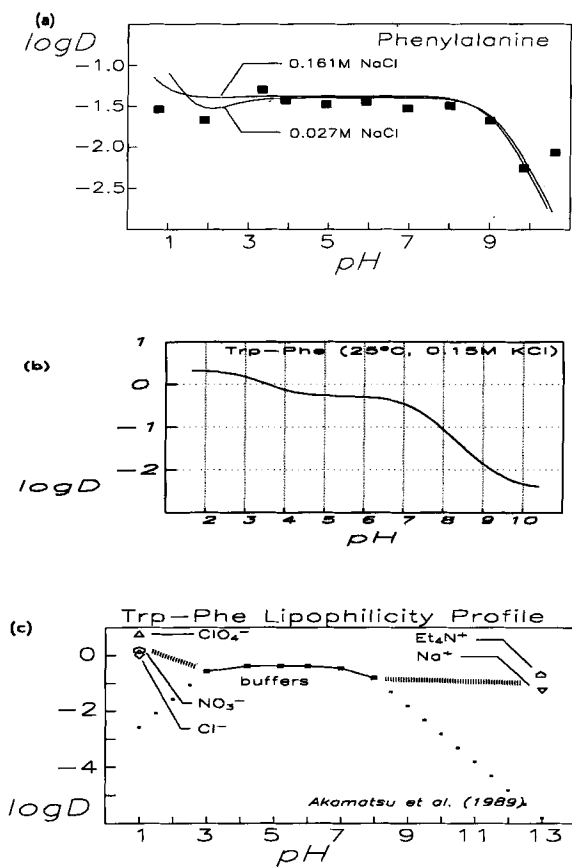

Figure 16. Lipophilicity profiles of amino acids. (a) Phenylalanine at two different ionic strengths and with radiolabelled shake-flask measured log D values (squares). See text. (b) Tryptophanylphenylalanine. See text (c) Tryptophanylphenylalanine using data of Akamatsu et al. [35]. See text.

plot of Trp-Phe constructed from the data of Akamatsu et al. [35]. Using conventional buffers they saw a maximum log D at the isoelectric point. However, when they attempted to measure log D in low pH solutions in the presence of chloride and other anions more lipophilic than phosphate, they observed log D values higher than that found at the isoelectric point. Fig. 16c shows such points at pH 1 and 13. It would seem that salt has a key influence on the lipophilicity of charged species such as amino acids. Our measured lipophilicity profiles of hydrophilic amino acids (e. g., glycine, aspartic acid, tyrosine) in the 0.15 M NaCl or KCl medium often look like cases (j), (k), or (l) in Fig. 12.

7.3.14 Partitioning of Morphine Derivatives and Metabolites

We would like to briefly present a summary of an extensive study of the partitioning behavior of morphine and a series of morphine derivatives, including 3- and 6-glucuronides (M3G and M6G). Details of the study will soon appear elsewhere (S. S. Davis et al., in preparation). Fig. 17 illustrates the lipophilicity curves of the family of compounds we examined. It is interesting to see how the ranking of lipophilicities varies with pH. At pH 7.4, the molecules have the following lipophilicity ranking, with $D_{7.4}$ values in parentheses: buprenorphine (5600) $\gg$ diacetylmorphine (7.1) > 6-acetylmorphine (4.1) $>$ codeine (1.7) $>$ morphine (0.9) $>$ morphine-6β-D-glucuronide (0.2) $>$ morphine-3β-D-glucuronide (0.08) $>$ norcodeine (0.06) $>$ normorphine (0.03). We were able to characterize the partitioning of the zwitterionic form of M3G and M6G under a variety of conditions. It is interesting and unexpected that M3G has virtually identical lipophilicity in octanol, chloroform and PGDP. This may be the consequence of the acid and base properties of the M3G zwitterion molecule offsetting proton donor/acceptor properties of the three partition solvents. The lipophilicity curve clearly indicates enhanced partitioning when the background KCl concentration was raised from 0.15 M to 0.5 M: the log P of the zwitterion increased from -1.06 to -0.88 with the increased salt. This is about the magnitude to effect expected if the zwitterion partitioned as an ion triplet, $(K^+M3GH^\pm Cl^-)_{OCT}$. When the charge centers are close to each other, as in phenylalanine, the log P of the zwitterion does not appear to show dependence on ionic strength (cf. Fig. 16a). However, in the M3G zwitterion, the two charge centers are far apart; this may be the reason we see ionic
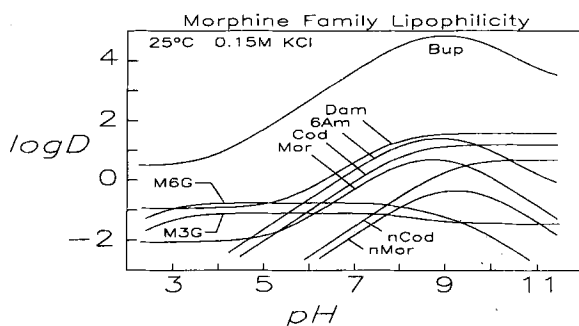

Figure 17. Morphine and derivatives profiles. Bup, buprenorphine; Dam, diacetylmorphine; 6 Am, 6-acetylmorphine; Cod, codeine; Mor, morphine; M3G, morphine-3β-D-glucuronide; M6G, morphine-6β-glucuronide; nCod, norcodeine; nMor, normorphine.
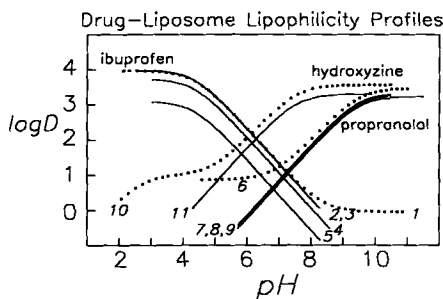
Figure 18. Drug-multilamellar liposome lipophilicity profiles. The dotted curves are octanol/water profiles for the three drug ibuprofen (curve 1), propranolol (curve 6), and hydroxyzine (curve 10). The solid curves are the calculated drug-liposome profiles: (2) ibuprofen-DOPC; (3) ibuprofen-eggPC; (4) ibuprofen-eggPC/0.3 cholesterol; (5) ibuprofen-eggPC/0.6 cholesterol; (7) propranolol-eggPC/0.3 cholesterol; (8) propranolol-eggPC; (9) propranolol-DOPC; (11) hydroxyzine-DOPC.

strength dependence. The high-pH region shows the most dramatic increase, presumably with the partitioning of $(K^+M3G^-)_{OCT}$.

7.3.15 Drug-Liposome Partitioning, First Look

Fig. 18 shows lipophilicity plots for three drug substances: ibuprofen, propranolol, and hydroxyzine, in octanol/water (dotted curves) and multilamellar liposome/water (solid curves). The research is jointly pursued with Leo Herbette and his coworkers and will be the subject of a future publication. Noteworthy is the effect of cholesterol on the partitioning of ibuprofen into the liposomes, as shown in the figure.

7.4 Outlook

We have described in a very comprehensive way the experimental and mathematical characteristics of the lipophilicity profile. The benefits of the pH-metric technique for the thorough assessment of the lipophilicities of a large variety of molecules are substantial. In an hour or so a whole lipophilicity profile for a drug substance can be determined, often with very high precision. We have reliably measured pK_as as high as 13.3 and as low as 0.6, log P as high as 7.4 and as low as -2.3. The span in log P is nearly ten orders of magnitude! For example, the complete lipophilicity profile of chlorpromazine, which has a log P 5.40, can be determined reliably in about 40 minutes. Factors limiting the measurement of log $P > 7$ include (a) extreme water insolubility, (b) poorly reproducible micelle or other type of aggregation reactions, and (c) surface activity of the substance. Hydrophilic molecules are very difficult to characterize by the technique if their log P is less than -2, the approximate limit of detection. The log P region between -1.5 and -2 is difficult for the pH-metric technique and care is required in the design of the assay.

The subtle interplay between the lipophilicity and pH depends critically on the pK_as and log Ps of the molecule, effects that cannot be broadly understood from partial -characterizations of the lipophilicity. Single log D measurements can be especially limiting.

The physiological medium is loaded with salt, and the effect of salt on the ion-pair partitioning of drug substances may be important to understand. The prospect of being able to quickly and reliably characterize drug-liposome partitioning is particularly exciting.

Acknowledgements

References

[1] Dyrssen, D., *Sv. Kem. Tidsks.* **64**, 213–224 (1952)
[2] Avdeef, A., Fast simultaneous determination of log p and pK_a by potentiometry: para-alkoxyphenol series (methoxy to pentoxy). In: *QSAR: Rational Approaches to the Design of Bioactive Compounds.* Silipo, C., and Vittoria, A. (Eds.). Elsevier: Amsterdam; 119–122 (1991)
[3] Avdeef, A., *Quant. Struct.-Act. Relat.* **11**, 510–517 (1992)
[4] Avdeef, A., *J. Pharm. Sci.* **82**, 183–190 (1993)
[5] Avdeef, A., Comer, J. E. A., and Thomson, S. J., *Anal. Chem.* **65**, 42–49 (1993)
[6] Avdeef, A., Comer, J. E. A., Measurement of pK_a and log P of water-insoluble substances by potentiometric titration. In: *QSAR and Molecular Modelling*, Wermuth, C. G. (Ed.). Escom: Leiden; 386–387 (1993)
[7] Slater, B., McCormack, A., Avdeef, A., and Comer, J. E. A., *J. Pharm. Sci.* **83**, 1280–1283 (1994)
[8] Takács-Novák, K., Avdeef, A., Box, K. J., Podányi, B., and Szász, G., *J. Pharmaceut. Biomed. Anal.* **12**, 1369–1377 (1994)
[9] Avdeef, A., Box, K. J., and Takács-Novák, K., *J. Pharm. Sci.* **84**, 523–529 (1995)
[10] Shim, C. K., Nishigaki, R., Iga, T., and Hanano, M., *Int. J. Pharm. acent.* **8**, 143–151 (1981)
[11] Dyrssen, D., *Acta Chem. Scand.* **8**, 1394–1397 (1954)
[12] Dyrssen, D., Dyrssen, M., and Johannson, E., *Acta Chem. Scand.* **10**, 341–352 (1956)
[13] Dyrssen, D., *Acta Chem. Scand.* **10**, 353–359 (1957)
[14] Dyrssen, D., *Acta Chem. Scand.* U11, 1771–1786 (1957)
[15] Rydberg, J., *Sv. Kem. Tidskr.* **65**, 37–43 (1953)
[16] Brändström, A., *Acta Chem. Scand.* **17**, 1218–1224 (1963)
[17] Seiler, P., *Eur. J. Med. Chem.-Chimica Therapeutica* **9**, 665–666 (1974)
[18] Kaufman, J. J., Semo, N. M., and Koski, W. S., *J. Med. Chem.* **18**, 647–655 (1975)
[19] Briggs, T. N., and Stuehr, J. E., *Anal. Chem.* **46**, 1517–1521 (1974)
[20] Avdeef, A., and Bucher, J. J., *Anal. Chem.* **50**, 2137–2142 (1978)
[21] Avdeef, A., Kearney, D. L., Brown, J. A., and Chemotti, A. R., Jr, *Anal. Chem.* **54**, 2322–2326 (1982)
[22] Avdeef, A., *Anal. Chim. Acta* **148**, 237–244 (1983)

[23] Scherrer, R. A. The treatment of ionizable compunds in quantitative structure-activity studies with special consideration to ion partitioning. In: *Pesticide Synthesis Through Rational Approaches*, American Chemical Society Symposium Series 255. Magee, P. S., Kohn, G. K., and Menn, J. J. (Eds.). American Chemical Society: Washington, DC; 225–246 (1984)

[24] Clarke, F. H., *J. Pharm. Sci.* **73**, 226–230 (1984)

[25] Avdeef, A., STBLTY: Methods for Construction and Refinement of Equilibrium Models. In: *Computational Methods for the Determination of Formation Constants*. Leggett, D. J. (Ed.). Plenum: New York; 355–473 (1985)

[26] Clarke, F. H., and Cahoon, N. M., *J. Pharm. Sci.* **76**, 611–620 (1987)

[27] Bjerrum, J., *Metal-Ammine Formation in Aqueous Solution*. Haase: Copenhagen 1941

[28] Irving, H. M., and Rossotti, H. S., *J. Chem. Soc.* 2904–2910 (1954)

[29] Davies, C. W., *Ion Association*. Butterworths: London 1962

[30] Scherrer, R. A., and Howard, S. M., *J. Med. Chem.* **20**, 53–58 (1977)

[31] Scherrer, R. A., Crooks, S. L. Titrations in water-saturated octanol: a guide to partition coefficients of ion pairs and receptor-site interactions. In: *QSAR. Quantitative Structure-Activity Relationships in Drug Design*, Fauchère, J.-L. (Ed.). Alan R. Liss; 59–62 (1989)

[32] Westall, J. C., Johnson, C. A., and Zhang, W., *Environ. Sci. Technol.* **24**, 1803–1810 (1990)

[33] Mannhold, R., Dross, K. P., and Rekker, R. F., *Quant. Struct.-Act. Relat.* **9**, 21–28 (1990)

[34] Murthy, K. S., and Zografi, G., *J. Pharm. Sci.* **59**, 1281–1285 (1970)

[35] Akamatsu, M., Yoshida, Y., Nakamura, H., Asao, M., Iwamura, H., and Fujita, T., *Quant. Struct.-Act. Relat.* **8**, 195–203 (1989)

[36] Takács-Novák, K., Józan, M., Hermecz, I., and Szász, G., *Int. J. Pharm. acent.* **79**, 89–96 (1992)

Appendix

Molecules Mentioned in Review
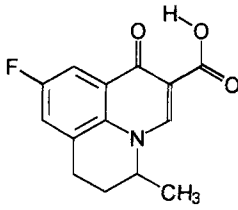
flumequine
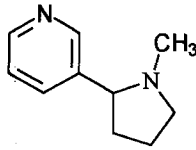
nicotine
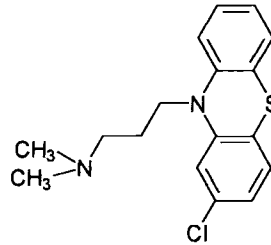
chlorpromazine
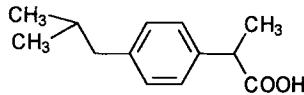
ibuprofen
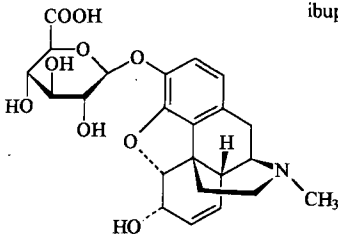
morphine-3β-D-glucuronide (M3G)
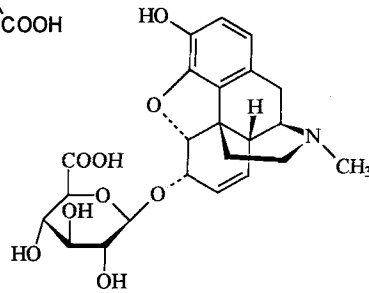
morphine-6β-D-glucuronide (M6G)
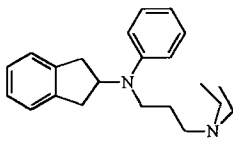
aprindine
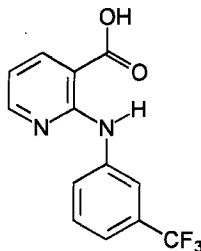
niflumic acid
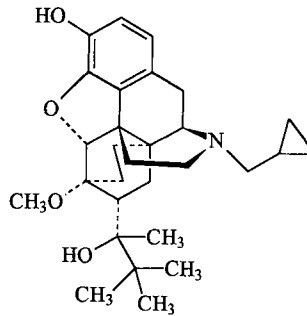
buprenorphine
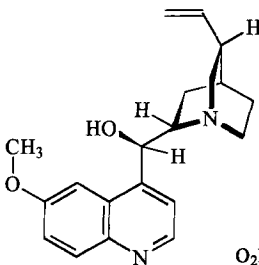
quinine
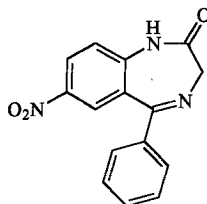
nitrazepam
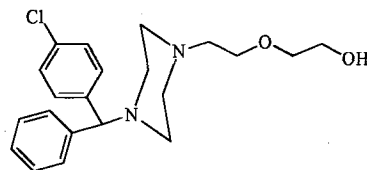
hydroxyzine

8 Estimation of Lipophilicity by Reversed-Phase Thin-Layer Chromatography

Raimund Mannhold, Karl Dross and Christoph Sonntag

Abbreviations

ASCLOGP	Conformation-dependent log P, based on approximate surface calculation
KLOGP	Calculated log P values, based on fragmental contributions
CHEMICALC-2	Computer program to calculate log P, based on atomic contributions
CLOGP	Calculated log P values, based on fragmental contributions
HINT	Computer program to calculate log P, based on atomic contributions
HPLC	High-performance liquid chromatography
ODS	Octadecylsilane
QSAR	Quantitative structure-activity relationships
RP-HPLC	Reversed-phase high-performance liquid chromatography
RP-TLC	Reversed-phase thin-layer chromatography
Σf	Calculated log P values, based on fragmental contributions
SMILOGP	Computer program to calculate log P, based on atomic contributions
TLC	Thin-layer chromatography

Symbols

$\alpha \times \beta$	Product of hydrogen-bond donor and acceptor capacity
β	Hydrogen-bond acceptor capacity
R_F	Retention factor in TLC
R_M	Linear physico-chemical descriptor in TLC
R_{Mw}	R_M-values, extrapolated to modifier-free conditions
$\log k'$	Logarithm of the isocratic capacity factor
$\log k_w$	Logarithm of the polycratic (extrapolated) capacity factor
$\log P_{oct}$	1-octanol/water partition coefficient
φ	Fraction of the modifier in binary solvent mixtures
ψ	Fraction of water in binary solvent mixtures

8.1 Introduction

Lipophilicity is the one physico-chemical parameter that continually attracts prime interest in QSAR studies, as befits its role as a predominant descriptor of pharmacodynamic, pharmacokinetic and toxic aspects of drug activity. Numerous monographs and

reviews treat this topic in adequate detail [1–6]. The partition coefficient P between water (or buffer) and 1-octanol was for a long time used as the preferential experimental expression of the lipophilic properties of a compound. 1-Octanol/water partitioning is losing this role as the method of choice due to methodological drawbacks and limitations, which have been extensively discussed in the literature [7].

Chromatographic approaches (HPLC and TLC) are very important experimental alternatives to 1-octanol/water partitioning. Retention of solutes in chromatography is mainly governed by adsorption and partitioning processes. In order to derive lipophilicity descriptors by chromatographic approaches it is mandatory to limit the influence of adsorption on retention. Adaptation of chromatographic procedures to the experimental determination of lipophilic properties resulted in the development of reversed-phase chromatography, in which the commonly hydrophilic, polar stationary phase is replaced by a hydrophobic, nonpolar phase.

The use of chromatographic approaches in the QSAR field goes back to the early work of Martin and Synge [8] as well as Consden et al. [9], who established relationships between the R_F values obtained from partition chromatography and partition coefficients.

The R_F value of a compound x is defined as the ratio of the migration distance of the solute x $(Z_x–Z_0)$ to that of the mobile phase $(Z_F–Z_0)$:

$$R_F = \frac{Z_X - Z_0}{Z_F - Z_0} \tag{1}$$

As linear correlate between chromatographical behavior and chemical structure Bate-Smith and Westall [10] introduced the R_M value:

$$R_M = \log \frac{1 - R_F}{R_F} \tag{2}$$

Combining Eqs. 1 and 2 yields:

$$R_M = \log \frac{(Z_F - Z_0) - (Z_X - Z_0)}{Z_X - Z_0} = \log \frac{Z_F - Z_X}{Z_X - Z_0} \tag{3}$$

Thus, R_M represents the logarithmic ratio of the distances between solute spot and solvent front (i.e., a measure of the solute interaction with the lipophilic, stationary phase) on the one hand and the migration distance of the solute (i.e., a measure of the migration of the solute with the hydrophilic, mobile phase) on the other hand.

The less polar a solute x, the stronger will be its interaction with the stationary phase, which is expressed by decreasing R_F values and increasing R_M values. According to Eq. (2), R_M values below 0.5 give positive R_M values and vice versa. Thus, R_M represents the direct correlate of the lipophilicity of a solute, provided its estimation is based on pure or at least preferential partition chromatography. Correct R_M values can only be obtained on the basis of precisely measured R_F values, which is achieved under two experimental prerequisites. First, starting and running points have to be evaluated densitometrically and not by hand. Second, for the determination of the true front, which is not identical with the visible front, front markers have to be used (for details, see [11]).

Biagi and coworkers were the first to use reversed-phase TLC for lipophilicity measurements systematically and to develop them further. The introduction of the R_{Mw} value, i.e., R_M extrapolated to modifier-free conditions, is the merit of this group. According to current view, only the R_{Mw} values exhibit significant comparability with log P. Biagi et al. [12–14] were able to show such interrelations for several pharmacological classes. The comprehensive work of this group is reviewed in several recent papers [15–17].

Physico-chemical conditions of reversed-phase TLC as well as its relevance for QSAR studies have been thoroughly discussed in an excellent review by Tomlinson [18]. He demonstrated the additive, constitutive character of R_M and the analogy of ΔR_M to π as a fragmental constant for lipophilic group contributions. In addition, the impact of steric and electronic effects on R_M was discussed in detail.

In 1965 Boyce and Milborrow [19] published the first QSAR paper applying R_M values as lipophilicity parameters in correlations with biological activity. In addition, the linear increase of R_M with alkyl chain length of congeneric compounds impressively demonstrated the applicability of R_M as a substitute of log P.

In the following sections we want to provide the reader with the theoretical and methodological background enabling the practitioner to derive precise and reproducible TLC data. Given up-to-date technology (e.g., densitometry) TLC represents an attractive experimental alternative to both HPLC and octanol/water partitioning.

8.2 Stationary Phase

As mentioned above, in reversed-phase chromatography – i. e., the chromatographic approach to deriving lipophilicity descriptors – hydrophobic, nonpolar material is used as the stationary phase. In some cases [20, 21] cellulose plates were used. According to Geiss [22] this material has no reversed-phase properties. Silica-gel plates, impregnated by hand with paraffin, silicone or other oils deserve mention here; in addition, commercially available octadecyl silylated silica-gel (ODS) materials are employed. Some typical examples of stationary phases, being far from complete, are summarized in Table 1.

The major part of TLC work has been done by using silica-gel plates, impregnated by hand either with silicone oil (in particular the group of Biagi and coworkers) or with paraffin. Such plates can be moistened with pure water, which explains their common use in lipophilicity determinations by TLC.

For the sake of reproducibility and comparability between laboratories, however, the use of commercially available plates is recommended. Nevertheless, the latter material can also suffer from some drawbacks. Impregnated silica-gel 60 plates from Merck, e. g., seem to decompose, resulting in irregularities of the running of the plates. In addition, this plate material cannot be moistened with solvent mixtures containing more than 60 % water. In the case of all impregnated plates comparability of the TLC data to log P_{oct} values is rather limited (for a detailed comment see section 8. 6.2). HPTLC plates, coated with highly C18 etherified silica-gel, can only be moistened with solvents containing at most 40 % water [23–25]. In our opinion, an extrapolation of TLC data to modifier-free conditions is mandatory (see section 8. 4); small unavoidable

Table 8.1.   Typical examples of stationary phases, used in RP-TLC

Stationary phase	References
Kieselguhr/oleyl alcohol	[51, 68]
Silanized Kieselguhr	[52]
Silica gel impregnated with liquid paraffin	[19, 23, 33, 36, 53, 54]
Silica gel impregnated with silicone oil	[14–17, 41, 55–57]
Silica gel impregnated with octanol	[55, 58]
Silica gel impregnated with tricaprylmethylammonium	[33, 35]
Silanized precoated silica gel Merck	[41, 42]
RP8 HPTLC Merck	[25, 58]
RP18 HPTLC Merck	[23, 25, 58]
Whatman KC 18 F	[25, 32, 60–63]
RP2 TLC Merck	[25, 37, 64]
RP8 TLC Merck	[25, 37, 64]
RP18 TLC Merck	[25, 32, 33, 37, 64–66]

errors in estimating R_M at low water contents of the solvent will result in unacceptable errors in extrapolating to R_{Mw}.

In recent years many investigators prefer the use of ODS-coated plate material. The silica-gel material of RP 18 plates suffers from the low etherification (22 %) of the silica-gel OH-groups [26]. Thus, in the case of solvents with low water contents, silano-philic forces become prevalent, resulting in relatively too high R_M (Fig. 1).

Taken together, the use of RP 18 TLC plates demands a high accuracy in extrapolating to R_{Mw}, which we discuss in detail in section 8. 4.
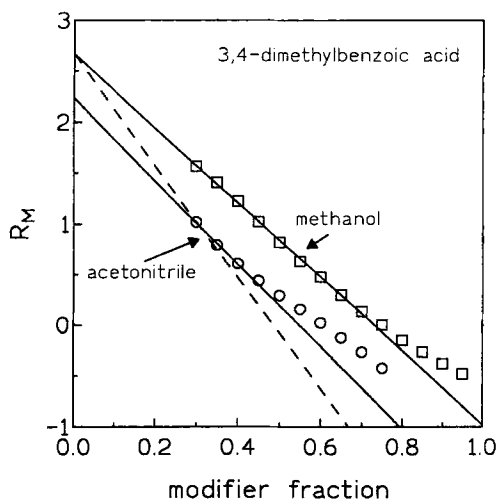

Figure 1. R_M values of 3,4-dimethoxy-benzoic acid as a function of the modifier fraction of methanol (□) or acetonitrile (o). Solid lines represent the extrapolation lines. The dashed extrapolation line corresponds to a theoretical, complete independence of R_{Mw} from modifier.

8.3 Mobile Phase

8.3.1 The Influence of the Organic Modifier on R_M

Mobile phases represent mixtures of water (or buffer) and an organic solvent (modifier) of various composition. The application of binary mixtures is mandatory even on those stationary phases, allowing the use of pure water. Solutes with intermediate or high lipophilicity would exhibit marginal migration distances in pure water, counteracting their exact measurement.

The most commonly used organic modifiers are methanol, acetone, and acetonitrile, while the use of dioxane or tetrahydrofuran is less frequent. Braumann [27] recommends methanol as the modifier of choice due to the quite pronounced similarity of its physico-chemical properties to water. It provides both strong hydrogen-bond donor and acceptor properties so that the addition to an aqueous mobile phase over a wide range of volume fractions will change the ordering of water molecules only in a limited manner. Our own experience supports the above recommendation (Fig. 1).

From a theoretical point of view, the extrapolation of R_M values, obtained with different solvent systems, to R_{Mw} (i. e., modifier-free conditions) should yield identical results for a given solute. This was experimentally proven by Biagi et al. [15] on silicone oil impregnated silica-gel plates. From their experimental findings the latter authors conclude, that it is solely the surface tension of the modifier which impacts the R_M values [15]. On the other hand, Cserháti [28] using paraffin-coated silica-gel showed substantial differences in R_{Mw} depending on the modifier used (methanol, acetone, or acetonitrile).

Retention on stationary phases with free silanol groups, such as ODS, will also be due to silanophilic hydrogen bond and dipole interactions, which vary with the modifier used [29, 30]. In these cases the electronic and H-bonding properties of the modifier will influence the R_M values in addition to surface tension.

8.3.2 The Influence of Solvent pH and Ionic Strength on R_M

The pronounced impact of ionization on partitioning processes is undisputed. As far as TLC is concerned, Biagi et al. [31] using silicone oil-impregnated plates and Cserháti [28] using paraffin-impregnated silica-gel showed that R_M depends on pH. Also de Voogt and coworkers [32] call for proper pH buffering in the case of ionizable solutes. In most cases investigators avoid pK correction by using experimental pH values which surmount the pH of the test compounds by at least 2 log units.

Several investigators have observed deviations from the generally applied rules for pK correction in TLC. For some organic acids Wilson [33] detected no variation in R_F related to pH (2–11). Negligible impact of salt concentration and pH (3–12) on retention was reported by Cserháti et al. [34] for some peptides and amino acids, using paraffin-impregnated silica-gel as the stationary phase. Kovács-Hadady and Szilágyi [35], using tricaprylmethylammonium-impregnated silica-gel plates, also found no effect of ionic strength and pH (2.4–9.4) on retention of minoxidil and its intermediates. Varying the pH between1 and 13 does not affect the retention of weak acids and bases

on RP18 silica-gel, as reported by Dross et al. [11]. According to these authors stronger bases (pK 8.0–10.7) exhibit no pH-related variation in mobile phases with high water content (75 %), whereas some pH-related effects are seen in the case of reduced water content (40 %). This latter finding is attributed by the authors to polar adsorption, which becomes more prominent in solvents with low water content. The pH-dependence of polar adsorption was shown by Cserháti and Szögyi [36] as well as Dross et al. [24], while Dingenen and Pluym [37] speculated on the influence of silanol dissociation in this context.

Above-described observations might be explained on the basis of concepts of Horváth et al. [38] on solvophobic interactions with octadecyl silica-gel. According to these authors it is solely the "hydrocarbonaceous moiety" of the solute "that can enter into hydrophobic interactions with the octadecyl chains of the stationary phase". Thus, the polar groups of the solute would remain outside the lipidic part of the stationary phase; correspondingly, ionization should not influence the chromatographic process. These conditions contrast profoundly with partitioning processes due to shaking out into an organic phase.

Polar interactions, which take place between the solute and SiOH groups of the plate material, are labeled silanophilic effects [39, 40]. An effect of ionization on partitioning behavior is therefore expected in the case of strong bases in modifier systems containing high amounts of methanol.

Whether the effects of ionization under the experimental conditions of HPLC and TLC are similar or not remains to be clarified. Even in the case of identical stationary phases, an essential difference between TLC and HPLC is the fact that for HPLC the columns have to be equilibrated before the runs. Correspondingly, the stationary phase of the column is adjusted to the pH of the solvent, while in the case of a TLC-plate a pH gradient is formed.

8.4 R_{Mw} and Extrapolation Methods

Since the introduction of R_{Mw}, i.e., the R_M extrapolated to modifier-free conditions, by Biagi et al. [12], this approach has attracted steadily increasing acceptance. Nevertheless in some papers [32, 41. 42] R_M values, estimated at a given modifier concentration, are still used for correlations with log P_{oct}; Kuchař and Jelínková [43] even state that R_M and R_{Mw} values are mutually interchangeable as lipophilicity descriptors.

The main reason for preferring R_{Mw} values can be derived from Fig. 2. Plots of R_M versus mobile phase content (φ) can vary considerably in slope for various test compounds. Thus, an appropriate expression of lipophilicity differences is only given at 100 % water, i. e., the extrapolated value, which better approximates the experimental conditions of lipid-phase/water partitioning systems. Another drawback of R_M values – in particular those estimated at rather low water contents in the mobile phase – is that they may be partially affected by polar adsorption and not exclusively by partitioning. Finally, using R_{Mw} values is accompanied by extended scales allowing better resolution in characterization of lipophilicity.

Most common linear extrapolation procedures are based on trial-and-error principles. As alternative approaches, some nonlinear methods based on equations,
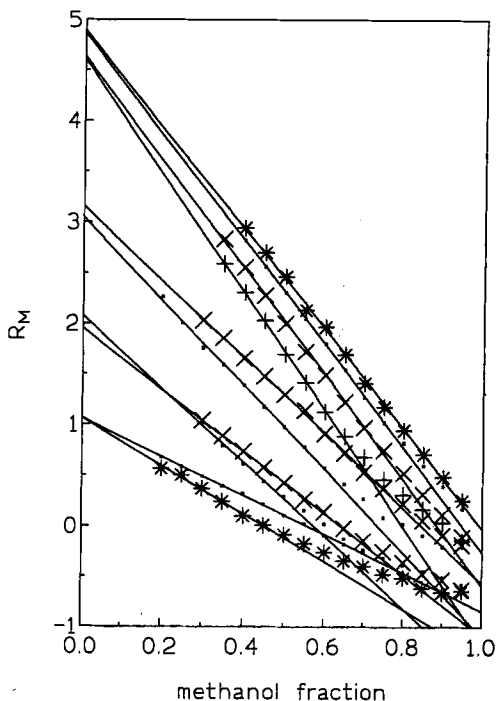
Figure 2. R_M values of some test compounds [24] as a function of the modifier fraction in the binary solvent. Extrapolation lines for five pairs of compounds are shown exhibiting identical or very similar R_{Mw} but significantly different slopes of the extrapolation lines. In general, deviations from linearity are found at low methanol content for those compounds with steeper extrapolation lines.

describing the entire dependence of R_M on modifier content including solvophobic and silanophilic interactions, may be useful [24, 44].

8.4.1 Quadratic Function

On the basis of an extended solubility parameter model [44] Schoenmakers et al. [45] describe the relation between $\log k'$ and the fraction of the modifier (φ) in the binary solvent mixture by the following equation:

$$\log k' = \log k_w + A\varphi^2 - S\varphi \tag{4}$$

Log k_w values calculated according to Eq. (4) have been correlated with $\log P_{oct}$ data by Braumann [27]. From these correlations it is concluded that, in the case of quite lipophilic compounds, the above approach yields overestimated k_w. This observation also holds when using R_M instead of $\log k'$ for the sake of R_{Mw} calculations [24].

8.4.2 Exponential Function

The plots of R_M versus φ resemble in their shape a decreasing exponential in the first part followed by an increasing one. Hence, one could describe this pattern empirically by the following equation:

$$R_M = \log (Ae^{-B\varphi} + Ce^{D\varphi}) \tag{5}$$

The decreasing exponential in Eq. (5) expresses the contribution of hydrophobic interactions between the test compound, the stationary hydrophobic phase and the aqueous mobile phase to R_M, while the increasing exponential corresponds to the contribution of polar adsorption. Parameters A, B, C, and D were calculated by nonlinear regression. If A, B, C, and D are given, R_{Mw} is calculated by setting $\varphi = 0$:

$$R_{Mw} = \log (A + C) \tag{6}$$

Eq.(6) is calculated by an iteration programme, which in some cases calculates negative values for D, indicating that the programme fits the plot of R_M versus φ to two decreasing exponentials. Although most R_{Mw}, calculated according to Eq.(6), agree well with those obtained by linear extrapolation [24] this approach is not generally applicable.

8.4.3 Mixed Exponential/Linear Function

An equation that separates two contributions to log k' was developed by Nahum and Horváth [39].

$$k' = A\ e^{B\psi} + (C + D\psi)^{-1} \tag{7}$$

The symbol ψ defines the water fraction in a binary mixture in the case of lipophilicity measurement by HPLC. For use in TLC, Eq. (7) has to be formulated as follows:

$$\text{antilog } R_M = A\ e^{B\psi} + (C + D\psi)^{-1} \tag{8}$$

Calculation of parameters A, B, C, and D by an iterative procedure and setting $\psi = 1$ (i.e., modifier-free), then gives:

$$R_{Mw} = \log [A\ e^{B} + (C + D)^{-1}] \tag{9}$$

R_{Mw} data calculated in this way ($= R_{Mw; Hor}$) are almost identical to those obtained by linear extrapolation [24].

Thus, we conclude that in most cases R_{Mw}, carefully calculated by linear extrapolation, is sufficiently precise. In critical cases the use of Eqs. (8) and (9) is recommended.

8.5 Analysis of the R_M/φ Relation

As pointed out by van de Waterbeemd and Testa [3] lipophilicity is mainly governed by the size (volume or surface) and the polarity of a solute:

lipophilicity = bulkiness – polarity $\tag{10}$

The same relations seem to be given for chromatographic retention mechanisms. Horváth and coworkers [38] assume that the slope of the R_M/φ relation gives the hydrophobic contact surface area of a solute. It should be noted, that Horváth and coworkers [38] define the modifier fraction φ in the solvent by a volume/volume relation, while Soczewinski and Golkiewicz [46] prefer the use of the logarithm of the molar modifier fraction.

Only a small number of authors [47, 48] follow the above interpretations of Horváth; nevertheless, the results of other groups investigating the correlation of slope and R_{Mw} in R_M/φ relations are not contradictory to the theoretical considerations of Horváth. Both Biagi et al. [15] and Kuchař and Jelínková [43] have shown high correlations of slope and R_{Mw} in the case of structurally related compounds, which differ only marginally in polarity. In the slope/R_{Mw} correlations of Biagi et al. [15, 49] the two outliers observed among 15 β-carbolines were the two nonaromatics in the series. Dross and Sonntag [48] have shown that, in contrast to benzoic acids, their test set of non polar structures exhibited lower slopes as compared with R_{Mw}; in terms of Eq. (10) this is easily understandable when inserting R_{Mw} as the lipophilicity parameter and defining the slope as an indicator of bulkiness as corresponds to Horváth.

Our group is currently exploring the possibility of deriving a polarity parameter from the above relation. Extrapolation of the linear part of the R_M/φ relation to water-free conditions gives a theoretical value for an R_M exclusively governed by partitioning processes. The difference between this value and the R_M derived experimentally with pure modifier is similar to β in the case of compounds with pure hydrogen bond acceptor properties, and very similar to the product $\alpha \times \beta$ in the case of compounds exhibiting both acceptor and donor properties.

8.6 Comparison with Other Lipophilicity Data

8.6.1 The Comparison of R_{Mw} with log k_w

Despite significant differences in the experimental procedures, the basic conditions are similar in RP-TLC and RP-HPLC, provided the stationary phases used are identical or at least similar. This is the case in the lower four rows, given in Table 2, as evidenced by intercepts around zero and regression coefficients near one, pointing to an at least numerical agreement between these two lipophilicity scales. In the upper four examples, stationary phases used in HPLC and TLC are very different; correspondingly, intercepts vary between -2 and $+2$ and regressors between 1 and 2.

8.6.2 The Comparison of R_{Mw} with log P_{oct}

Log P_{oct} represents the prime reference parameter as lipophilicity descriptor in QSAR; correspondingly, many investigators compare their TLC data (both R_M and R_{Mw}) by means of Collander equations with octanol/water partitioning data. Some typical examples for such correlations are summarized in Table 3. We list only correlations applying R_{Mw}, since this is the only case, for which a numerical comparability with log P_{oct} can be expected. The quality of correlation increases significantly when chemically

Table 8.2. $R_{Mw} = a + b \log k_w$. Correlations of RP-TLC data (R_{Mw}) with RP-HPLC data ($\log k_w$) are listed. Provided the stationary phases used in TLC and HPLC are very similar, the regression data (slope b and intercept a) indicate a rather good 1:1 fit; see lower four rows. In the case of different stationary phases (upper four rows) slopes and intercepts deviate significantly from 1 and 0.

Stationary phase		Modifier	a	b	r	n	Compounds	Reference
RP-HPLC	RP-TLC							
Bondapak C 18	Silicone	Acetone	−1.49	2.11	0.947	44	Cardiac glycosides	[67]
LiChrosorb C 18	Silicone	Acetone	2.11	1.68	0.938	28	Cardiac glycosides	[67]
Corasil C 18, silylated	Oleyl alcohol	Methanol	−1.97	0.95	0.938	11	Benzodiazepines	[68]
Porasil C, oleyl alcohol	Oleyl alcohol	Methanol	−1.01	1.01	0.966	11	Benzodiazepines	[68]
LiChrosorb C 18	RP 18-HPTLC	Methanol	0.01	1.39	0.978	45	Phenols	[23]
Bondapak C 18	KC 18 F	Acetonitrile	0.18	0.89	0.931	21	Benzo-dia-ze-pines	[60]
Bondapak C 18	KC 18 F	Methanol	0.28	1.09	0.954	17	β-carbolines	[69]
various C 18 silicagels	RP 18	Methanol	0.11	0.93	0.987	23	Various compounds	[24]

Table 8.3. $R_{Mw} = a + b \log P_{oct}$. Correlations of RP-TLC data (R_{Mw}) with experimental otanol/water partition coefficients ($\log P_{oct}$) are listed. In general, comparability is better in the case of chemically homogeneous sets of test compounds. Provided ODS material is used, acceptable correlations are even found for chemically diverse sets (see Eqs. (25) and (26)).

Stationary phase	a	b	r	n	Compounds	Reference	Equation
Silanized Kieselguhr	2.19	0.86	0.994	6	Hydrocoumarins	[52]	11
Silica-gel/octanol	0.17	0.51	0.868	31	Cardenolides	[55]	12
Silica-gel/silicone oil	0.97	0.54	0.693	31	Cardenolides	[55]	13
Silica-gel/silicone oil	−1.21	0.93	0.961	28	Phenols	[14]	14
Silica-gel/silicone oil	0.25	0.55	0.975	15	Penicillins	[31]	15
Silica-gel/silicone oil	1.25	0.63	0.914	23	Cardenolides	[67]	16
Silica-gel/silicone oil	0.49	0.37	0.916	40	Amines	[63]	17
Silanized precoated	0.47	0.26	0.861	48	Various drugs	[42]	19
Silanized precoated	0.32	0.80	0.991	11	Arylaliphatic acids	[41]	20
C 8-HPTLC	−1.26	1.21	0.985	17	NSAIDs	[59]	21
C 18-HPTLC	−0.35	1.07	0.983	17	NSAIDs	[59]	22
C 18-HPTLC	−0.24	1.10	0.963	28	Phenols	[23]	23
KC 18 F	1.24	0.60	0.929	39	Amines	[63]	24
C 18-TLC	−0.06	1.00	0.983	40	Simple compounds	[24]	25
C 18-TLC	0.26	0.93	0.961	49	Various drugs	[70]	26

related compounds are used for comparison. The R_{Mw} data used for Eqs. (14–17) in Table 3 have been measured under almost identical experimental conditions; corresponding intercepts (a) and regression coefficients (b) differ profoundly and indicate a putative decrease in significance when merging all these compounds for comparison. Nevertheless, the following equation for 415 chemically diverse heterocycles shows an acceptable correlation between R_M and log P, but the slope proves the significant deviation from a 1:1 fit [50].

$$R_M = 0.565(\pm0.011) \log P + 0.231(\pm0.037)$$

$$n = 415, s = 0.377, r = 0.933, F = 2803 \tag{18}$$

In contrast to the above conditions intercepts near 0 and regression coefficients approaching 1 are obtained in comparisons of log P_{oct} with R_{Mw} data estimated on plates with ODS as stationary phase (Eqs. (19–26)). The numerical agreement of log P_{oct} with R_{Mw} data is impressively demonstrated for a test set of 121 chemically diverse molecules (identical with the test set in Chapter 23 of this volume) by the following equation:

$$R_{Mw} = 0.959(\pm0,046) \log P_{oct} + 0.067(\pm0.147)$$

$$n = 121, s = 0.353, r = 0.967, F = 1703 \tag{27}$$

A linear Collander relation between log P_{oct} and R_{Mw} data is shown in Fig. 3. It should however be noted that such a linearity is not generally given. Pliška et al. [20] using cellulose plates and t-butanol as modifier could show nonlinear relations between log P_{oct} and R_{Mw} for a set of amino acids.

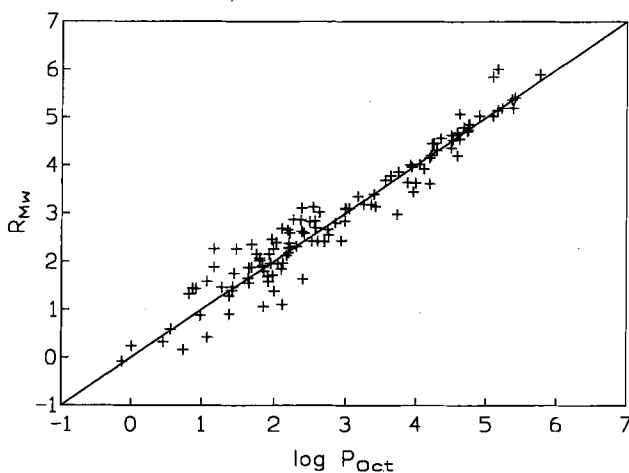**Figure 3.** Plot of R_{Mw} versus log P_{oct} data for a set of 121 chemically diverse molecules including both simple organic standard compounds and complicated drug molecules. TLC-data have been estimated on ODS material as stationary phase and methanol as modifier. The solid line represents the theoretical 1:1 fit and indicates linearity between R_{Mw} and log P_{oct} data; for details, see [70].

8.6.3 The Comparison of R_{Mw} with Calculated log P

Recently, computerized calculations of log P have attracted increasing interest as substitutes for experimental approaches. According to the methodology applied calculation procedures can be classified into fragmental methods (Σf, CLOGP), atom-based (CHEMICALC, SMILOGP, HINT) and molecular property-based approaches (ASCLOGP). For a detailed treatise the reader is referred to Chapter 23 of this volume.

For two test sets of 92 simple, rather homogeneous organic standard compounds and 68 chemically quite diverse drug molecules, we have correlated R_{Mw} data with calculated log P representing the three calculation procedures mentioned above. The results are summarized in Table 4. These calculations substantiate the better fit between chromatographic and partitioning data in the case of homogeneous as compared to chemically heterogeneous compounds. In addition, for both test sets fragmental calculation procedures exhibit higher interrelations with R_{Mw} data than atom – or molecular property – based procedures.

Table 8.4. $R_{Mw} = a + b \log P_{calc}$. Correlations of RP-TLC data (R_{Mw}) with calcuated otanol/water partition coefficients (log P_{oct}) are listed. Atom-based (CHEMICALC-2, HINT) and fragmental calculation approaches (Σf, CLOGP 4.34, KLOGP) were included. In general, the interrelation is better for simple organic standard compounds than for complicated drug molecules (included compounds are in detail given in Chapter 23 of this volume).

Calculation method	a	b	r	n
Organic standard compounds				
Σf	0.01	0.91	0.959	92
CLOGP 4.34	0.00	0.93	0.960	90
KLOGP	–0.05	1.00	0.959	90
CHEMICALC 2	0.24	0.89	0.959	92
HINT	0.52	0.77	0.912	91
Drug molecules				
Σf	1.18	0.66	0.834	65
CLOGP 4.34	0.95	0.72	0.911	56
KLOGP	0.71	0.75	0.836	67
CHEMICALC 2	1.87	0.58	0.803	68
HINT	1.29	0.59	0.823	65

8.7 Concluding Remarks

Exactness and reproducibility of chromatographic R_M data strictly depends on the application of standardized experimental conditions. Control of temperature and humidity, the use of front markers, the densitometric evaluation of starting and running points and the extrapolation to modifier-free conditions deserve mention here. Provided the TLC data are obtained under these experimental conditions they represent an attractive alternative to the tedious and time-consuming measurement of partition coefficients.

Important advantages of TLC are inter alia that test compounds need not be pure and that only trace amounts of test material are necessary. Compounds can be investigated over a broad lipophilicity range and a quantitative determination of their concentration (often posing analytical problems) is not necessary.

Also in comparison with RP-HPLC we view RP-TLC as a feasible alternative. One of the major advantages of RP-TLC is its speed. As described previously [24], 30 compounds can be tested simultaneously. This number can even be increased by double use of at least some starting positions. One only has to guarantee that the compounds sharing a starting position should differ in lipophilicity by at least one log unit which can easily be determined ahead of time by calculating their Σf or CLOGP values.

Another advantage of lipophilicity determination by TLC might be the somewhat broader range of measurable lipophilicities. According to Braumann [27] this measurable range comprises log k_w values from 0.0 to 7.0 in the case of RP-HPLC, while the range in the case of RP-TLC includes R_{Mw} values from about −1.0 to + 7.0.

An important disadvantage of TLC is its lack of applicability to liquid test compounds except those with a very high boiling point.

Taken together, from the authors' point of view TLC represents a convenient experimental alternative to RP-HPLC and octanol/water partitioning for measurement of molecular lipophilicity and application in QSAR studies. Table 5 summarizes some illustrative examples for successful applications of TLC data in QSAR. Additional examples are given by Tomlinson [18].

Table 8.5. Application of RP-TLC in QSAR studies

Compounds	Biological activity	References
Cardiac glycosides	Inhibition of Na-K-ATPase	[58]
Cardiac glycosides	Positive inotropic action	[71]
Dihydropyridines	Binding to Ca-channels	[73]
Dihydropyridines	Negative inotropic action	[74]
Phosphothionates	Fungicidal activity	[75]
Sulfonamides	Antimalarial activity	[78]
Benzodiazepines	Behavioral activity	[79]
Arylaliphatic acids	Fibrinolytic and antihemolytic activity	[80]
Acridines	Antitumor activity	[81, 82]
Bis-guanyl-hydrazones	Antileukemic activity	[83]
Heteroatomic hydrocarbons	Activity of *Photobacter* phosphorus	[32]
Heteroatomic hydrocarbons	Toxicity in *Poeciliea* reticulata	[32]

Table 8.5. Continued

Compounds	Biological activity	References
Heteroatomic hydrocarbons	Toxicity in *Daphnia magna*	[32]
Phenols	Toxicity in rats	[23]
Phenols	Toxicity in guppies	[23]
Phenols	Toxicity in *Daphnia magna*	[23]
Benzoic acids	Toxicity in mosquito larvae	[21]
Naphthols	Toxicity in chick embryo	[76]
Nitro-imidazoles	Biliary excretion	[77]
Class I antiarrhythmics	Therapeutic dose in humans	[72]

References

[1] Hansch, C., and Leo, A., *Substituent Constants for Correlation Analysis in Chemistry and Biology.* Wiley: New York, 1979

[2] Rekker, R. F., and Mannhold, R., *Calculation of Drug Lipophilicity. The Hydrophobic Fragmental Constant Approach.* VCH: Weinheim, 1992

[3] van de Waterbeemd, H., and Testa, B., *Adv. Drug Res.* **16**, 85–225 (1987)

[4] Kubinyi, H., *Progr. Drug Res.,* **23**, 97–198 (1979)

[5] Leo, A., Hansch, C. and Elkins, D., *Chem. Rev.* **71** 525–616 (1971)

[6] Taylor, P. J. Hydrophobic Properties of Drugs. In: *Quantitative Drug Design.* Ramsden, C. A. (Ed.). Pergamon Press: Oxford; 241–294 (1990)

[7] Kubinyi, H. QSAR: Hansch Analysis and Related Approaches. VCH: Weinheim, 1993

[8] Martin, A. J. P., and Synge, R. L. M., *Biochem. J.* **35**, 1358–1368 (1941)

[9] Consden, R., Gordon, A. H., and Martin, A. J. P., *Biochem. J.* **38**, 224–232 (1944)

[10] Bate-Smith, E. C., and Westall, R. G., *Biochem. Biophys. Acta* **4**, 427–440 (1950)

[11] Dross, K., Sonntag, Ch., and Mannhold, R., *J. Chromatogr.* **639**, 287–294 (1993)

[12] Biagi, G. L., Barbaro, A. M., Guerra, M. C., and Gamba, M. F., *J. Chromatogr.* **41**, 371–379 (1969)

[13] Biagi, G. L., Guerra, M. C., and Barbaro, Am. M., *J. Med. Chem.*, **13**, 944–948 (1970)

[14] Biagi, G. L., Gandolfi, O., Guerra, M. C., Barbaro, Am. M., and Cantelli-Forti, G., *J. Med. Chem.* **18**, 868–882 (1975)

[15] Biagi, G. L., Barbaro, A. M., Sapone, A., and Recanatini, M., *J. Chromatogr. A* **662**, 341–361 (1994)

[16] Biagi, G. L., Barbaro, A. M., Sapone, A., and Recanatini, M., *J. Chromatogr. A* **669**, 246–253 (1994)

[17] Biagi, G. L., Barbaro, A. M., and Recanatini, M., *J. Chromatogr. A* **678**, 127–137 (1994)

[18] Tomlinson, E., *J. Chromatogr.* **113**, 1–45 (1975)

[19] Boyce, C. B. C., and Milborrow, B. V., *Nature* **208**, 537–538 (1965)

[20] Pliška, V., Schmidt, M., and Fauchère, J. -L., *J. Chromatogr.* **216**, 79–92 (1981)

[21] Pyka, A., *J. Planar Chromatogr.,* **7**, 108–116 (1994)

[22] Geiss, F., *Fundamentals of Thin Layer Chromatography.* Hüthig: Heidelberg, 1987

[23] Butte, W., Fooken, C., Klußmann, R., and Schuller, D., *J. Chromatogr.* **214**, 59–67 (1981)

[24] Dross, K., Sonntag, Ch., and Mannhold, R., *J. Chromatogr. A.* **673**, 113–124 (1994)

[25] Brinkman, U. A. Th., and de Vries, G., *J. Chromatogr.* **192**, 331–340 (1980)

[26] MERCK, Darmstadt, Germany, personal communication

[27] Braumann, T., *J. Chromatogr.* **373**, 191–225 (1986)

[28] Cserháti, T., *Chromatographia* **18**, 318–322 (1984)

[29] Verzele, M., van Damme, F., Dewaele, C., and Ghijs, M., *Chromatographia* **24**, 302–308 (1987)

[30] Lamparczyk, H., Atomura, M., and Jinno, K., *Chromatographia* **23**, 752–759 (1987)

[31] Biagi, G. L., Barbaro, A. M., Guerra, M. C., Sapone, A., and Recanatini, M., Lipophilic Character of Penicillins and Cephalosporins: R_M and Log P Values as Lipophilicity Parameters. In: *QSAR in Design of Bioactive Compounds. Second Telesymposium on Medicinal Chemistry.* Kuchar, M., Biagi, G. L., Hopfinger, A. J., and Pliska, V. (Eds.). J. R. Prous Publishers: Barcelona; 47–63 (1992)

[32] de Voogt, P., van Zijl, G. A., Govers, H., and Brinkman, Th., *J. Planar Chromatogr.* **3**, 24–33 (1990)

[33] Wilson, J., *J. Chromatogr.* **354**, 99–106 (1986)

[34] Cserháti, T., Bordás, B., and Ösapay, G., *Chromatographia* **23**, 184–197 (1987)

[35] Kovács-Hadady, K., and Szilágyi, J., *J. Chromatogr.* **553**, 459–466 (1991)

[36] Cserháti, T., and Szögyi, M., *J. Chromatogr.* **520**, 249–256 (1990)

[37] Dingenen, J., and Pluym, A., *J. Chromatogr.* **475**, 95–112 (1989)

[38] Horváth, C., Melander, W., and Molnár, I., *J. Chromatogr.* **125**, 129–156 (1976)

[39] Nahum, A., and Horváth, C., *J. Chromatogr.* **203**, 53–64 (1981)

[40] Nikolov, R. N., *J. Chromatogr.* **286**, 147–162 (1984)

[41] Kuchař, M., Rejholec, V., Kraus, E., Miller, V., and Rábek, V., *J. Chromatogr.* **208**, 279–288 (1983)

[42] Dross, K., Mannhold, R., and Rekker, R. F., *Quant. Struct.-Act. Relat.* **11**, 36–44 (1992)

[43] Kuchař, M. and Jelínková , M., Some Problems of Chromatographic Evaluation of Lipophilicity. In: *QSAR in Design of Bioactive Compounds, Second Telesymposium on Medicinal Chemistry.* Kuchar, M., Biagi, G. L., Hopfinger, A. J. and Pliska, V. (Eds.) J. R. Prous Publishers: Barcelona; 65–81 (1992)

[44] Hildebrand, J. H., Prausnitz, J. M., and Scott, R. L., *Regular and Related Solutions*, van Nostrand-Reinhold: New-York, 1970

[45] Schoenmakers, P. J., Billiet, H. A. H., Tijssen, R., and de Galan, L., *J. Chromatogr.* **149**, 519–537 (1978)

[46] Soczewinski, E., and Golkiewicz, W., *Chromatographia* **6**, 269–273 (1973)

[47] Darwish, Y., Cserháti, T., and Forgacs, E., *J. Planar Chromatogr.* **6**, 458–462 (1993)

[48] Dross, K., and Sonntag, Ch., The estimation of physicochemical properties by RP18-TLC. In: *Trends in QSAR and Molecular Modelling '92*, Wermuth, C. G., (Ed.). ESCOM: Leiden; 370–372 (1993)

[49] Biagi, G. L., Pietrogrande, M. C., Barbaro, A. M., Guerra, M. C., Borea, P. A., and Cantelli-Forti, G., *J. Chromatogr.* **469**, 121–126 (1989)

[50] Biagi, G. L., Recanatini, M., Barbaro, A. M., Guerra, M. C., Sapone, A., Borea, P. A., and Pietrogrande, M. C., Chromatographic R_M values as lipophilicity index In: *QSAR: Rational Approaches to the Design of Bioactive Compounds.* Elsevier: Amsterdam 83–90 (1992)

[51] Hulshoff, A., and Perrin, J. H., *J. Chromatogr.* **120**, 65–80 (1976)

[52] van der Giesen, W. F., and Janssen, L. H. M., *J. Chromatogr.* **237**, 199–213 (1982)

[52] Green, J., Marcinkiewicz, S., and McHale, D., *J. Chromatogr.* **10**, 158–183 (1963)

[54] Wallerstein, S., Cserháti, T., and Fischer, J., *Chromatographia* **23**, 184–188 (1987)

[55] Cohnen, E., Flasch, H., Heinz, N., and Hempelmann, F. W., *Arzneim.-Forsch./Drug Res.* **28**, 279–288 (1983)

[56] Anton-Fos, G. M., Garcia-March, F. J., Perez-Gimenez, F., Salabert-Salvador, M. T., and Cercos-del-Pozo, R. A., *J. Chromatogr. A* **672**, 203–211 (1994)
[57] Kuchař, M., Kraus, E., and Jelínková, M., *J. Chromatogr.* **557**, 399–411 (1991)
[58] Dzimiri, N., Fricke, U., and Klaus, W., *Br. J. Pharmacol.* **91**, 31–38 (1987)
[59] LaRotanda, M. I., Amato, G., Barbato, F., Silipo, C., and Vittoria, A., *Quant. Struct.-Act. Relat.* **2**, 168–173 (1983)
[60] Pietrogrande, M. C., Borea, P. A., and Biagi, G. L., *J. Chromatogr.* **447**, 404–409 (1988)
[61] Pietrogrande, M. C., Dondi, F., Borea, P. A., and Bighi, C., *J. Chromatogr.* **471**, 407–419 (1989)
[62] Bruggeman, W. A., van der Steen, J., and Hutzinger, O., *J. Chromatogr.* **238**, 335–356 (1982)
[63] Biagi, G. L., Guerra, M. C., Barbaro, A. M., Barbieri, S., Recanatini, M., Borea, P. A., and Pietrogrande, M. C., *J. Chromatogr.* **498**, 179–190 (1990)
[64] Cserháti, T., Kiss-Tamas, A., and Mikite, Gy., *Chromatographia* **25**, 82–86 (1988)
[65] Gocan, S., Irimie, F., and Cimpan, G., *J. Chromatogr. A* **675**, 282–285 (1994)
[66] Glen, R. C., Rose, V. S., Lindon, J. C., Ruane, R. J., Wilson, I. D., and Nicholson, J. K., *J. Planar Chromatogr.* **4**, 432–438 (1991)
[67] Biagi, G. L., Barbaro, A. M., Guerra, M. C., Borea, P. A., and Recanatini, M., *J. Chromatogr.* **504**, 163–178 (1990)
[68] Hulshoff, A., and Perrin, J. H., *J. Chromatogr.* **129**, 263–176 (1976)
[69] Pietrogrande, M. C., Borea, P. A., Lodi, G., and Bighi, C., *Chromatographia* **23**, 713–716 (1987)
[70] van de Waterbeemd, H. and Mannhold, R., Lipophilicity Descriptors for Structure-Property Correlation Studies. In: *Lipophilicity in Drug Action and Toxicology.* Pliska, V., Testa, B., and van de Waterbeemd, H. (Eds.). VCH: Weinheim; 399–416 (1996)
[71] Dzimiri, N., and Fricke, U., *Br. J. Pharmacol.* **93**, 281–288 (1988)
[72] Mannhold, R., Voigt, W., and Dross, K., *Cell Biol. Int. Rep.* **14**, 361–368 (1990)
[73] Mannhold, R., Jablonka, B., Voigt, W., Schönafinger, K., and Schraven, E., *Eur. J. Med. Chem.* **27**, 229–235 (1992)
[74] Rodenkirchen, R., Bayer, R., Steiner, R., Bossert, F., Meyer, H. and Möller, E., *Naunyn-Schmiedeberg's Arch. Pharmacol.* **310**, 69–78 (1979)
[75] Gupta, R. L., and Roy, N., *Pesticide Sci.* **22**, 139–144 (1988)
[76] Biagi, G. L., Barbaro, A. M., Guerra, M. C., Andreotti, D., and Cantelli-Forti, G. The developing chick embryo as an alternative model in toxicity testing: a QSAR approach. In: *Pharmacochemistry Library*, Vol 10. Timmerman, H. (Ed.). Elsevier: Leiden; 349–351 (1987)
[77] Biagi, G. L., Cantelli-Forti, G., Barbaro, A. M., Guerra, M. C., Hrelia, P., and Borea, P. A., *J. Med. Chem.* **30**, 420–423 (1987)
[78] Saxena, A. K., and Seydel, J. K., *Eur. J. Med. Chem.* **15**, 241–246 (1989)
[79] Biagi, G. L., Barbaro, A. M., Guerra, M. C., Babbini, M., Gaiardi, M., Bartoletti, M., and Borea, P. A., *J. Med. Chem.* **23**, 193–201 (1980)
[80] Kuchař, M., Rejholec, V., Brunová, B., and Jeláünková, M., *J. Chromatogr.* **195**, 329–338 (1980)
[81] Denny, W. A., Atwell, G. J., and Cain, B. F., *J. Med. Chem.* **22**, 1453–1460 (1979)
[82] Baguley, B. C., Denny, W. A., Atwell, G. J., and Cain, B. F., *J. Med. Chem.* **24**, 520–525 (1981)
[83] Denny, W. A., and Cain, B. F., *J. Med. Chem.* **22**, 1234–1238 (1979)

9 The Future of log *P* Calculation

A. J. Leo

Abbreviations

AM-1   Molecular orbital program (by Dewar)
CLOGP   Program to calculate log *P* from structure (by Leo)
CoMFA   Comparative Molecular Field Analysis (by Cramer, Tripos, Inc.)
MO   Molecular orbital
PM-3   Molecular orbital program (by Stewart)
QSAR   Quantitative structural-activity relationship
SASA   Solvent-accessible surface area
SCAP   Solvent-dependent conformational analysis procedure
ZwI   Zwitterion

Symbols

$\mathring{A}$ Angstrom unit
$\log P_{oct}$ Experimental octanol/water partition coefficient
$\log P_{alk}$ Experimental alkane/water partition coefficient
α hydrogen-bond acceptor strength
β hydrogen-bond donor strength

9.1 Introduction

Almost a century ago Meyer [1] and Overton [2] laid the foundations for the use of partition coefficients as a measure of the way chemicals travel through and distribute themselves inside living systems. This work was followed by that of Meyer and Hemmi [3] in the 1930s and by Collander [4] in the early 1950s. So the time was ripe in the early 1960s for Fujita, Iwasa, and Hansch [5] to concentrate on the octanol/water system, and use log P_{oct} as a hydrophobic parameter. Combining this parameter with Hammett-Taft electronic and steric parameters, they were successful in constructing quantitative structure-activity relationships (QSAR) which rationalized some simple biological end-points, such as growth rate of oat shoots and inhibition of fly head cholinesterase [6]. It is doubtful that either Hansch or Fujita could have foreseen the torrent of research papers their work would initiate through the following three decades. Just a list of the fields in which the log P_{oct} of solutes have found a role is impressive: 1) protein binding and specific roles of the solute in enzyme stabilization, denaturation or enantioselectivity [7–11]; 2) inhibition of bacteria [12–15], molds protozoa and organelles [17–20]; 3) action as mutagens, carcinogens, or antineoplastics [16–18]; 4) action as

anesthetics, anticonvulsants, or pesticides [19–21]; 5) their effect on skin penetration, host-guest complexation and multiple drug resistance [22–24]; 6) in predicting behavior of dyes in photographic emulsions, histochemistry, and radiopaque media [25–27].

In the attempt to maintain a comprehensive database of measured log P values, references are made to over 500 different journals, and this does not include sources such as graduate degree theses, monographs, and private communications. It is safe to conclude that interest in the hydrophobic parameter is still on the rise with no end in sight.

Of course the number of reported measurements underestimates the level of use of log P_{oct} as a hydrophobic parameter. Many QSAR are constructed using calculated log P values, and scarcely a month goes by without a report of several new methodologies for their estimation. Almost all of them have their strong and weak points. An adequate discussion of each would take far too much space, and so reference must be made to a recent review article [28], even though that is fast becoming out of date.

9.2 Methods

Before presenting a brief review of the various procedures for calculating log P, it is well to clear up some confusion in labeling and nomenclature used in the past. The "atom-based" methods must take into account the number and types of bonds to surrounding atoms, i.e. their environment. Thus "atom-centered fragments" is a more accurate designation, and it takes over 100 (and preferably over 250) of them to adequately handle a data file of moderate size and complexity. At least one method using molecular properties derived from quantum chemical theory claims a sharp contrast to the methods employing fragments of substituents "which have no scientific basis" [29]. However, the parameters derived from molecular orbital (MO) calculation must first be validated through regression analysis and then they work well only when combined with indicator variables. There can be no doubt of the desirability of calculating log P from *molecular* properties, rather than from the summation of parts, but that requires knowledge of the solvent-directed forces which influence conformation, hydrogen bonding, tautomeric equilibria, etc.

9.2.1 The Substituent Method

Calculating log P by the "hydrogen replacement" (or π) method was the first to be published [5] and still should be considered the most dependable procedure whenever the log P for the "parent" structure is available. The reason it is dependable is that the parent structure frequently contains most of the needed "polar interaction factors", and one needs only to determine if the desired substituents create any new interactions.

9.2.2 Atom-Based Methods

Log P calculations which have been given this heading have be proposed by Broto and his colleagues [30], by Ghose and Crippen [31] by Moriguchi and coworkers [32], by

Table 9.1. Methods of calculation of log $P_{octanol/water}$ from structure

1. Substituent constants, π: Hansch and Fujita: log P (Ph–X) = log P (Ph–H) + π–X

2. Fragment-based: log $P = \Sigma\, a_n f_n + \Sigma\, b_m F_m$
   A. Rekker and Nauta (by regression; i.e., "reductionist")
   B. Pomona CLOGP ("constructionist"; i.e., greater weight give simplest analogs)
   C. Moriguchi and Iwase: originally SASA + Σf (see 3G)
   D. van de Waterbeemd (by regression; computerized)
   E. Rekker and Mannhold (by regression; manual)

3. Atomic: (i.e., atoms and environment of atom-base clusters)
   A. Broto and Moreau (regression, 222 variables)
   B. Ghose and Crippen; later Viswanadhan (regression 120 variables)
   C. Dubost and Croizet (computerized Broto and Moreau)
   D. Klopman (regression 39 variables; computerized in CASE)
   E. Suzuki and Kudo (groups: basic, extended an user defined)
   F. King (using molecular transform index)
   G. Moriguchi and Hirono: regression with $\Sigma a + \Sigma F + I$

4. Molecular properties
   A. Rogers and Cammarata
   B. Hopfinger and Battershell (SCAP)
   C. Bodor (regression: I_{alk}, 4th power ovality, 4th power atomic charges; total 15 parameters)
   D. Kasai (regression: charge transfer + electrostatic energies)
   E. Umeyama and Sasaki (4B + SASA)
   F. Richards and Essex (log P difference by molecular dynamics)
   G. Niemi and Basak (regression graph-theoretical invariant + molecular connectivity + H-bond)
   H. Politzer and Brinck (SASA + electrostatic potential)

Viswanadhan et al. [33] and by many others (see Table 1). The atom in question is characterized by the "cluster" of which it is the center. Somewhere between 100 and 300 such atom-centered clusters are chosen by means of regression analysis to cover most of the structural variation encountered. As can be anticipated, the variation in the training set is crucial to the method's ability to predict entirely new structures. This method of defining "clusters" is unable to cover more than a four-bond pathway between electronically interactive pairings, and this can lead to difficulty at times. Intramolecular hydrogen bonding is also difficult to deal with [30]. Calculation by computer is available for most of these.

9.2.3 Methods Based on Molecular Properties

The ratio of the free energies of solvation in the two phases, water and wet octanol, ought to yield a dependable partition coefficient. An early attempt to calculate this value with quantum chemical methods was made by Rogers and Cammarata [34], fol-
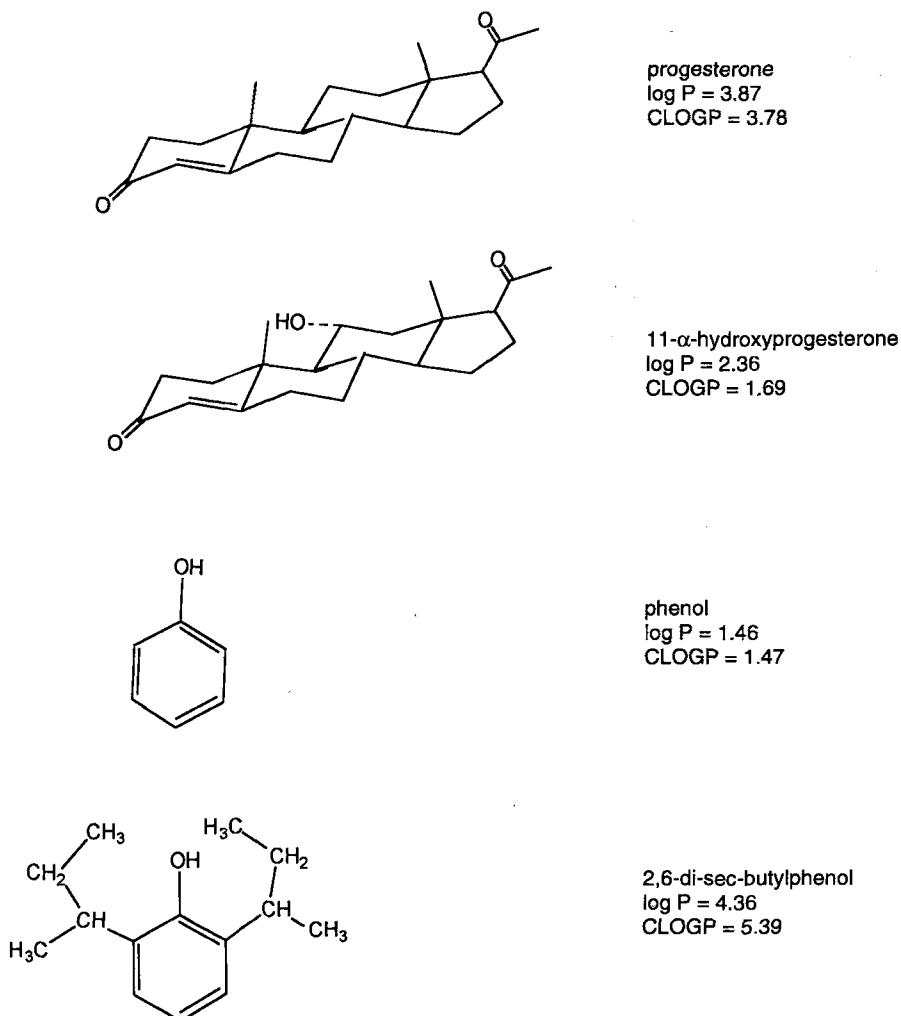
progesterone
log P = 3.87
CLOGP = 3.78

11-α-hydroxyprogesterone
log P = 2.36
CLOGP = 1.69

phenol
log P = 1.46
CLOGP = 1.47

2,6-di-sec-butylphenol
log P = 4.36
CLOGP = 5.39

Figure 1. Anomalies with sterically hindered hydroxyl fragments.

lowed by Hopfinger and Battershell using SCAP (solvent-dependent conformation analysis procedure) [35]. It must be kept in mind that there is a demand for methods to cope with very large databases of complex structures such as antibiotics and peptides. The present calculation rate of 6000 structures per minute (CLOGP operating on an SGI INDY) is, perhaps, faster than necessary. In contrast, good quantum chemical calculations take time, and currently may be applied to only a few of the most interesting structures.

As might be expected, solvent-accessible-surface area (SASA) is a calculated molecular property which appears in most log P calculations of this type. There is good evidence that solute size is one of the primary determinants of log P [36]. Although it

would seem logical to do so, there are serious pitfalls in attempting to separate SASA into two components – one hydrophobic and one hydrophilic [37]. One quite often finds a polar group in certain structural contexts in which it appears to be not as hydrophilic as usual. A prominent example is the 11-α or β substituents in steroids. When this position is unsubstituted, as in progesterone or estradiol, the calculations are normal, as seen in Fig. 1. With a hydroxyl group or carbonyl oxygen at this position, it appears that a correction of about +1.0 log units is required. Using the SASA as calculated by the SAVOL program [38], allowing for a water radius of 1.5 Å, it was found that the oxygen of an 11-β-hydroxyl group exposed only 6.46 square Angstroms ($Å^2$) to the solvent while in cyclohexanol 25.97 $Å^2$ was exposed. This lent support to the postulate that the decrease in hydrophilicity resulted from the shielding of the polar oxygen by surrounding hydrocarbon. However, this explanation fails to consider the partition equilibrium as a competing process. The oxygen in 2,6-di-*sec*-butylphenol exposes less than half the area to solvent, as does the oxygen in the parent phenol, but in this case the measured log P is one log units *lower* than calculated. It makes more sense to propose that octanol, as an H-donor, has more difficulty reaching a hindered H-acceptor than does water. Thus the decreased hydrophilicity of oxygens at the 11 position in steroids remains an anomaly.

9.2.4 Fragment-Based Methods

The first published fragment method of calculating log P_{oct} was proposed in 1973 by Nys and Rekker [39–41], and is still widely applied as a manual procedure. A variation of it is available on computers. The Medchem method, developed at Pomona College and also based on the additivity of fragments, considers so many interaction factors that it is not recommended as a manual procedure but should be applied via the CLOGP program. The two methods fragment the solute structure in different ways, but each requires some knowledge of the fragment's attachment bonds before lookup in a table provided. CLOGP uses five bond types while the Rekker procedure utilizes only two. The fragment interaction factors are crucial to both, and until these are understood more fully, it is impossible to say which treatment of them can be judged superior. Therefore, the balance of this paper will focus on these common problems, which, in a different guise perhaps, are also the stumbling blocks to the proper application of atom-based procedures as well as quantum chemical methods.

9.3 Common Problems

9.3.1 How is the "True" Structure to be Represented?

If a structure can exist in two tautomeric forms, the equilibrium is often quite different in water than it is in wet octanol. Of course the partitioning equilibrium constant is necessarily dependent upon the tautomeric ratio in the separate phases and this is not easily predicted *ab initio* in every case. In the case of keto/enol tautomers, it is the keto form which is promoted by water, because it has the higher H-bond acceptor
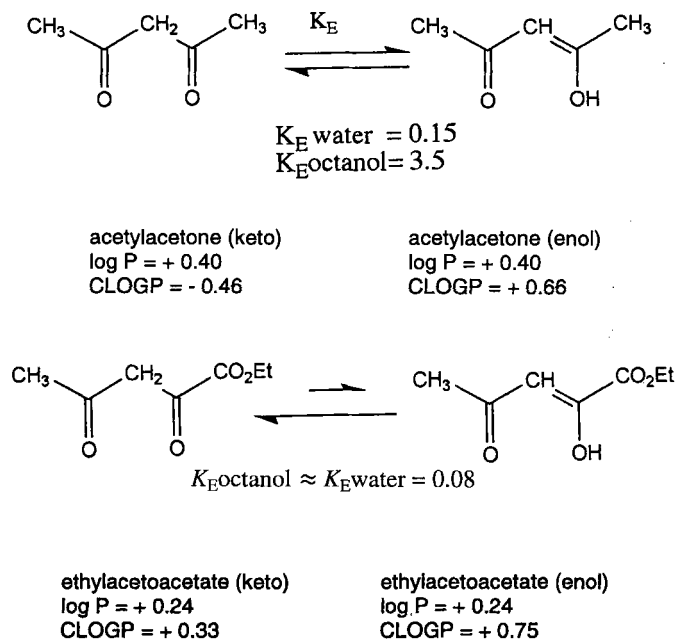
$$K_E \text{ water} = 0.15$$
$$K_E \text{octanol} = 3.5$$

acetylacetone (keto)
log P = + 0.40
CLOGP = - 0.46

acetylacetone (enol)
log P = + 0.40
CLOGP = + 0.66

$$K_E \text{octanol} \approx K_E \text{water} = 0.08$$

ethylacetoacetate (keto)
log P = + 0.24
CLOGP = + 0.33

ethylacetoacetate (enol)
log.P = + 0.24
CLOGP = + 0.75

Figure 2.   Calculation of keto/enol tautomers.

strength, β. Conversely, it is the enol form which is preferred by octanol. Generally, the log P calculated for the two structures bracket the measured value, as is seen for acetylacetone in Fig. 2. When neither phase supports the enol form, the calculated value for the keto should closely match the measured, as is seen for ethylacetoacetate. Most chemists know that the 2-pyridone form predominates over the 2-pyridinol, and would enter the former to get the correct calculation, as seen in Fig. 3. It is not as well known that electron withdrawing substituents near the pyridine nitrogen favor the enol form [42], and one might not realize that the observed value for the 6-chloro analog ought to lie about midway between those calculated for the two forms.

One cannot totally discount the possibility that certain hydroxy/keto combinations may exist largely as a hemiacetal in one solvent and not it the other. In calculating log P for the drug, celiprolol, it was noted that the deviation between CLOGP and measured value was +0.29, which is not unreasonable for an H-bonded ring of eight atoms *not* considered by CLOGP. However, as seen in Fig. 4, the CLOGP value for the hemiacetal structure happens to agree very well with measured. This is not presented as evidence for the presence of hemiacetal, but probably a measurement using C-13 NMR would decide if that possibility had merit, if anyone were so motivated.

If the two pK_as in a zwitterion are separated by seven log units, as they are in glycine, for example, there is little question but what both functionalities should be given their respective charges if the log P value desired is for a pH in the range of 5 to 7.5.

2-pyridone
log P = - 0.58
CLOGP = - 0.57

2-pyridinol
log P = -0.58
CLOGP = + 0.93

6-Cl-2-pyridone
log P = 0.93
CLOGP = 0.34
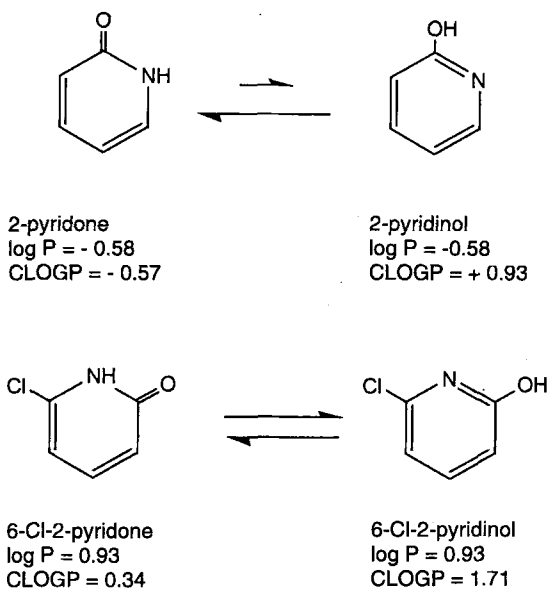
6-Cl-2-pyridinol
log P = 0.93
CLOGP = 1.71

Figure 3. Calculation of pyridones.

For convenience, in CLOGP both acid and amine are given the unionized notation, and a ZwI correction factor is applied later, (see Table 2). Versions 3.55 and later of CLOGP consider the α-carbon of amino acids as part of a "superfragment", for the simple reason that it gets "lost" in the intense field and cannot be considered as "isolating" or "insulating". This is most important in the case of the amino acids with polar side chains, (Table 2, 1B) When the α-carbon is part of a hydrocarbon ring, the cyclopropyl analog appears anomalous. (Table 2, 1C) When a hydrocarbon chain separates the opposing charges, there is evidence that this correction should be sequentially
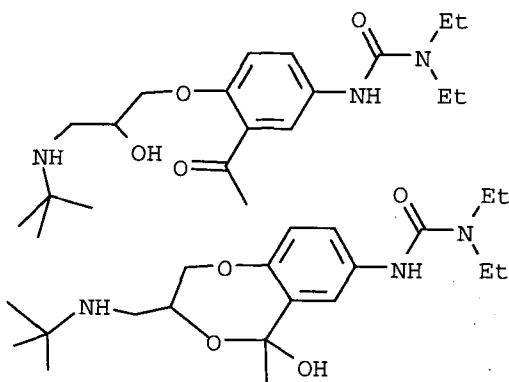
Figure 4. Celiprolol (log P = 2.07): intramolecular H-bond (CLOGP = 1.78) or hemiacetal (CLOGP = 2.09)?

Table 9.2. Calculation of zwitterions

Zwitterion: as neutral + correction

1. α-Amino acid as "Superfragment"

$$H_2N\text{-}CH\text{-}Co_2H$$
$$|$$
$$R$$
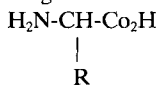
A.   R = H to C_4H_9;   Average Deviation = 0.05

B.   R   Deviation
    1. CH_2OH −0.33
    2. $CH_2CH_2SCH_3$ −0.14
    3. CH_2CONH_2 +0.22
    4. $CH_2CH_2CONH_2$ +0.31   Av. = 0.25

C. R = Alkane ring of C_n (CH → C)

n	Deviation
2	+0.92(?)
4	+0.30
5	+0.15

2. α–ω-Amino acids: $H_2N(CH_2)_nCO_2H$
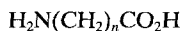
n	Deviation
1	0.00
2	+0.12
3	+0.16
4	+0.17
5	+0.67

reduced when n exceeds 3 or 4 (Table 2), but since a greater than three-atom separation is most frequent in peptides and the scaling of the correction is not needed there, it has not been implemented in CLOGP. As seen in Fig. 5, currently available data infers that the fifth methylene group separating the amine from the acid moieties actually *reduces* log *P*, a fact which is hard to reconcile unless this length of chain promotes a folded ion-sandwich.

Some important solutes, such as tetracycline shown in Fig. 6, have three or more ionizable moieties, and establishing the charge distribution in the molecule is no small task. Even if the charge distribution were known, predicting its effect on the partitioning equilibrium would not be a simple matter. CLOGP makes no allowance for either formal charges or H-bonding in structures of this type, and it estimates log *P* at −1.86. The average of seven measured values is −1.42, and a deviation of +0.42 may be con-
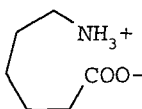
Figure 5. Possible (intramolecular) folding in α–ω-amino acids. Experimental log *P* = −2.63 (*n* = 4) and −2.95 (*n* = 5).
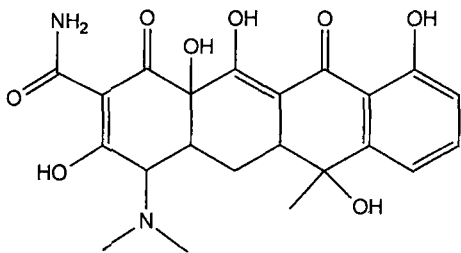
Figure 6. Problems with solutes having multiple pK_as. Tetracycline (pK_as = 3.3, 7.7, 9.6; averaged log P = –1.42; CLOGP = –1.86) has five sites for internal H-bonding. This causes a positive deviation between log P and CLOGP, while a negative deviation is due to the zwitterionic character of the compound. Both effects may partially cancel each other.

sidered satisfactory for a structure of this complexity. However, this moderately good agreement may be fortuitous – the result of factors not considered but whose effect happens to be compensating (i.e., weak zwitterion counter-balanced by H-bonding). The fragment methodology is, perhaps, inappropriate for this type of structure, but current MO calculations may not be a great deal more reliable.

Even carboxyl-containing solutes without a strong protonation site, such as methotrexate with pK_as of 3.76, 4.83, and 5.60, seem to cause trouble, as seen in Fig. 7. With the two carboxyls esterified, as in the diethyl ester, the calculation is quite satisfactory (log P = 0.98; CLOGP = 0.84). The measured values from the literature for methotrexate itself (–1.8 at pH 2.2, –2.59 at pH 2.2, –2.52 at pH 7.5), indicate some difficulty in measurement and perhaps a strong dependence upon counter-ion type. The CLOGP value of –0.87 seems to indicate that considerable charge exists on this molecule at both pH 2.2 and 7.5, but better data are needed before any reliable conclusion is reached.

9.3.2 Intramolecular H-bonding

The normal criteria for estimation of the strength of H-bonds do not show a direct and clear relationship with the effect they have on log P_{oct}. Sometimes one forgets that, in partitioning, we are dealing with a *competition* between solvation forces. *o*-Nitro-
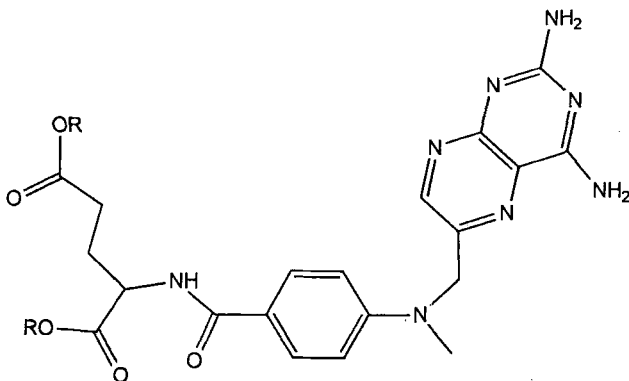

Figure 7. Problems with solutes having multiple acidic pK_as. Methotrexate (pK_as = 3.76, 4.83, 5.60; R = H, log P ≈ –2.5; CLOGP = –0.87) and its diethyl ester (R = Et, log P = 0.98; CLOGP = 0.84).
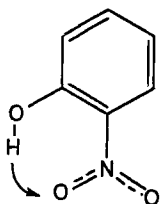
Figure 8. Intramolecular H-bond difference between solvent systems. *Ortho* effect in nitrophenols: log P_{oct} = 1.96 (average for *meta* and *para*) and 1.79 (*ortho*); log P_{alk} = -1.85 (average for *meta* and *para*) and +1.40 (*ortho*).

phenol has long been considered a prototype of intramolecular H-bonding, as witnessed by the undergraduate laboratory experiment of separating it by steam distillation from its *meta-* and *para*-isomers. Even though its water-solubility is reduced by internal H-bonding, the octanol-solubility must be reduced even more, because its log P_{oct}, as seen in Fig. 8, is *lower* by about 0.1 log unit compared with the *meta* and *para* isomers. In contrast, the log P_{alk} for the *ortho* analog is over 3.0 log units *higher*. The type of structures that do show the expected positive shift in *octanol* log P due to H-bonding have a carbonyl group *ortho* to an H-bond donor, such as OH or NH, (see Fig. 9). Another H-bonding combination pairs a strong donor, such as C(=O)NH-, with an acceptor, such as OMe. In all of these, but especially with the C(=O)/OH pair, the angle and distance is far from ideal for hydrogen bonding. In fact, there is good evidence to indicate that some twisting of the carbonyl out of the ring plane is required for the H-bond to form. Fragment pairs, each having donor/acceptor functions but each being rather bulky, may be prevented from hydrogen bonding internally because of steric constraints. As seen in Fig. 9, this seems to be the case with *o*-phthalic acid and *o*-phthalamide where the "twisting" prevents significant H-bonding.

ortho correction

+0.70

+0.60

-1.12

-1.28

Figure 9. The *ortho* factor: positive for H-bonding, negative for twisting.
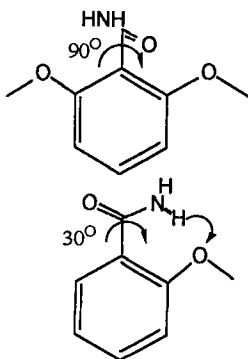
Figure 10. Hydrophobic reversal for 2,6-disubstituted benzamides. π (OMe) on benzene = –0.02; *meta* and *para* on benzamide = +0.21; *ortho* on benzamide = +0.20; 6- on 2-MeO-benzamide = –1.04.

Steric "twisting" can become a problem of some significance when a substituent is placed at the 6- position to groups that ordinarily H-bond in the 1,2- positions. Depending upon the steric parameters of the 2 and 6 substituents, the central group may be pushed so far out of plane that it becomes much more hydrophilic than an ordinary aromatic group, and furthermore the positive effect of the H-bonding may be lost. This can be thought of as a "steric reversal" of hydrophobicity. As seen in Fig. 10, benzamides provide a good illustration of this rather strange "reversal" behavior when methoxyl groups are placed successively in the 2 and 6 positions. One expects that electronic effects will make the methoxyl appear more hydrophobic on benzamide than it is on benzene, and, indeed, its π-value is +0.21 in the *meta* or *para* position as compared with a π-value of –0.02 on benzene. Its π-value is the same (+0.2) in the *ortho*-benzamide (log P = 0.84), but it appears that this is due to the cancellation of opposing effects. Repulsion of the electron lone pairs of the oxygens results in a partial twisting out of plane of the carbonyl and methoxyl groups (less delocalization and greater hydrophilicity), but this is compensated by the formation of an H-bond, with NH as donor. Initially it came as a surprise that the log P for the 2,6-dimethoxy analog was found to be –0.22, which makes the effective π-value of the 6-methoxyl group, at –1.04 – over a log unit lower than on benzene and 1.25 log units lower than the first *ortho*-methoxy. It seems very likely that further steric repulsion by the second methoxyl makes H-bonding impossible.

An apparent "twisting" of this sort in the antipsychotic, remoxipride, was extensively studied by Hogberg [43] and his associates at Astra Laboratories. They concluded that in the solid state the carbonyl group was almost perpendicular to the ring, but in the 6-hydroxy analogs (i.e., the salicylamide analogs) the benzamide portion was rigidly planar and stabilized by two intramolecular H-bonds. The research group at Lausanne collaborated with Hogberg in a further study of a number of the salicylamide analogs, including eticlopride and raclopride. As seen in Fig. 11, the measured log P for remoxipride is over a log unit lower than that estimated by CLOGP, which is similar to the deviation for 2,6-dimethoxybenzamide. In both cases the current version of CLOGP sees H-bonds that are not there. However, for the 6-hydroxy analogs (FLA-797 measured by Hogberg, and raclopride measured at Lausanne) there is still a large negative deviation between measured log P and that calculated allowing for H-
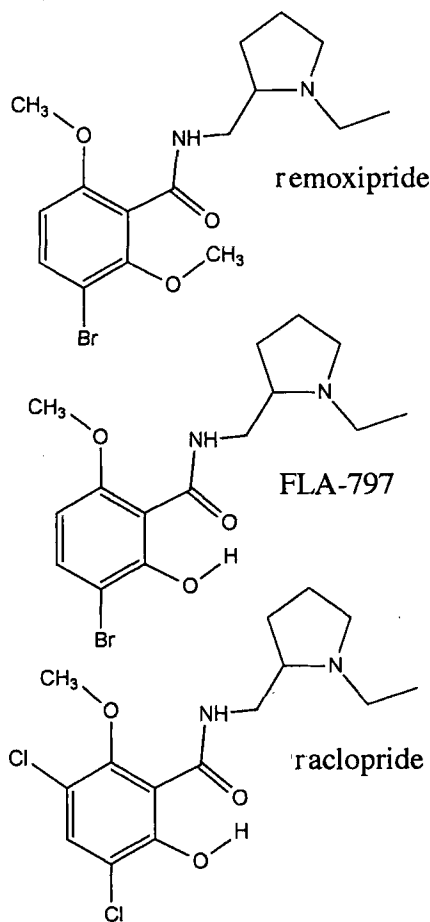
Figure 11. Hydrophobic reversal for antipsychotic benzamides. Remoxipride (log P = 2.10; CLOGP = 3.26), FLA-797 (log P = 2.0?; CLOGP = 3.73), raclopride (pK_as = 5.81, 9.21; log P = 1.33; CLOGP = 4.04).

bonding. This deviation may be due, at least in part, to the need for a zwitterion correction, since the phenolic hydroxyl in the analogs measured has been activated by halogen substitution.

The most important point to keep in mind is that in order to make dependable hydrophobicity estimates of solutes of actual interest, complex structures such as raclopride demand that many other factors, not present in simple models, must be considered. The research done at Lausanne [44] showed that steric parameters of groups seemingly distant from the protonation site affected the phenolic pK_a and thus the amount of zwitterion present at physiological pH, which is rather important knowledge to a drug designer. They also found that the rotational barriers of the amide moiety, which can directly affect activity as well as hydrophobicity, seem to be calculated better by molecular dynamics force field methods rather than by AM1 or PM3.

In view of all this, it is only realistic to acknowledge that the CLOGP program – and all others based on fragmental or atomic summations – have important limitations.
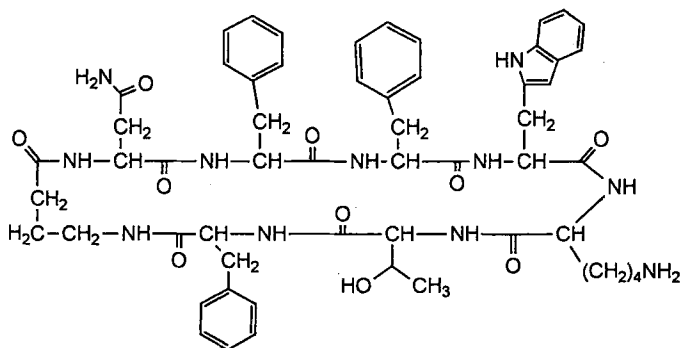
Figure 12. Intramolecular H-bonding in cyclic peptides. Mini-somatostatin (log P = 2.69; CLOGP = 2.30). In the CLOGP calculation H-bonds and decreased solvent accessible surface area have not been considered.

However, current quantum chemical calculations also have their limitations, and when measured log Ps differ markedly from those estimated by a program such as CLOGP (and if one also knows what the program considers and what it does not) it may focus one's attention on which one of a variety of MO calculations treats that structural problem best. It is likely that worthwhile information can flow between empirical and theoretical methods in *both* directions.

Since a frequent (if not the most frequent) problem to be faced in log P calculations involves structures which may form intramolecular H-bonds, it should be possible to "correct" a computation which neglects them with information from NMR or IR spectra which detects them. Obviously these spectra-based corrections can only be made subsequent to synthesis and then only for those structures which are of special interest. Some attempts to make H-bond corrections on calculated log Ps of cyclic peptides using spectral data have been carried out by Peter Moser and his colleagues at Ciba-Geigy (P. Moser, personal communication). The early work, at least, indicates some pitfalls when this procedure is applied to very complex structures, because the loss of *hydrophilic* contribution of the H-bonding groups can often be balanced by a loss of *hydrophobic* contribution of overlapping hydrocarbon moieties.

The cyclic peptide minisomatostatin, which is depicted in Fig. 12, has a free lysine side chain, giving it a pK_a of 10.2. The partition coefficient measured above pH 11 gave a log P for the undissociated solute of 2.69. Moser presented evidence for the formation of three intramolecular H-bonds which he postulated would raise the estimated
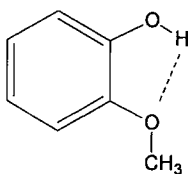

Figure 13. Intramolecular H-bond difference measured by spectral methods. Internal H-bonding measured for *o*-methoxyphenol in CCl_4 (100 %), $CHCl_3$ (50 %) and octanol (c. 10 %). CLOGP uses the correction factor F_{ortho} = –0.25, i.e., "twisted", but no H-bond.
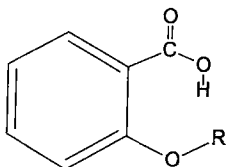
Figure 14. Some steric factors not easily programmed. Twist, but no H-bond giving as *ortho* interaction for R = methyl (–0.43) and R = phenyl (–1.09). The current CLOGP program does not consider size-dependency of the R-group when setting an *ortho* correction.

log *P* by 1.80 log units. Obviously these cross-links reduce the solvent-accessible surface area (SASA), and I know of no current program which will estimate with confidence how that reduction will affect the hydrophilic/hydrophobic ratio. The CLOGP estimate is 2.30. The difference of +0.39 might indicate that the reduction in the overall solute size is more than counteracted by the cross-links which reduce H-bonding to solvent, but the loss of SASA of the remaining groups may or may not be evenly balanced between the hydrophilic and hydrophobic fragments. Obviously a number of other interpretations of these results could be proposed.

M. Berthelot and C. Laurence at the University of Nantes have been obtaining H-bonding constants by measuring the equilibrium constants for the 1:1 complex using standard donors and acceptors. This is an extension of the solvatochromic approach initiated by Kamlet et al. [45]. This study of the fundamentals in hydrogen bonding can be of real utility to those trying to include it in calculation of log *P*. A simple illustration of its use comes from the solute, *o*-methoxy phenol. M. Berthelot and C. Laurence (personal communication) have shown it to be 100 % "chelated" (i.e., internally H-bonded) in carbon tetrachloride, 50 % chelated in chloroform, and approximately 10 % chelated in wet octanol (see Fig. 13). Presumably the amount of chelation would be negligible in water. This is compatable with the *ortho* effect previously programmed into CLOGP, which is –0.25 and indicates a loss of conjugation by the methoxyl is not compensated by an internal H-bond.

Unfortunately, some information about the partitioning process has not yet made its way into the CLOGP program. In the original algorithm the assignment of "ortho-class" was made for the generic form of a fragment; that is, for a fragment with more than one valence bond, it applied to all groups attached on the other bond(s). The problem is illustrated in Fig. 14 by phenyl ethers with an *ortho* carboxyl group. The calculations for *ortho*-*n*-alkoxy benzoic acids are quite satisfactory using a factor of –0.43, but the *o*-phenoxy analog needs a much larger negative factor because of its greater bulk.

9.4 Conclusions

This paper highlights only a few problems which will be critical in determining *if* log P_{oct} calculation has a future, and if so, what *direction* it will take. The following points bear emphasizing:
1. There will be a continuing need for *rapid* calculation of hydrophobic parameters for large databases. Methods based on "fragments", "substituents" or "atom-centered" groups will always have a speed advantage over *ab initio* MO calculations, but the

latter will be used more and more frequently for structures of special interest at they become better at dealing with polar and nonpolar solvent effects.

2. Spectral data, which determines the separate hydrogen bond donating strength (α) and accepting strength (β) of eash polar fragment and also their effective sum, will become increasingly important in improving the accuracy of log P estimates.

3. The need for accurate measurements of log P (especially over a range of pHs) will increase rather than decline as calculation methodologies improve.

4. So many factors must be considered in obtaining reasonably accurate log P estimates that computer calculation is already essential. However, researchers who use only the bottom line" from the computer output (i.e., use it as a "black box") will be ignoring its greatest value; learning how *each part* of the structure is affected by polar and non polar solvent forces. This is now being appreciated by those designing sophisticated modeling programs such as CoMFA [46].

References

[1] Meyer, H., *Arch. Exp. Pathol. Pharmakol.* **110**, 42–50 (1899)

[2] Overton, E., *Studien über die Narkose*, Fischer, Jena, Germany, 1901

[3] Meyer, K. and Hemmi, H., *Biochem. Z.* **277**, 39–71 (1935)

[4] Collander, R., *Acta Chem. Scand.* **4**, 1085–1098 (1950)

[5] Fujita, T., Iwasa J., and Hansch, C., *J. Am. Chem. Soc.*, **86**, 5175–5180 (1964)

[6] Hansch, C., Muir, R., Fujita T., Maloney, P., Geiger, F., and Streich, M., *J. Am. Chem. Soc.* **85**, 2817–2824 (1963)

[7] Helmer, F., Kiehs, K. and Hansch, C., *Biochemistry* **7**, 2858–2863 (1968)

[8] Hansch, C., and Klein, T., *Acc. Chem. Res.*, **19**, 392–400 (1986)

[9] Matsumura, M., Becktel, M., and Mathews, B., *Nature* **334**, 406–410 (1988)

[10] Mozhaev, V., Khmelnitsky, Y. Serveeva, M., Belova, A., Klyachko, N., Levashov, A, and Martinek, K., *Eur. J. Biochem.* **184**, 597–602 (1989)

[11] Sakurai, T., Margolin, A., Russell, A., and Klibanov, A., *J. Am. Chem. Soc.* **110**, 7236–7239 (1988)

[12] Martin, Y., and Lynn, K., *J. Med. Chem.* **14**, 1162–1166 (1971)

[13] Kirino, O., Takayama, C., and Inoue, S., *J. Pest. Sci.* **12**, 79–84 (1987)

[14] Craig, P., and Hansch, C., *J. Med. Chem.* **16**, 661–667 (1973)

[15] Hansch, C., and Cornell, N., *Arch. Biochem. Biophys.* **151**, 351–352 (1972)

[16] Debnath, A., Lopez de Compadre, R., Debnath, G., Shusterman, A., and Hansch, C., *J. Med. Chem.* **34**, 786–796 (1991)

[17] Rippman, F., *QSAR*, **9**, 1–5 (1990)

[18] Hansch, C., Smith, R., Engle, R., and Wood, H., *Cancer Chemother. Rep.* **56**, 443–456 (1972)

[19] Gupta, S., *Chem. Rev.* **89**, 1765–1800 (1989)

[20] Yamagami, C., Takao, N., Tanaka, M., Horisaka, K., Asada S., and Fujita, T., *Chem. Pharm. Bull.* **33**, 2403–2410 (1985)

[21] Hansch, C., and Fujita, T. (Eds.). *Classical and 3-D QSAR in Agrochemistry and Toxicology.* ACS: Washington, DC, 1995

[22] Driessman, J. B., Ridout, G., and Guy, R. H. Delivery system technology. In: *Comprehensive Medicinal Chemistry*, Vol. **5**. Hansch, C., Sammes, P. G., and Taylor, J. B. (Eds.). Pergamon: New York; 615–660 (1990)

[23] Newcomb, M., Moore, S., and Cram, D., *J. Am. Chem. Soc.* **99**, 6405–6410 (1977)

[24] Selassie, C., Hansch, C., Khwaja, T. A., Structure-activiting relationships of multidrug resistance. In: *Chemistry and Biology of Pteridines*. Curtius, H., Ghisla, S., and Blau, N. (Eds.). De Gruyter: Berlin; 1217–1220 (1990)

[25] Eastman Kodak Co., private communication

[26] Rashid, F., and Horobin, R., *Histochemistry*, **94**, 303–308 (1990)

[27] Havaldsen, J., Nordal, V., and Kelly, M., *Acta Pharm. Suec.* **20**, 219 (1983)

[28] Leo, A., *Chem. Rev.* **93**, 1281 (1993)

[29] Bodor, N., Gabanyi, Z., and Wong, C.-K., *J. Am. Chem. Soc.* **111**, 3783–3786 (1989)

[30] Broto, P., Moreau, G., and Vandycke, C., *Eur. J. Med. Chem.* **19**, 71–78 (1984)

[31] Ghose, A., and Crippen, G., *J. Comput. Chem.* **7**, 565–577 (1986)

[32] Moriguchi, I., Hirono, S., Liu, Q., Nakagome, I., and Matsushita, Y., *Chem. Pharm. Bull.* **40**, 127–130 (1992)

[33] Viswanadhan, V., Ghose, A., Revankar, G., and Robins, R., *J. Chem. Inf. Comput. Sci.* **29**, 163–168 (1989)

[34] Rogers, K., and Cammarata, A., *Biochim. Biophys. Acta*, **193**, 22–28 (1969)

[35] Hopfinger, A., and Battershell, R., *J. Med. Chem.* **19**, 569–573 (1976)

[36] Leahy, D., *J. Pharm. Sci.*, **75**, 629–636 (1986)

[37] Dunn, W., III, Gigoras, S., and Koehler, M., *J. Med. Chem.* **30**, 1121–1126 (1987)

[38] SAVOL program written by R. Pearlman, Univ. of Texas, Austin; furnished courtesy of Tripos Assoc., St. Louis, MO.

[39] Nys, G., and Rekker, R., *Chim. Therap.* **8**, 521–527 (1973)

[40] Rekker, R., *The Hydrophobic Fragmental Constant*. Elsevier: New York, 1977

[41] Rekker, R., and Mannhold, R., *Calculation of Drug Lipophilicity*, VCH, Weinheim, 1992

[42] Gordon, A., Katritzky, A., and Roy, S., *J. Chem. Soc. B*, 556–559 (1968)

[43] Högberg, T., Rämsby, S., Ögren, S.-O., and Norinder, U., *Acta Pharm. Suec.* **24**, 289–328 (1987)

[44] Tsai, R.-S., Carrupt, P.-A., Testa, B., Gaillard, P., and El Tayar, N., *J. Med. Chem.* **36**, 196–204 (1993)

[45] Kamlet, M., Doherty, R., Abraham, M., Marcus, Y., and Taft, R., *J. Phys. Chem.* **92**, 5244–5255 (1988)

[46] Cramer, R., Classical and 3-D QSAR. In: *Agrochemistry and Toxicology*. ACS: Washington, DC 1995, and private communication.

10 Theoretical Calculation of Partition Coefficients

W. Graham Richards

Abbreviations

DMPC   2,3-Dimyristoyl-D-glycerol-phosphatoryl choline

Symbols

$H(r,p)$	Classical Hamiltonian ($r =$ position; p momentum)
K	Equilibrium constant
q	Molecular partition function
Q	Canonical partition function
ε_i	Energy level
n_i	Population of particles
k	Boltzmann constant
T	Temperature

10.1 Introduction

Why should anyone want to calculate a partition coefficient when the measurement of the quantity is relatively straightforward? The answer becomes clear when one realizes why partition coefficients are of such interest in medicinal chemistry: they relate to the passage of drugs through the lipid bilayers of a cell membrane. Hence, the conventional measurement of partition is between water and liquid *n*-octanol, with the latter taken as a model membrane. Clearly, this is a very poor representation of a reality where the lipid bilayer has ordered hydrocarbon chains, charged head groups and contains cholesterol, protein, and some water. Ideally we are seeking partition between water and the membrane. Experimentally, this is difficult except for very simplified model membranes.

Theoretically, the tools are all in place to compute partition coefficients between water and realistic membrane representations using the methods of statistical thermodynamics and computer simulation. At present, such a calculation is right on the edge of what is achievable with available computing resources, but in the near future very detailed questions about partition, such as the effect of nearby protein, will be asked and answered.

In this chapter the tools required will be reviewed and an indication of future directions will be given.

10.2 Statistical Thermodynamics

Statistical thermodynamics can be summarized in a none-too-rigorous way by starting with the Boltzmann distribution law which indicates the population of particles, n_i, in an energy level ε_i by

$$n_i = n_0 \exp\left(-\varepsilon_i / kT\right) \tag{1}$$

with k being Boltzmann's constant and T the temperature. The total number of particles

$$N = \sum_i n_i = n_0 \sum_i \exp\left(-\varepsilon_i / kT\right) = n_0\, q \tag{2}$$

where q is called the partition function, and

$$q = N / n_0 \tag{3}$$

It is a number with minimum value unity at absolute zero, when all particles are in level zero, $N = n_0$. At higher temperatures q indicates how the particles are spread out among energy levels and we can have $q_{\text{electronic}}$; $q_{\text{vibrational}}$; $q_{\text{rotational}}$; $q_{\text{translational}}$, etc.; a number for each type of level. Analytical expressions for partition functions are to be found in standard elementary texts [1–3]. If we have N indistinguishable particles then the canonical partition function becomes

$$Q = q^N / N! \tag{4}$$

This important number is related to the thermodynamic functions U, H, S, A and G (see Chapter 3) and hence also to equilibrium constants.

10.3 Equilibrium Constants

We have to remind ourselves that q, the inaptly named (in English) partition function, (German *Zustandssumme* – state sum) is a number indicating how particles are spread out over a set energy levels. If we consider a situation such as that shown in Fig. 1, where there is en equilibrium between a state A and another state B then the partition functions give us the equilibrium constant, K_n, in terms of the number of molecules in each state
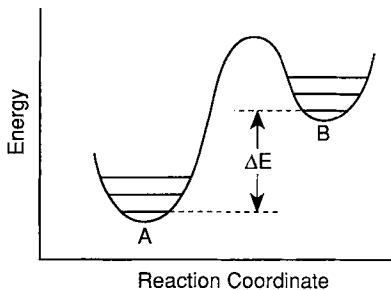
Figure 1. A schematic potential energy surface showing the separate energy of states A and B which are in equilibrium

$$q_A = n_A / n_0 \qquad q_B = n_B / n_0 \tag{5}$$

and

$$\frac{q_B}{q_A} = \frac{n_B}{n_A} = K_n \tag{6}$$

This simple result assumes that level zero is the same for both states, so that n_0 may cancel. If, on the other hand, we compute the partition function for state B from its own lowest level

$$q_B = \sum_i \exp -(\Delta E + \varepsilon_i) \tag{7}$$

and

$$K_n = \frac{n_B}{n_A} = \frac{q_B}{q_A} \exp (-\Delta E / kT) \tag{8}$$

Thus for an equilibrium

$$X = Y$$

the ratio of partition functions together with the zero point energy difference yield the equilibrium constant. So far, we have been assuming that the energy of a system is the sum of the energies of the individual molecules making up the system. That implies that there are no intermolecular interactions. If on the other hand the molecules so interact, then we are forced to treat the system classically and the energy of the system is given by the classical Hamiltonian, $H(r,p)$, the sum of kinetic and potential energies and hence a continuous function dependent on the positions (r) and momenta (p) of the constituent atoms. The partition function is then transformed from a summation over energy levels to an integral over all possible positions and momenta of particles

$$Q = \frac{1}{N!} \frac{1}{h^3 N} \int \cdots \int \exp [-H(r, p) / kT] \, dr \, dp \tag{9}$$

The coefficient $h^{-3}N^{-1}$ in front of the integral is merely a volume element (a unit of phase space, momentum times position in three dimensions, r x p) which ensures that Q is, as ever, a unitless number of which logarithms may be taken unlike a quantity with units. Integration is over all positions (r) and momenta (p) of each of the N particles.

Sadly, the canonical partition function Q is not obtainable, so in order to calculate equilibrium constants for systems of interacting particles we have to resort to a perturbation method and one which is particularly suitable for computing equilibrium constants and free energy changes.

10.4 Free Energy Perturbation Calculations

For our equilibrium

$$X = Y$$

the Helmholz free energy
$$\Delta A = -kT \ln K \tag{10}$$

$$= -kT \ln \left(Q_Y / Q_X \right) \tag{11}$$

$$= -kT \ln \left[\frac{\int \ldots \int \exp[-H_Y / kT] \, dr \, dp}{\int \ldots \int \exp[-H_X / kT] \, dr \, dp} \right] \tag{12}$$

If we now assume that the Hamiltonian of *B* is given by

$$H_Y = H_X + \Delta H \tag{13}$$

then

$$\Delta A = -kT \ln \left[\frac{\int \ldots \int \exp[-H_X / kT] \exp[-\Delta H \, kT] \, dr \, dp}{\int \ldots \int \exp[-H_X / kT] \, dr \, dp} \right] \tag{14}$$

Although this expression may seem just as formidable to evaluate it is greatly simplified when one realises that the portion outlined in the tinted box is in fact the Boltzmann distribution law and represents a probability. Hence the equation may be rewritten as

$$\Delta A = \left\langle -kT \ln \exp \left[-\Delta H / kT \right] \right\rangle_X \tag{15}$$

where $\left\langle \ldots \right\rangle_X$ indicates an ensemble average over system X which is just a simple average taken over configurations generated according to the probability density for system X [4, 5].

The generation of configurations over which averages may be taken can be achieved either using Monte Carlo or molecular dynamics methodologies. We start with a box of molecules whose intermolecular and intramolecular energy expression is a molecular mechanics potential such as can be found in the AMBER [6] or CHARMM [7] software. Either random moves are made and with Metropolis sampling to give configurations over which one averages. Alternatively, the molecular mechanics potentials provide forces and the motion of particles follows Newton's equations of motion. After equilibration, thousands of configurations are indicated over which again averages may be taken.

The perturbation, ΔH, should be small. If it is not, then a successfion of small changes are made with state A gradually changing to state B via hybrid states λ.

$$H_\lambda = (1 - \lambda) H_X + \lambda H_Y \tag{16}$$

All the terms in the molecular mechanics potential for state X thus gradually transform themselves into those appropriate for Y.

If we run our simulations in boxes which ensure constant pressure conditions we will derive Gibbs free energies ΔG.

10.5 Partition Coefficients

With this theoretical background we can now turn our attention to partition coefficients and in particular to the calculation of the differences in partition coefficients of molecules A and B between water and another solvent. We consider the cycle

$$A_{water} \quad \xrightarrow{\Delta G_3} \quad B_{water}$$

$$\Delta G_1 \downarrow \qquad\qquad \downarrow \Delta G_2 \qquad\qquad\qquad (17)$$

$$A_{solvent} \quad \xrightarrow{\Delta G_4} \quad B_{solvent}$$

The partition coefficients of A and B relate to free energy charges:

$$\Delta G_1 = -RT \ln P_A, \ \Delta G_2 = -RT \ln P_B \qquad\qquad (18)$$

so

$$\Delta \log P = \frac{1}{2.303}(\Delta G_2 - \Delta G_1) \qquad\qquad (19)$$

but $(\Delta G_2 - \Delta G_1) = (\Delta G_4 - \Delta G_3)$ since over the complete cycle ΔG must equal zero.

It is ΔG_3 and ΔG_4 which are computed using free energy perturbation. The former represents solvent molecule A being converted into solvent molecule B in water. The set-up for the simulation has periodic boundary conditions as in Fig. 2, so as to keep the solvent concentration constant.

During the course of the simulation the parameters for molecule A in the potential are mutated to those for B. In this way ΔG_3 and ΔG_4 are readily calculated while ΔG_1 and ΔG_2 are very difficult.
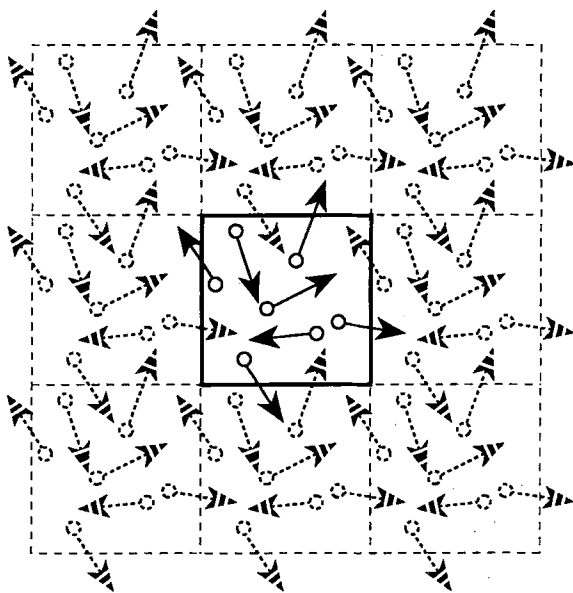
Figure 2. Periodic boundary conditions. Solid lines represent the actual system being simulated while the dotted lines indicate periodic images of that system

One example of this type of calculation was work on the differences of partition co-
efficient between methanol and ethanol partitioned between water and the solvent car-
bon tetrachloride [8]. The difference between calculated and experimental values was
less than 0.06 log P units.

Such calculations can be made absolute if molecule A is taken as a sizeless, charge-
less particle which grows by perturbation into molecule B and on into a series of other
molecules. On the other hand this is not necessarily a wise usage of computer time
since differences in partition coefficients may be all that one needs.

More important would be to have as "solvent" something more realistically repre-
senting a membrane than mere liquid n-octanol.

10.6 Membrane Simulations

The molecular dynamics technique mentioned above is also suitable for simulating the
properties of membranes, complete with head groups and surrounding water layers as
well as with incorporated cholesterol or protein.

In particular we have taken DMPC (2,3-dimyristoyl-D-glycero-1-phosphatoryl cho-
line) as a typical membrane in its L_α phase [9, 10]. Fig. 3 shows the basic unit and Fig. 4
the dynamics frame after 400 ps of dynamics on all the atoms.
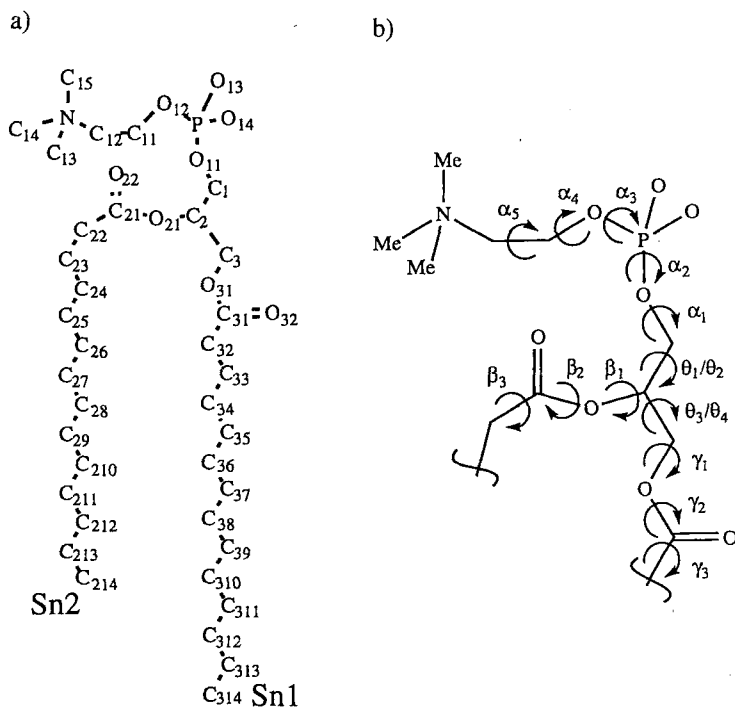
Figure 3.   Schematic structure of DMPC illustrating a) the atom numbering, and b) some the
variable torsion angles
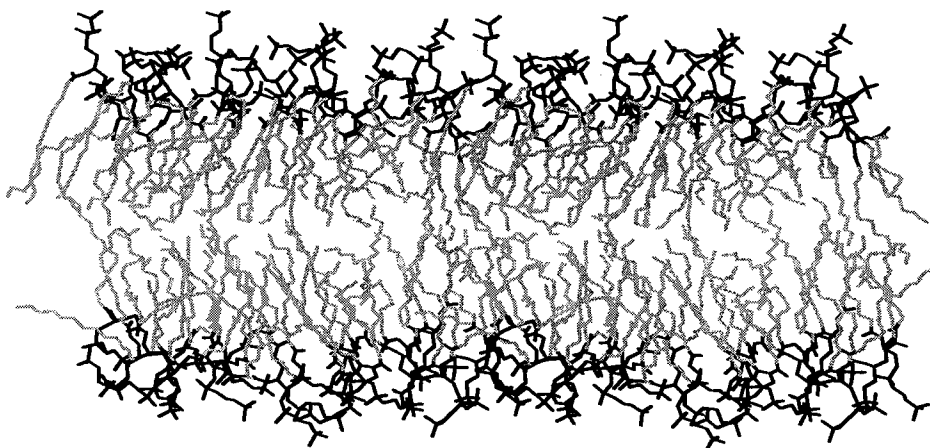
Figure 4. A snapshot structure of a DMPC membrane after 400 ps dynamics. Water molecules are not shown; head groups are colored black and the rest of the lipid is colored gray

The experimental results which confirm that this is a realistic model of a membrane include order parameter profiles from NMR quadrapolar splittings and neutron diffraction studies. Work currently in progress involves placing cholesterol molecules in the membrane as in Fig. 5 and also protein.

In any of these models it would be possible to insert solutes at a specific position and to compute differences in partition coefficients just as was done for the alcohols in carbon tetrachloride above. Calculations of this type should indicate just how drugs may tend to have preferential locations within membranes and diffuse to membrane-bound receptors.
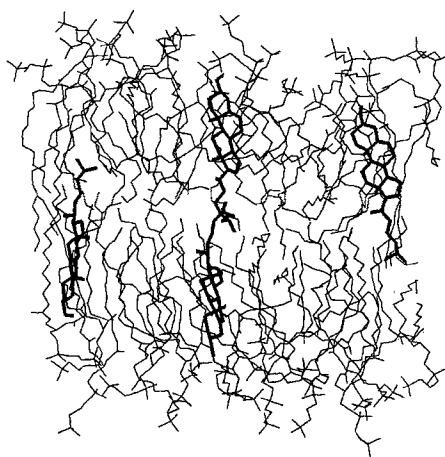
Figure 5. The structure of a DMPC membrane containing cholesterol molecules (in black)

10.7 Future Outlook

The work outlined here is not the way to go if one merely wants partition data on thousands of compounds. It is, however, the direction to go if one wants to ask very specific questions and to have details of real membranes included.

These computations are not inexpensive. Indeed, some of the work described here has taken hours of CPU time on the world's largest supercomputers. Nonetheless, if one wants understanding at the molecular level then this technique may provide it.

References

[1] Gasser, R. P. H., and Richards, W. G., *Entropy and Energy Levels.* Oxford University Press: Oxford, 1986
[2] Gasser, R. P. H., and Richards W. G., *Introduction to Statistical Thermodynamics*, World Scientific Publ. Singapore, 1995
[3] Chandler, D., *Introduction to Statistical Thermodynamics.* Oxford University Press: Oxford 1987
[4] Zwanzig, R. W., *J. Chem. Phys.* **22**, 1420–1431 (1954)
[5] Reynolds, C. A., King, P. M., and Richards, W. G., *Molec. Phys.* **76**, 251–275 (1992)
[6] Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. *J. Comput. Chem.* **7**, 230–243 (1986)
[7] Brooks, S. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M., *J. Comput. Chem.* **4**, 187–191 (1983)
[8] Essex, J. W., Reynolds, C. A., and Richards, W. G., *J. Chem. Soc. Chem. Commun.* 1152–1154 (1989)
[9] Essex, J. W., Hann, M. M., and Richards W. G., *Phil. Trans. R. Soc. Lond. B.* **344**, 239–260 (1994)
[10] Robinson, A. J., Richards, W. G., Thomas, P. J., and Hann, M. M., *Biophysics J.* **67**, 2345–2354 (1994)

11 Cellular Automata Model of Partitioning Between Liquid Phases

Lemont B. Kier and Chao-Kun Cheng

Symbols

$P_B(W)$	Water-water breaking probability
$P_B(L)$	Ingredient-ingredient breaking probability
$P_B(WL)$	Water-ingredient breaking probability
$L_1, L_2 \ldots$	Ingredient designations
$J(W), J(L), J(WL)$	Corresponding joining parameters for the above probabilities
$f_0(L)$	Fraction of L molecules not bound to any other L molecules
$f_4(L)$	Fraction of L molecules bound to four other L molecules
G_U	Ratio of switching probabilities of a molecule above another relative to movement to a vacant cell
G_D	Same as GU for a molecule below another

11.1 Introduction

The traditional approach to the study of aqueous solution phenomena has been based on the reductionist philosophy in which the whole is considered to be the sum of the parts and the route to understanding a system is to analyze it to identify the parts. Understanding is approached by summing these parts to create a model of the whole. As an example, we dissect a solution into isolated molecules, then employ various methods to evaluate their electronic and steric attributions. From there we assemble this information into models of systems using linear combinations. At the other end of the size scale, we observe thermodynamic properties of vast ensembles of molecules and then attempt to ascribe some average contribution to the individual molecular ingredients, identified again by a reductionist process.

This reductionist philosophy has dominated science in general and the study of solution phenomena in particular for a long time. In contrast, there has evolved in recent years an alternative way of viewing nature based on the belief that the whole is, in many cases, much more than the sum of the parts. When ingredients interact or transact, there emerges in the whole a set of properties not clearly recognizable as additive contributions of ingredients. There is formed a complex system that possesses emergent properties. The ingredients of a system are identified by analysis or reduction, but an equally important process must then follow. That process is the dynamic synthesis of a model that mirrors enough of the emergent properties of the complex system so as to provide for some understanding. The subjects of complexity and emergent properties in drug research have recently been introduced by Kier and Testa [1].

In our laboratory we have long recognized the essential part that water plays in drug phenomena, including drug-effector events. We view it as an active ingredient in the trio: drug, effector, water. It is essential then, that we add to our understanding of this precious fluid and the complex systems that form when solutes enter its embrace. The complex nature of water and solutions has been recognized in recent times and has prompted a few investigators to attempt to derive models as non-linear combinations of ingredients. In particular, we have witnessed the growth of Monte Carlo and molecular dynamics simulations of water that have added to our understanding of its complex character [2–6]. While molecular dynamics is a step in the direction towards a better understanding of water as a complex system, some difficulty lies in the inability to model large numbers of molecules with significant diversity. The vast amounts of computer time required coupled with the assumptions of force fields based on two-molecule interactions produces limitations to the general understanding of the variety of complex solution phenomena of interest to the drug research scientist. This has led us to embark upon the study of water and solution phenomena using an alternative method of dynamic synthesis called cellular automata.

11.2 Cellular Automata

11.2.1 The Model

Cellular automata are dynamical systems that are discrete in space, time, and state and whose behavior is specified completely by rules governing local relationships. It is an attempt to simplify the often numerically intractable dynamics simulations into a set of simple rules that mirror intuition and that are easy to compute. As an approach to the modeling of emergent properties of complex systems it has a great benefit in being visually informative of the progress of dynamic events. From the early development by von Neumann [7] a variety of applications ranging from gas phenomena to biological applications have been reported [8]. We have viewed cellular automata as an opportunity to advance our understanding of water and solution phenomena and have embarked upon a series of studies with this goal in mind [9–11].

Our model is composed of a grid of spaces called cells on the surface of a torus to remove boundary conditions. Each cell i has four tesselated neighbors, j, and four extended neighbors, k, in what is called an extended von Neumann neighborhood (Fig. 1 d). Each cell has a state governing whether it is empty or is occupied by a water or other molecules. The contents of a cell move, join with another occupied cell or break from a tesselated relationship according to probabilistic rules generated as random numbers. These rules are established at the beginning of each simulation. The rules are applied one after another to each cell at random, the complete application of the rules to all cells constituting one iteration. The rules are applied uniformly to each cell and are local, thus there is no action at a distance. The cellular automata model is thus kinematic, asynchronous, and stochastic. The initial conditions are random; hence, they do not determine the ultimate state of the cells, called the configuration. The same initial conditions do not yield the same set of configurations after a certain number of iterations expect in some average sense. The configurations achieved after
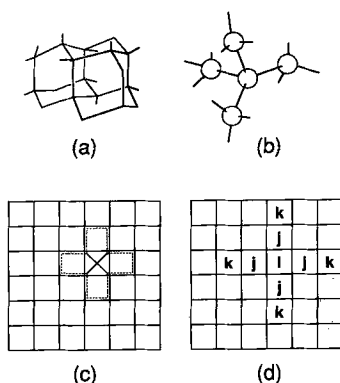
Figure 1. (a) The hexagonal structure of ice taken as a model of the trajectories of individual molecules; (b) an isolated fragment depicting the probable directions a central water molecule may take; (c) the trace of (b) on a plane, coincident with the von Neumann neighborhood; (d) the extended von Neumann neighborhood

many iterations reach a collective organization that possesses a relative constancy in appearance and in reportable counts of attributes. What we observe and record from the cellular automata simulations are emergent attributes in a complex system.

11.2.2 The Molecular System

It must be made clear just what the cells, the configurations generated, and the cellular automata models represent. This is important in order to derive any understanding of the results of a simulation and to dispel misunderstanding based on direct comparisons with molecular methods. A cell with a state value encrypting occupation by a particular object is not a model of a molecule with specified electronic and steric features. It is a statement of the existence in space-time of an object that has certain rules governing its trajectories. These rules govern the transaction of the object in the cell with all other objects in its tesselated environment called a von Neumann neighborhood. The object, in our studies, is interpreted to be a molecule, specified or not, that is defined only by its state and transition functions. Those rules are considered to be sufficient to allow the dynamics to proceed. Electronic and topological characterizations are considered to be subsumed into these rules. The transactions in the dynamics results in configurations after several iterations that embrace more than just one molecule. There is dynamically simulated the emergent attributes of a *molecular system*. We define a molecular system as the minimum number of molecules necessary to model a phenomenon which is recognizable as emergent behavior. This emergent behavior is epitomized by visual patterns or calculable attributes such as average counts or sizes of configuration features. The molecular system is intermediate between molecular level and bulk phase models of systems and is modeled with molecular dynamics or cellular automata. Molecular level phenomena are modeled with molecular orbital theory, topological indices, or fragment methods. At the bulk level we use descriptions based on statistical and thermodynamic methods.

Many solution phenomena should be studied using molecular system models in order to understand the processes whereby single molecules achieve the configurations measurable as bulk properties. Example include crystal dissolution, partitioning

among immiscible liquids, drug-receptor encounters, and solute binding phenomena. With this background in mind we proceed with the use of cellular automata dynamics to model aqueous solution phenomena.

11.2.3 The Dimensional Relationship in Cellular Automata Models

In our model the trajectory of a liquid water molecule is assumed to approximate the paths of the hexagonal ice lattice (Fig. 1 a). Each vertex in that figure denotes a water molecule while each edge denotes a bonding relationship. This three-dimensional network can be dissected into a contiguous series of tetrahedral fragments of five vertices each (Fig. 1 b). Some or all of these vertices in each fragment may be representative of a water molecule. The trace of each of these tetrahedral fragments can be mapped onto a two dimensional grid (Fig. 1 c). We equate this mapping with a cellular automata von Neumann neighborhood. The cellular automata transition functions operate on the central cell, i in each von Neumann neighborhood (Fig. 1 c). The rules in our studies composing these transition functions operate on the states of cells i, j, and k in the extended von Neumann neighborhood (Fig. 1 d). In our model the rules are executed for each cell in the cellular automata grid asynchronously and at random over the grid. As a consequences, the new configuration for each cell, i, and its neighborhood is derived independent of all other cells remote from cells i, j, and k, in Fig. 1 d. The configuration of the system, achieved after all cells respond in random order to the rules, constitutes one iteration. This configuration is a composite of the collective configurations achieved in all of the von Neumann neighborhoods. Each of these neighborhoods is a two-dimensional mapping of a tetrahedral fragment of the original three-dimensional model. Our model is a representation of the configuration of a three-dimensional system on the basis of it being a summation of discrete, orthogonal events occurring within that system. We can lay claim to the conjecture that the dimensionality of our model is fractional between two and three. Other studies using this approximation have been reported [12, 13].

11.2.4 The Rules

In previous reports we have described the specifics of the calculations used [9–11]. We repeat these here in abbreviated form but include the new rules governing the gravity influence on the modeling of immiscible liquids. There are two kinds of cell occupants, active and boundary molecules. The active molecule can move so that the state of a cell occupied by an active molecule may change; however, a boundary molecule cannot move, thus, the state of a cell occupied by a boundary molecule will never change. In this paper, cells in two selected rows of the grid are occupied by boundary molecules. It transforms the torus model into a cylinder model. Those two rows occupied with boundary molecules serve as the top and the bottom of the cylinder. This introduces the notion of vertical direction to the model, therefore, a "gravity" concept can be incorporated into the model.

Four parameters are adopted for our model to govern the probabilities for moving molecules in the grid. The breaking probability, P_B, used in our previous studies is the

probability for an active molecule to break away from the molecule at j cells when there is exactly one occupied j cell (see Fig. 1 d). The value for P_B lies in the closed unit interval. The second parameter, J, describes the movement of the active molecule at cell i toward or away from the molecule at a k cell in the extended von Neumann neighborhood (Fig. 1 d) when the intermediate j cells is vacant. It represents the ratio of the probability that a molecule at i cell will move toward an occupied k cell while the intermediate i cell is vacant, and the probability that a molecule at j cell will move toward a vacant k cell while the intermediate j cell is vacant. J is a positive real number. When $J = 1$, it indicates that the molecule i has the same probability of movement toward or away as for the case when k is empty. When $J > 1$, it indicates that i has a greater probability of movement away from k than when k is empty. When $J < 1$, it indicates that i has a greater probability of movement toward k relative to when k is empty.

The third parameter, G_u for cell content i, describes the movement of the active molecule at cell i relative to an active molecule in the j cell above it. G_u represents the ratio of the probability that an active molecule at i cell will switch places with the molecule in the j cell above it, and the probability that an active molecule at i cell will move toward a vacant k cell while the intermediate j cell is vacant. G_u is a non-negative real number. When $G_u = 1$, it indicates that the active molecule i has the same probability to switch with the active molecule in the j cell above it as for the case when the molecule will move toward a vacant j cell when k is empty. The cases in which $G_u > 1$ and $G_u < 1$ have a similar specification. When $G_u = 0$, it indicates that switching is not allowed with respect to the types of molecules involved. The fourth parameter, G_D for cell content i, describes the movement of the active molecule at cell i which switches cells with the active molecule in the j cell below it. This parameter is applicable only when the j cell below it is occupied with an active molecule. It represents the same information as G_u except this measures the switching probability with an active molecule below cell i.

11.3 Models of Solution Phenomena

11.3.1 A Model of Water

Our objective has been to test systematically the ability of cellular automata simulations to create models having some linkages to reality. We summarize here our previous work on water and solution phenomena as a prelude to our description of our recent work on immiscibility and partitioning. Our initial simulation of liquid water [9] revealed a configuration resembling the extended network patterns shown by molecular dynamics simulations [14,15]. By evaluation the fraction of water molecules unbound to other water molecules, $f_0(W)$ and comparing this with earlier estimates and predictions, we were able to draw a close parallel between the value of $P_B(W)$ and the temperature, $T(\,^\circ C) = 100\, P_B(W)$. With this relationship we established two equations that closely related the average water cluster size to the viscosity and the fraction of water molecules unbounded to other water molecules, f_0, to the vapor pressure.

11.3.2 A Model of a Solution

In this study rules were introduced for a second substance, L, in addition to water, W [10]. Designing a simulation experiment to model a small concentration of solute L in water. W, produced several configurations in agreement with other models and experiments. Increased values of $P_B(W)$ produced an increase in the $f_0(L)$ in spite of the fact that no rule including an L parameter has changed. This simulation can be interpreted as a model of an increased solubility of L upon heating the water. The second observation from our simulation experiments dealt with changes in the concentration of L in the water. It was found that by simulating an increase of the concentration of L in water there was produced a change in attributes of water relating to its structure. Specifically, there was a decrease in the average number of water-to-water tessellations, interpreted as the average number of hydrogen bonds, and an increase in $f_0(W)$. This corresponds to the disruption of the water patches in the presence of an increasing solute concentration.

The third observation concerned the breaking parameter relating the water and the solute, $P_B(W)$. With low values of this parameter, the simulations produced a relatively low count of the average number of molecules bonded to four other water molecules. Concurrently, there is revealed a relatively high value of waters not bonded to other water molecules. The graphics reveal that most of the solute molecules are within the patches of water. This simulation produces a configuration which may be interpreted as a solution of a polar solute, L. In contrast, when the $P_B(WL)$ is high, there is a relatively high count of bonded water molecules. The graphics reveal that most of the solute molecules are within the water cavity areas. At this end of the $P_B(WL)$ range we can say that the rule is characteristic of nonpolar solutes. A conclusion from these simulation experiments was that the $P_B(WL)$ rule influences the solute polarity while the emergent effect on the water structure corresponds to the prevailing view that nonpolar solutes increase the structure of the solvent, water. This has been termed, the hydrophobic effect.

11.3.3 A Model of the Hydrophobic Effect

Further simulations have confirmed the configurations produces by changing the $P_B(WL)$ rule [11]. This establishes a definitive influence of this rule on a recognizable attribute which we can liken to the hydrophobic effect. This opens up the possibility of being able to rationally select rules in order to direct simulations toward a specific outcome. We have explored further the influence of other rules on solubility, polarity, and structuration with the outcome of having a clearer impression of their influences on simulation attributes.

11.3.4 A Model of Dissolution

In a recently submitted study by Kier and Cheng, the emergent properties of solute block disruption, solute dissolution, and solute diffusion through the solvent have been dynamically simulated using cellular automata. The solute block disruption was found to be dependent upon the transition functions describing the relationship of one solute to another. This disruption was found to be due to the presence of water cavities

penetrating the solute block rather than water molecules. This occurred in almost very variant of the parameters. The dissolution was found to be strongly related to the solute-solute and solute-water relationship. The diffusion is influenced primarily by the solute-water transition function. The processes are all enhanced in the simulation by the increase in water temperature. A rich variety of configurations, interpretable as variations in the way a block of solute may disintegrate and pass into solution, were simulated when the transition functions were varied.

11.4 A Cellular Automata Model of Immiscibility

11.4.1 Immiscible Liquids

The two-phase liquid system has aroused considerable interest for many years because of its ability to quantify solutes on the basis of their relative partitioning between the two liquids. This partitioning behavior has come to be identified with a number of biological phenomena such as drug absorption and binding, membrane passage, and some aspects of intermolecular interaction. This has led to the common practice of quantifying the propensity of a molecule to partition selectively between two immiscible liquids. This propensity called *lipophilicity*, expressed quantitatively as the partition coefficient, has become a prominent entry in the rubric of properties of molecules, especially those of biological interest. Considerable effort has been made to identify the salient features of molecules that influence the partitioning between the liquids of an immiscible system. These reductionist approaches are conducted in an effort to identify the ingredients responsible for the partitioning behavior followed by the effort to model the property with a linear combination of information from these ingredients. We have asserted in our previous studies, that water, solutions, and certainly a two-phase system with a solute is a complex system. Understanding may then arise from a dynamic synthesis of the emergent properties of these complex systems.

11.4.2 A Model of Immiscible Systems

We approach this study, as we have earlier, with the goal of simulating the complex system as a configuration emerging from the parts in a dynamic synthesis. Our first effort is directed towards the modeling, using cellular automata, of an immiscible system. Several computer simulations of the liquid-liquid interface have been reported. Simulations using Monte Carlo [16, 17] molecular dynamics [18, 19] a lattice model [20] and a review [21] have focused attention on hydrophobic interactions and water-membrane structure. The most recent molecular dynamics simulation by Benjamin [22], studied the water-dichloroethane interface.

The dynamic simulation that we employ takes into account the relative influence of gravity upon immiscible two liquids. For purposes of formulating rules and executing transition functions, we must encode the gravity influence within the specifications of single molecule-occupied cells in the simulation. Further, we must resort to grid surface that possesses a boundary condition that produces an up and down relationship. We use the surface of a cylinder for our grid as described in section 2.4.

11.4.3 An Immiscible Liquid Simulation

We have systematically studied and have optimized the choice of parameters for the simulation of an immiscible system. One ingredient is designated as water using the parameters established in earlier studies. For this study we use $P_B(W) = 0.3$ and $J(W) = 1$. For the other liquid, we use $P_B(L_1) = 0.5$ and $J(L_1) = 0.5$. For the cross-terms, we invoke a nonpolar character to the liquid, L_1 by using $P_B(WL_1) = 0.9$ and $J(WL_1) = 0.25$. The gravity terms are $G_D(W) = 0.25$, $G_u(L) = 0.25$ and $G_D(L) = 0.20$, $G_u(W) = 0.20$; thus, there is a modest preference for water to favor a position below the liquid, L_1, in a vertical encounter in the von Neumann neighborhood. The initial conditions are a random mix of 1050 water-occupied cells and 1050 liquid (L_1)-occupied cells in a grid of 3025 cells. At about 12 000 iterations, the fractions of ingredients in their respective positions in the grid have become relatively stable. This distribution profile reveals an increased concentration of each liquid in the other, near the interface. This increase in concentration near the interface is seen to be due to the presence of stacks of liquid-occupied cells arising from the bulk of each liquid, directed into the other liquid, (see Fig. 2). They move laterally and vertically as the dynamics proceeds. These stacks are very similar to what others have described from dynamic simulations as capillaries [16–20] or fingers [22], arising from each liquid and penetrating into the other phase.

Figure 2.   The partial separation of two immiscible liquids showing the elongated stacks of each liquid. The water is black, the nonpolar liquid is gray and the cavities are white

We find this agreement with other dynamic simulations quite significant in that it confirms our confidence in the ability of this cellular automata experiment to model at least some aspects of two-solution behavior.

In our studies we find that the stacks of liquid (capillaries) extend into the other liquid by as much as ten cells, depending upon the parameters chosen. We further find that in the early stages of the dynamics prior to the achievement of a stable configuration, there are formed isolated stacks of liquid W above the interface and liquid L_1 below the interface (Fig. 3). These gradually drift to the interface, forming connections with liquids of the same composition. These isolated stacks are roughly 5–10 cells in height and 2–4 cells across, although these dimension are rough averages, there being a rapid change in their structure and position. This configurational patterns has not been detected in other dynamic simulations because each of those began with two homogenous liquids abutting each other with no mixing simulation. Our simulation more completely models the mixing phenomena of any two immiscible liquids, since experimentally there is an inescapable amount of turbulence at any newly formed interface that is going to produce some amount of heterogeneity. We should note, however, that the lack of direct knowledge about the structure of the interface and the difficulty of experimentally detecting the events is a situation facing investigators in this field [21].

Figure 3. The interface of two immiscible liquids

11.5 A Model of Partitioning between Immiscible Liquids

With our dynamic model simulating the behavior of immiscible liquids, we can proceed to introduce a small number of solute molecules, L_2, to follow their emergent behavior as a result of probabilistic rules. Using the values of rules for W and L_1 described in the previous section, we have systematically varied the rules governing the transitions L_2L_2, L_1,L_2 and L_2W. These values are shown in Table 1 along with the codes to simplify descriptions of the parameter sets. The initial conditions are 1025 cells occupied by water, W, 1025 cells occupied by liquid L_1, and 50 cells occupied by solute L_2. We have elected not to use any gravity term relating L_2 to W or L_1. Linse [16] has addressed this issue and recognizes the near-negligible influence on molecules moving a few molecular diameters. Dynamic simulation runs of various lengths were made; in each case the fraction of solute, L_2, moving to the upper, L_1, liquid phase was fairly constant after about 20 000 iterations. Using the parameter combinations encoded in Table 1, we recorded the fraction of L_2 solute molecules in the upper phase, shown in Table 2. About half of these parameter combination produced results favoring the partitioning into the upper phase, L_1. The partition coefficient is influenced in these simulations by a com-

Table 11.1. Parameters for solute partitioning studies

Breaking parameters	Value	Joining parameter	Value	Code
$P_B(W)$	0.3	$J(W)$	1.00	
$P_B(WL_1)$	0.9	$J(WL_1)$	0.25	
$P_B(L_1)$	0.5	$J(L_1)$	0.50	
$P_B(WL_2)$	0.3	$J(WL_2)$	1.00	a
$P_B(WL_2)$	0.7	$J(WL_2)$	1.00	b
$P_B(L_1L_2)$	0.3	$J(L_1L_2)$	1.00	c
$P_B(L_1L_2)$	0.7	$J(L_1L_2)$	1.00	d
$P_B(L_2)$	0.3	$J(L_2)$	1.00	e
$P_B(L_2)$	0.7	$J(L_2)$	1.00	f

Table 11.2. Partitioning using various parameter sets

Parameter code set[a]	Partition coefficient, P[b]
bcd	3.17
bde	2.33
ace	2.13
bce	1.78
acf	1.08
bdf	0.85
ade	0.32
adf	0.28

[a] Parameter combinations described in Table 1.
[b] Ratio of solutes in upper, nonaqueous layer to those in lower, aqueous layer.

Figure 4. The initial stages of immiscible liquid separation showing the preferential occupancy of the nonpolar stacks with the solute (cross-hatched cells)

bination of parameters reflecting the self-association of the solute, $P_B(L_2)$, and the affinity of the solute for water, $P_B(WL_2)$, plus the non aqueous solute, $P_B(L_1L_2)$. Further systematic variation of the parameters may refine this conclusion more quantitatively, however, the basic understanding of the simulation is at hand from these results.

The statistical results are revealed from counts of molecules after certain numbers of iterations (Table 2). A detailed understanding of how the partitioning process is occurring requires a continuous visual observation of the dynamics over an extended period of time. These observations have revealed an interesting and heretofore unrecorded prediction about what the sequence of events may be between an initial random mixture of solvents and solute, as in a shake-flask experiment, and the subsequent state of these three ingredients after a partitioning has presumed to occur. From a random initial state, the three ingredients slowly move toward a configuration in which there are small stacks of L_1 in W and W in L_1. We described this configuration in section II 4.3. The size and duration of these stacks is a function of the parameter sets. We observe that the solutes, L_2, preferentially partition into the L_1 or W stack depending upon the parameters chosen. The stack favored is the same liquid phase in which L_2 ultimately preferentially resides. We can say that the partitioning process begins early with the appropriate liquid stacks capturing many of the solute molecules well before an interface has formed. These stacks, one type with a significant concentration of so-

Figure 5. The organization of a positional relationship among the two liquids with the solute predominantly in the nonpolar liquids

lute L_2 molecules, slowly move toward the now-forming interface. The stacks ultimately join their companions to form dynamic interface with the capillary structures as we described in section II.4.3. The solute molecules, L_2, are now free to move through these capillaries and join the general concentration of solute in the, preferred solvent. This sequence of events is depicted in the series of Figs. 4–6.

11.6 Conclusion

In a series of studies we have developed models using cellular automata dynamic simulations of water, solution phenomena including the hydrophobic effect, and solute dissolution and diffusion. In this report studies are extended to the creation of a dynamic model simulating a two-liquid interface and a model in which a solute may partition between these two immiscible liquids. It has been encouraging along the way to find significant agreement between our recorded attributes and configurations, and experimental and simulation evidence from other studies. It appears that cellular automata simulations, in spite of their simplicity and ease of use, may provide an approach to the understanding of complex systems of chemical and biological interest.

Figure 6. A well-developed model of the immiscible liquid interface showing the preference of the solute for the nonpolar liquid. Note the extensions of each liquid in the other phase

To restate here the general nature of these simulations: they are not a dissection of a phenomena into discrete parts for analysis; they reveal the nature of the transition from one state to another with an evanescent series of configurations; they create a model of a process, not the end-points. Since all of nature is dynamic, a claim may be made that these models contribute to our understanding of some vital phenomena of interest in drug research. We shall continue along these lines to explore other phenomena to enrich our understanding.

Acknowledgements

References

[1] Kier, L. B., and Testa, B., *Adv. Drug Res.* **26**, 1–43 (1995)

[2] Geiger, A., Stillinger, F. H., and Rahman, A., *J. Chem. Phys.* **70**, 4186–4192 (1979)

[3] Mezei, M., and Beveridge, D. L., *J. Chem. Phys.* **74**, 622–632 (1981)

[4] Rossky, P. J., and Karplus, M., *J. Chem. Soc.* **101**, 1913–1921 (1979)

[5] Jorgensen, W. L., Gao, J., and Ravimohan, C., *J. Phys. Chem.* **89**, 3470–3477 (1985)

[6] Plummer, P. L. M., *J. Mol. Struct.* **237**, 47–61 (1991)

[7] Von Neumann, J., *Theory of Self-Reproducing Automata*, compiled and edited by Burks, A. W., University of Illinois Press, Urban, (1966)

[8] Ermentrout, F. B., and Edelstein-Keshet, L., *J. Theor. Biol.* **160**, 97–133 (1993)

[9] Kier, L. B., and Cheng, C.-K., *J. Chem. Inf. Comput. Sci.* **34**, 647–652 (1994)

[10] Kier, L. B., and Cheng, C.-K., *J. Chem. Inf. Comput. Sci.* **34**, 1334–1337 (1994)

[11] Kier, L. B., Cheng, C.-K., Testa, B., and Carrupt, P.-A., *Pharm. Res.* **12**, 615–620 (1995)

[12] Blumberg, R. L., Shlifer, G., and Stanley, H. E., *J. Phys. A: Math. Gen.* **13**, L 147–150 (1980)

[13] Brodsky, M. H., *Bull. Am. Phys. Soc.* **25**, 260–265 (1980)

[14] Rice, S. A., and Sceats, M. G., *J. Phys. Chem.* **85**, 1108–1119 (1981)

[15] Stanley, H. E., and Teixeira, J., *J. Chem. Phys.* **73**, 3404–3422 (1980)

[16] Linse, P., *J. Chem. Phys.* **86**, 4177–4187 (1987)

[17] Gao, J., and Jorgensen, W. L., *J. Phys. Chem.* **92**, 5813–5822 (1988)

[18] Meyer, M., Mareschal, M., and Hayoun, M., *J. Chem. Phys.* **89**, 1067–1073 (1988)

[19] Carpenter, I. L., and Hebre, W. J., *J. Phys. Chem.* **94**, 531–536 (1990)

[20] Smit, B., Hilbers, P. A. J., Esselink, K., Rubert, L. A. M., van Os, N. M., and Schlijper, A. G., *J. Phys. Chem.* **95**, 6361–6368 (1991)

[21] Hanna, G. J., and Noble, R. D., *Chem. Rev.* **85**, 583–598 (1986)

[22] Benjamin, I., *J. Chem. Phys.* **97**, 1432–1445 (1992)

12 The Molecular Lipophilicity Potential (MLP): A New Tool for log *P* Calculations and Docking, and in Comparative Molecular Field Analysis (CoMFA)

Pierre-Alain Carrupt, Patrick Gaillard, Frédéric Billois, Peter Weber, Bernard Testa, Christophe Meyer and Serge Pérez

Abbreviations[a]

APC   Antigen-presenting cells
CoMFA   Comparative molecular field analysis
HEL   Hen egg lysozyme
MD   Molecular dynamics
MHC   Major histocompatibility complex
MLP   Molecular lipophilicity potential
PLS   Partial least square
QSAR   Quantitative structure-activity relationships
RMS   Root mean square
SAS   Solvent-accessible surface area
TcR   T-cell receptor
3D   Three-dimensional

Symbols

K_i Dissociation constant of a receptor-inhibitor
q^2 Cross-validated correlation coefficient
$\Delta \log P$ Difference between lipophilicity indices measured in two different solvent systems
r Correlation coefficient
α H-bond donor acidity
ΣMLP^+ Parameter describing the polar part of a molecule
ΣMLP^- Parameter describing the apolar part of a molecule

12.1 Computational Approaches to Lipophilicity

12.1.1 Introduction

Lipophilicity, which expresses steric and polar intermolecular interactions between a solute and a biphasic liquid system [1,2], is an important molecular property largely

---

[a] For additional symbols and abbreviations see Chapter 4.

used in drug design. In the first stages of drug design, medicinal chemists must offer educated guesses as to the pharmacokinetic and pharmacodynamic behavior of compounds to be synthetized. Predicted lipophilicity is one of the tools on which medicinal chemists can rely in making such guesses.

Several approaches have been developed to allow the direct calculation of partition coefficients from molecular structure [3]. The most popular methods for calculating log P_{oct} values are based on the additivity of fragmental constants of lipophilicity, as described by Leo in this book (Chapter 9).

12.1.2 Limits of Fragmental Systems

Despite their great interest, additivity rules have fundamental limitations that did not remain undetected for long. The main discrepancies between predicted and measured values are due to the effects of intramolecular interactions as described by Testa and coworkers (Chapter 4). Thus, the many solutions proposed to take intramolecular effects into account are based on the use of a large number of corrections factors. However, too many correction factors not only render additive methods difficult to handle without powerful computational tools (e.g., the CLOGP software [4], cf. Chapter 9), but also prevent the intramolecular origin of the deviations to be understood.

All popular additive methods (e.g., the fragmental systems of Rekker [5] or Leo and Hansch [6], and the atomic systems of Broto and Moreau [7] or Ghose and Crippen [8]) are based on a two-dimensional description of molecular structure. Due to the significance of molecular rigidity and flexibility in modulating physico-chemical interactions underlying biological recognition processes, fast and reliable methods are needed in drug design to calculate lipophilicity also taking the tri-dimensional molecular topology into account.

12.2 The Molecular Lipophilicity Potential (MLP): a Tool to Compute Partition Coefficients from 3D Structures

12.2.1 Derivation of the MLP

The Molecular Lipophilicity Potential (MLP) offers a quantitative 3D description of lipophilicity and defines the influence of all molecular fragments on the surrounding space. At a given point in space, the MLP value represents the result of the intermolecular interactions encoded by the lipophilicity of all fragments in the molecule. Thus, the MLP is a potential of lipophilicity, i.e., the relative affinity of a solute for two immiscible solvents. Unlike the electrostatic potential, it is not necessary to have a probe to reveal the MLP in space, since all the necessary information is contained in implicit form in log P values.

Two components are needed to calculate the MLP, namely a fragmental system of lipophilicity [7–9], and a distance function [10, 11] describing the variation of lipophilicity in space. Based on these components, the MLP at the space point k is expressed by the following general equation:

$$\text{MLP}_k = \sum_{i=1}^{N} f_i \times fct(d_{ik}) \tag{1}$$

where k = label of a given point in space, i = label of a fragment, N = total number of fragments in the molecule, f_i = lipophilic constant of fragment i, fct = distance function, d_{ik} = distance between the fragment i and the space point k.

The MLP approach was first introduced by Audry et al. [12] who used the fragmental system of Rekker and De Kort [5] and a hyperbolic distance function. Their second version was based on the atomic fragmental system of Broto et al. [7] and an exponential distance function. Fauchère et al. [10] used the fragmental system of Rekker and De Kort [5] and/or Hansch and Leo [6] and an exponential distance function to study the qualitative variation of the spatial distribution of lipophilicity. Other approaches use the atomic fragmental system of Ghose and Crippen [8] and an hyperbolic distance to study the variation of lipophilicity on the molecular surface (Furet et al. [13]), or the fragmental system of Hansch and Leo (multiplying these values by the solvent-accessible surface (SAS) of each atom) and an exponential distance function like the HINT software (Kellogg et al. [14]).

The MLP presented here is based on the system of atomic increments of Broto et al. [7], and on a modification of the exponential function proposed by Fauchère and co-workers [10]:

$$\text{MLP}_k = \sum_{i=1}^{N} f_i \times e^{-d_{ik}/2} \tag{2}$$

The distance function should describe the decrease in space of the lipophilic contributions of atoms: at the atom core, the MLP value must be maximal and at large distances, it must approach zero. The exponential function proposed by Fauchère et al. [10] or the hyperbolic function proposed by Audry et al. [11] do not appear adequate near the SAS due to their low variation (less than 10% between c. 2.6 Å and 4 Å). A slightly modified exponential function ($e^{-d/2}$) overcomes this limitation. In addition, we have limited the distance function with a cut-off of 4 Å to avoid exaggerated effects at large distances.

12.2.2 Back-Calculation of Partition Coefficient

If spreading out lipophilicity over a given surface is valid, the original log P values should be recovered by integration of the MLP over that surface. This approach had led to a new method for predicting lipophilicity from the 3D structure of a molecule. The SAS appears as an adequate space to integrate the MLP back to log P values [15, 16]. However, log P calculations can be made only if a preliminary correlation with experimental lipophilicity data is established.

For this purpose, two parameters were derived from the MLP values generated on the water-accessible surface area of a molecule. These parameters (ΣMLP^+ and ΣMLP^-) are the partial summations of the positive and negative MLP values, respectively, and represent the *hydrophobic* and the *polar* parts of the molecule, i.e., the regions on the SAS where positive and negative atomic increments to lipophilicity are expressed [17]. A qualitative visualization of the MLP on the water-accessible surface

area has been made possible by color-coding (from red for polar parts to blue for hydrophobic parts) as illustrated by the MLP of phenolphthalein glucuronide presented on the cover of the present book.

The existence of a general correlation between experimental and calculated log *P* values for 114 compounds [18] demonstrates that our methodology allows a direct calculation of log *P* values from 3D structures without using any correction factors for intramolecular effects. It should be noted that the back-calculation of log *P* from the MLP gives an estimate of partition coefficients as precise as the one based on fragmental systems. Since the integration space (SAS) is sensitive to the 3D structure of the molecule, log *P* values calculated from the MLP may allow to explicate the effects of molecular conformation on lipophilicity [18].

12.3 The MLP: a Tool to Explore Conformational Effects on Lipophilicity

As described in Chapter 4, the effects of conformational changes on experimental lipophilicity are well known and can be approximated by the difference between the experimental log *P* value and the log *P* predicted with an additive fragmental system. However, the modeling of these effects prior to synthesis and experiment is not possible in drug design due to the lack of a good method for predicting lipophilicity from the 3D structure.

The pharmacokinetic behavior of a drug is markedly influenced by large variations in physicochemical properties such as lipophilicity. These variations correspond to the main valleys in conformational space rather than to fluctuations around a given conformer. Thus, the *virtual* log *P* of each conformer appears as a better descriptor than the average log *P* of a mixture of conformers for predicting the permeation of a drug across biological membranes. Unfortunately, no direct experiment can give access to the virtual log *P* of conformers of a flexible compound.

We demonstrate here that the method for calculating log *P* values from the MLP can be used within the conformational space of a molecule and can offer a new tool in drug design. The proposed approach relies on two distinct steps, namely the effective exploration of a conformational space, followed by the calculation of the virtual log *P* of each conformer identified.

12.3.1 Quenched Molecular Dynamics: an Effective Exploration of Conformational Space

To achieve the above objective, we have developed and tested [18, 19] a simplified conformational search strategy able to describe efficiently the main valleys of a conformational space. This method comprizes four steps.

1. Four to six different starting geometries are energy-optimized using the Tripos force field with Gasteiger-Marsili formal atomic charges in order to remove initial high-energy interactions. High-temperature molecular dynamics (MD) calculations are then carried out at 2000 K. Each simulation is run for 100 ps with steps of 1.0 fs.

The frame data are stored every 0.05 ps, giving 2000 frames. The starting velocities are calculated from a Boltzmann distribution. Finally, 10 % of all conformers are randomly selected and saved in a database ultimately containing 200 conformers.

2. All conformers in the database are then subjected to energy minimization using the same force field as for the MD calculations. The Powell minimization method is applied with the gradient value of 0.001 to test for convergence. The maximum number of iterations is set at 3000. The energy-minimized conformers are then classified according to increasing energy content.

3. The conformational similarity of the 200 energy-minimized conformers is investigated by comparing every possible pair of conformers. The two criteria of comparison are the force field energy and the RMS distance difference calculated by the option MATCH of SYBYL over all heavy atoms and polar hydrogens. Then an *ad hoc* Fortran program calculates the mean and standard deviations of the RMS values. Two conformers are considered identical when their energy difference is ≤ 3 kcal/mol and their RMS distance difference is less than or equal to the RMS mean minus the standard deviation. When this is the case, one of the two conformers is eliminated from the database, and it is always the one of higher energy.

4. The selected conformers are minimized a second time at the semi-empirical level with AM1 (or PM3) parametrization without the keyword PRECISE. The AM1-minimized conformers are again classified according to increasing heat of formation and selected by heat of formation and RMS distance difference. Identical conformers are eliminated again from the database using the same criteria as above. Finally, the AM1 calculations are repeated on the retained conformers, but with a higher level of precision (inclusion of the keyword PRECISE).

12.3.2 Conformation-Dependent Variations in Lipophilicity as Described by the MLP

Because the MLP is highly sensitive to the 3D structure of a molecule, it appears as the method of choice to calculate the virtual log P values of individual conformers and to explore lipophilicity variations within a conformational space. The calculation of virtual log P values is done quite easily using the calibration equation discussed above [18]. The ranking of conformers with respect to their virtual log P should be biologically more relevant than a ranking based on the relative gas-phase energies. Moreover, log P calculations based on the MLP allow to identify families of conformers according to their lipophilicity behavior, a classification of great importance for understanding the pharmacokinetic behavior of flexible drugs.

The strategy previously used by us for quantifying lipophilicity variations in morphine glucuronides [18] and for identifying the pharmacophore of nootropic agents [19] seems broadly applicable as illustrated by the two examples given below.

12.3.3 Applications

12.3.3.1 Lipophilicity Variation in GABA-receptor Antagonists

In their work on GABA-uptake inhibitors. N'Goka et al. [20] found that even if 6-(3,3-diphenylpropyl)guavacine (see Appendix[*]) was equipotent *in vitro* with its analog SKF 89976-A, it was inactive in *in vivo* models. Based on the $\Delta \log P_{oct-cyc}$ values, the authors attributed this phenomenon to a difference in permeability across the blood-brain barrier, the brain penetration of 6-(3,3-diphenylpropyl)guavacine being lower than those of SKF 89976-A. We propose that the difference in $\Delta \log P_{oct-cyc}$ values can be due either to a stabilization of the neutral form of SKF 89976-A by an intramolecular hydrogen bond, or to a stabilization of its zwitterionic form by an internal electrostatic bond (Fig. 1). Due to the geometric restrictions of the cyclic double bond, this structural effect cannot be present in 6-(3,3-diphenyl-propyl)guavacine.

To test this hypothesis, a conformational analysis was performed on a model of SKF 89976-A, namely the *N*-methylnipecotic acid (see Appendix). Non reproducible results were found, the conformer of lowest energy being highly dependent of the starting geometry. Hence, the conformation space of both the neutral and the zwitterionic forms was carefully investigated using our proposed conformational search strategy.

Intramolecular H-bond   Internal electrostatic bond
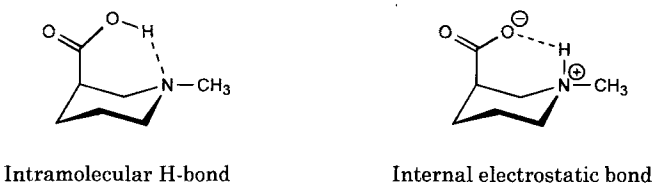
Figure 1. *N*-methylnipecotic acid: conformers stabilized by an intramolecular H-bond for the neutral form or by an internal electrostatic bond for the zwitterionic form

Neutral form

Four classes of conformers were defined for the neutral form of N-methylnipecotic acid according to the *axial* (**A**) or *equatorial* (**E**) position of the two ring substituents. In each class of conformers, the chair conformation was the most stable but boat and twist-boat conformations were only about 6 kcal/mol less stable. In general, the carboxylic acid group prefers the equatorial over the axial position (by ca. 1.2 kcal/mol).

The most stable conformers in the **AA** and **EA** classes (the first symbol corresponds to *N*-Me group and the second symbol to the carboxylic group) was easily identified by our strategy. In contrast, this was not the case for the **AE** class of conformers. Indeed, to allow an internal hydrogen bond, *N*-methylnipecotic acid must be in a **EA** conformation with a torsional angle (O = C-O-H) of 180°. But when the starting geometry was also internally H-bonded, only one conformer was found to have this intramolecular hydrogen bond, its energy being 1.0 kcal/mol (force field) or 3.8 kcal/mol (AM1) above the minimum of the **EA** class. The *trans* conformation of the carboxylic group is

---

[*] Appendix shows structural formulae of all substances presented in this chapter.

responsible for the destabilization of the hydrogen-bonded conformer with respect to the **EA** minimum. Nevertheless, by comparison with other *trans* conformers lacking an intramolecular hydrogen bond, we can quantify the stabilization due to the internal hydrogen bond, i. e., 2.6 kcal/mol (force field) and 2.3 kcal/mol (AM1).

Zwitterionic form

When applied to the zwitterionic form of *N*-methylnipecotic acid, our conformational search gave very different results depending on the method used to perform energy minimizations.

At the force field level, the most stable conformers were found to belong to the **EA** class with an internal electrostatic bond between the axial-carboxylate and the N-protonated group (methyl in *equatorial* position), the cycle being either in a *chair* or a *twist-boat* conformation. This stabilization of the internal salt conformation, also found in 6-methoxysalicylamide derivatives [21, 22], was clearly due to the conditions of calculation (*in vacuo* molecular mechanics calculations where the electrostatic energy is overestimated). We have already demonstrated that the large stabilization of internal ionic bonds with respect to the neutral form does not exist at the semi-empirical level [21].

In contrast to 6-methoxysalicylamides, the conformations of *N*-methylnipecotic acid having an internal ionic bond were not stable at the AM1 level, the geometry optimization leading to a neutral form with an intramolecular hydrogen bond. The lowering of the barrier for proton transfer between the ammonium and carboxylate groups is compatible with an increase of the exothermicity of the proton transfer reaction (Bell-Evans-Polanyi principle) [23], the neutral form being more stable than the zwitterion by 27.0 and 54.4 kcal/mol in 6-methoxysalicylamides and *N*-methylnipecotic acid, respectively. In addition to this crossing of the energy hypersurface between the zwitterion and the neutral form, the semiempirical step D in our procedure gave access to several conformations which do not lie on the neutral hypersurface, namely internally hydrogen-bonded conformations with the cycle in a boat or a twist-boat geometry.

These results show that our proposed search strategy cannot survey the entire conformational space of a molecule from only one starting geometry or electrical form. Indeed, the internally H-bonded conformers can very easily escape detection because the "OH flip" of the carboxylic group from *cis* to *trans* (Fig. 2) (necessary for a hydro-

Intramolecular H-bond   No Intramolecular H-bond
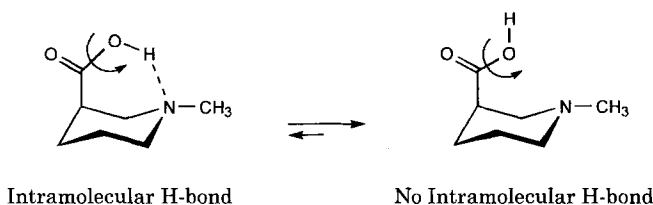
Figure 2. *N*-methylnipecotic acid: *Cis/trans* conformers of the carboxylic group: the rotation around the C-OH bond of the neutral form allows the formation of an intramolecular H-bond for the less stable *cis* geometry

gen bond) requires an amount of energy (ca. 16 kcal/mol determined by systematic search) greater than the rotation of the whole group around the piperidine ring (12 kcal/mol) or the interconversion of the piperidine ring (~ 10 kcal/mol). Hence, any high- or low-temperature dynamics simulation will have a poor probability of finding such local minima.

It should be also noted that the main limitation of the molecular dynamics is that the atoms are bound with a given *connectivity*. The crossing of hypersurfaces described above is thus impossible unless the subsequent step (step D) of semiempirical calculations can be achieved on the geometries selected by previous molecular dynamics and molecular mechanics. However, the three steps A-C are important in order to limit the length of semiempirical calculations.

The calculation of virtual log P values for all conformers identified by molecular dynamics supports the increase in lipophilicity of internally hydrogen bonded **EA** conformers with respect to the more stable **EE** or **AA** conformers. However, the lipophilicity increase remains small (0.1 log P unit), underlining that the octanol/water biphasic system is a bad tool to reveal lipophilicity variations due mainly to an internal hydrogen. This result is compatible with the solvatochromic analysis of octanol/water partition coefficients in which the H-donor parameter α is not significant.

12.3.3.2 Lipophilicity of L-Dopa Esters

To enhance the bioavailability of L-Dopa in the treatment of Parkinson's disease, the potentiality of esters was examined. Aryl and alkyl esters were prepared, and their stability with respect to chemical and enzymatic hydrolysis examined [23a].

The conformational space of each was explored using the Quenched Molecular Dynamics approach, revealing one family of folded conformers and one of extended conformers. The calculation of a virtual log P value for each conformer showed clearly that folded conformers are more polar than extended ones. This change in lipophilicity may be attributed to a decrease in the hydrophobic accessible surface in folded conformers as illustrated for the tetrahydrofurfuryl ester (Fig. 3).

A statistical analysis on all conformers defined the range of lipophilicity accessible for all these esters. Interestingly, the comparison with experimental log P values reveals a clear distinction between aryl and alkyl esters. The experimental partition coefficients are closer to the more polar virtual log P for alkyl esters and closer to the more lipophilic virtual log P for aryl esters, as illustrated in Fig. 4. This difference suggests that the conformational behavior in solution could be different in the two series of esters, the alkyl esters existing largely as folded conformers whereas the aryl esters prefer extended conformations. These conformational changes affect the accessibility of the ester function and thus could be responsible, at least for a part, for the differences in rates of hydrolysis exhibited by alkyl and aryl esters.

In summary the triad – experimental determination of partition coefficients, conformational analysis, and calculation of virtual log P values – appears as an additional tool to explore the conformational behavior of solutes.
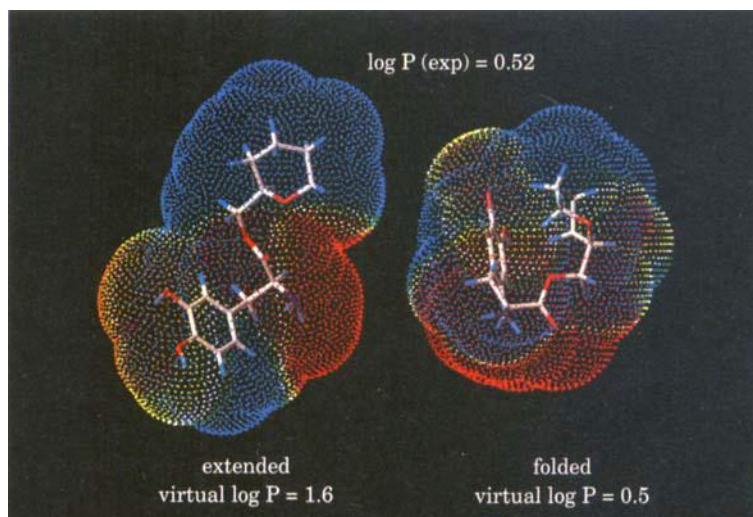
Figure 3. Examples of extended and folded conformers of the tetrahydrofurfuryl ester of L-DOPA. Due to an internal hydrophobic interaction, folded conformers are more polar (*virtual* log $P = 0.5$) than the extended conformers (*virtual* log $P = 1.6$). The MLP is displayed on the water-accessible molecular area. On all our MLP representations, the color coding follows a scale starting from the most polar regions to the most hydrophobic regions, namely red, orange, yellow, white, green, green-blue, blue
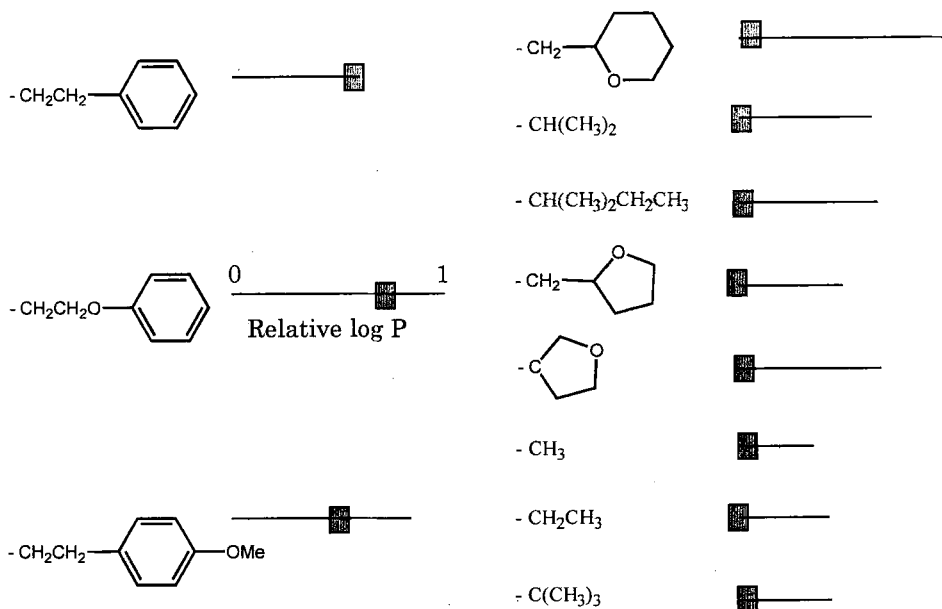
Figure 4. Normalized variations (represented by the line segment) of the calculated log P within the conformational space of alkyl (on the right) and aryl esters (on the left) of L-DOPA. The gray rectangles represent the experimental log P values

12.4 The MLP as a Docking Tool

When modeling binding of ligands to receptors, the major problems one encounters are: a) to select adequate starting points, and b) to estimate the free energy of binding in stable ligand-receptor complexes. Since ligand binding is controlled by a number of intermolecular forces many of which are expressed in lipophilicity (see Chapter 4), the most stable ligand-receptor complexes should be characterized by a maximal similarity between two MLPs, namely that generated by the ligand and that generated by the binding site.

12.4.1 Intrinsic MLP, Perceived MLP, and Similarities Between Them

Based on the above assumption, we have defined the van der Waals surface of the ligand as a working surface on which two MLPs are to be calculated. These are *intrinsic* MLP, i. e., the MLP generated by the ligand on this surface, and the *perceived* MLP, i. e., the MLP generated by the binding site. To quantitate the similarity between the two MLPs, a score function has also been defined:

$$\text{SCORE} = \sum_{k=1}^{Nb_{dots}} (\text{MLP}_k)^{\text{ligand}} \times (\text{MLP}_k)^{\text{receptor}} \qquad (3)$$

Thus, the more positive the score function, the greater the similarity between intrinsic and perceived MLP. In contrast, the greater the dissimilarity, the more negative the score function. At the time of writing, the score function was used to investigate the binding mode of the atypical D_2-agonist piribedil to the D_2 receptor, and to identify possible binding modes of the epitope HEL(52–61) to the I-A^k MHC class II protein, as discussed below.

12.4.2 Applications

12.4.2.1 Binding Modes of some D_2-receptor Agonists

1-(2-Pyrimidil)-4-piperonyl piperazine (piribedil; formula, see appendix) is a noncatechol analog of dopamine of value in the treatment of affective disorders [24]. This compound displays an affinity for the D_2 receptor which is higher ($pK_i = 6.2$) than that of its main metabolite (see Appendix) having a free catechol group ($pK_i = 4.9$) [25]. In order to understand the origin of this unexpected behavior, several strategies were followed to dock piribedil and its main metabolite to the D_2 receptor model developed by Livingstone et al. [26]. A number of interesting results were obtained for the most stable complexes involving piribedil (Fig. 5) and its metabolite (not shown):

a) To reinforce the ionic interaction between the basic amino group of the ligand and the carboxylic group of Asp114 in helix III, both piribedil and its metabolite should adopt a folded conformation leading to an enhanced accessibility of the basic nitrogen.
b) Additional anchor points were found, namely π-π stacking or hydrophobic interactions with Trp387 and Phe383 in helix VI, hydrophobic interactions with Ile158 in helix IV, and H-bonds with Ser194 and/or Ser197 in helix V.
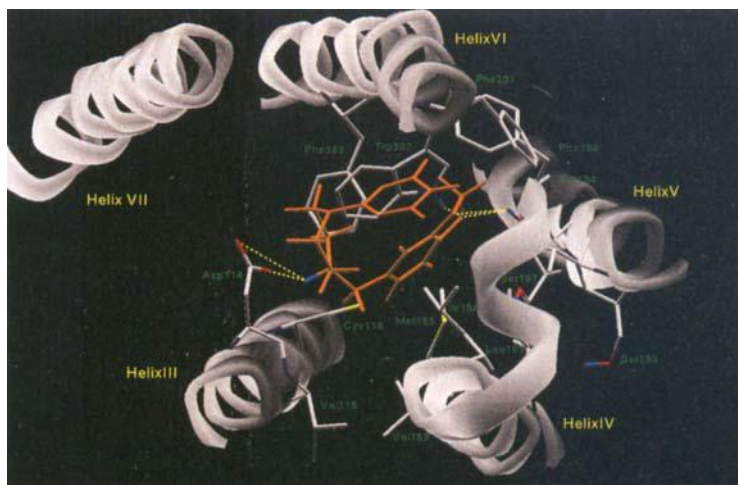
Figure 5. Binding mode of piribedil in the D_2 receptor model. Piribedil is represented in orange and the important side-chains of the receptor are colored by atom type. Only five transmembrane helices are displayed.

On the basis of interaction energies only, the relative importance of these intermolecular interactions was difficult to assess. Moreover, the score function (Fig. 6) suggests that the hydrophobic interactions of piribedil and its metabolite with the aromatic and aliphatic side-chains are more important than H-bonds for stabilizing the ligand-receptor complexes. The importance of hydrophobic interactions is also illustrated by
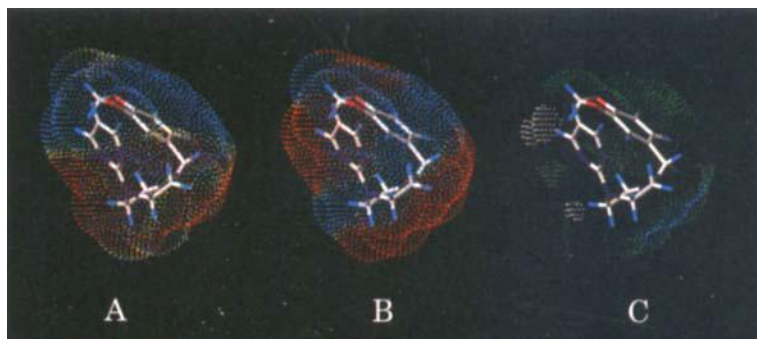
Figure 6. A, intrinsic MLP of piribedil in the bound geometry. For color coding, see Fig. 3; B, perceived MLP of piribedil in the bound geometry. For color coding, see Fig. 3; C, score function, namely similarity between the intrinsic and perceived MLP of piribedil in its bound conformation. On all the figure associated with the score function. MLPs and scores are displayed on the van der Waals surface of the molecule. The color coding for the score function follows a scale starting from the most dissimilar regions to the most similar regions with the following colors: red, orange, yellow, white, green, green-blue, blue.
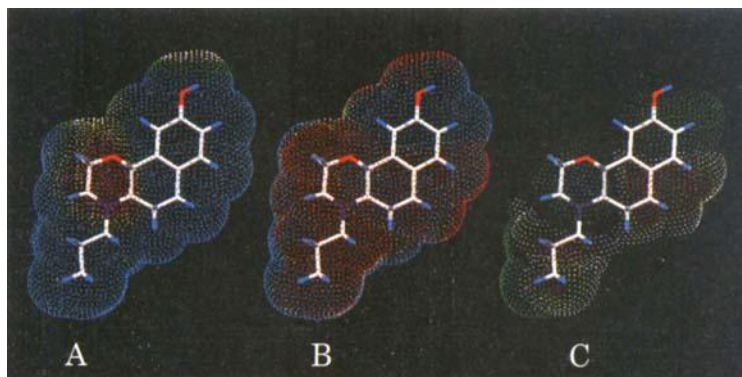
Figure 7. A, intrinsic MLP of naxagolide in the bound geometry. For color coding, see Fig. 3; B, perceived MLP of naxagolide in the bound geometry. For color coding, see Fig. 3; C, score function, namely similarity between the intrinsic and perceived MLP of naxagolide in its bound conformation. For color coding, see Fig. 6

the score function calculated for the binding mode of some rigid D_2-agonists (naxagolide and apomorphine) to the same D_2 receptor model (Fig. 7).

Interestingly, the ranking of the MLP-based score function follows the relative order of binding energies of piribedil and its metabolite. Work is in progress to determine if a QSAR tool can be derived from the score function of stable ligand-receptor complexes.

12.4.2.2 Binding Modes of HEL(52–61) to the I-A^k MHC II Protein

Much attention has been given recently to the crucial role played by peptides in the regulation of the immune response, and especially in the immune action of T lymphocytes. Indeed, the T-cell receptor (TcR) expressed at the surface of T lymphocytes binds to a bimolecular complex made of the major histocompatibility complex (MHC) and peptides derived from proteolytic degradation of the antigen. These peptide-MHC complexes are expressed at the cell surface of antigen-presenting cells (APC). To develop artificial vaccines and for generating antagonist peptides able to control the autoimmune reaction, it is important to uncover the rules governing the association of peptides with MHC proteins and the recognition of peptide-MHC complexes by the T-cell receptor.

As a model system, the binding of the decapeptide HEL(52–61) derived from the hen egg lysozyme (see Appendix) with the mouse I-A^k MHC class II molecule was studied experimentally [27]. The absence of an experimental 3D structure of this peptide-MHC class II complex renders difficult any interpretation of biological results in terms of both peptide binding and T-cell receptor recognition. To gain more structural information on these intermolecular processes, molecular modeling studies are being carried out.

In a first step, the 3D structure of the I-A^k MHC class II protein was built by standard homology approaches starting from the X-ray structure of a related system, the
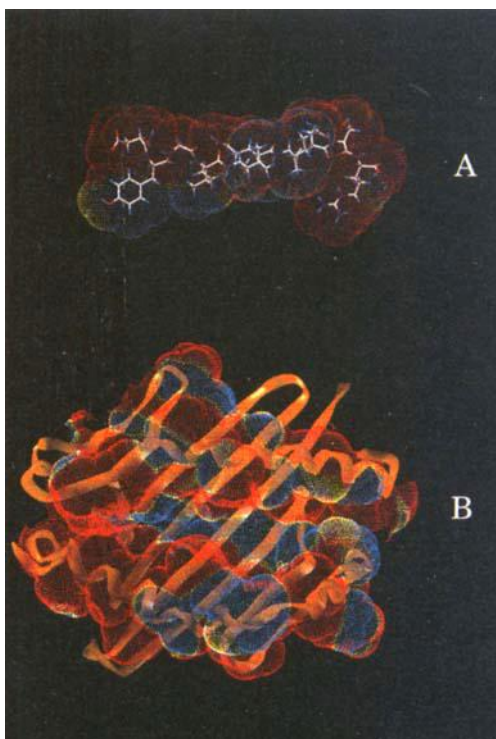
Figure 8. A, MLP of the HEL (52–61) peptide in its bound conformation seen from the face. For color coding, see Fig. 3; B, MLP of the empty binding groove of the I-A^k MHC II protein (partly represented with a ribbon) seen from the top. For color coding, see Fig. 3

human MHC class II HLA-DR1 molecule bound to its HA epitope [28]. Due to large variations in the nature of the peptide-binding groove, the geometry of the HA peptide can not be used to calculate a stable binding mode for the peptide HEL(52–61) to the I-A^k protein. Thus, a comparison between the MLP of the free protein (Fig. 8 B) and the MLP of the peptide in its extended conformation (Fig. 8 A) was used in order to identify two main anchor residues, the Tyr53 and Arg61 of the peptides.

In a second stage, several steps of coupled molecular dynamics and geometry optimization were performed in order to identify a possible binding mode of the HEL(52–61) peptide in the groove of the I-A^k protein. A detailed analysis of the most stable geometry obtained allows the identification of several hydrogen bonds between the backbone of the peptide and the protein (Fig. 9). However, an analysis based only on a single structure was not sufficient to rationalize all the experimental results obtained with modified HEL(52–61) peptides [27]. Therefore, systematic simulations were performed of the complexation of modified peptides with I-A^k. The first results were encouraging when the stabilization energy of the complexes was compared with the experimental affinities of peptides to the I-A^k.
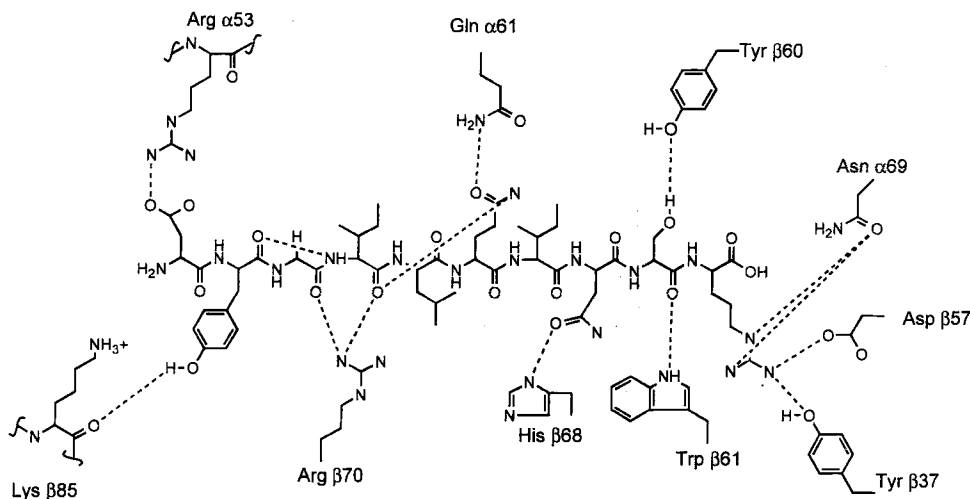
Figure 9. Schematic representation of the hydrogen bonds between the HEL(52–61) peptide and the I-A^k MHC II protein as identified after several runs of molecular dynamics and geometry optimizations

Moreover, the distribution of the MLP around the HEL(52–61)-I-A^k complex was postulated to be of critical importance for recognition by the T-cell receptor. Indeed, short-range intermolecular interactions should be controlled by hydrophobic forces, polar interactions, and hydration rearrangement, all of which are encoded in the MLP. Thus, MLP variations around several complexes were also studied in order to identify a MLP pattern which could rationalize the recognition by T-cell receptors. The preliminary results illustrated important changes in the MLP distribution around complexes due to the presence of the peptide in the groove (Fig. 10 A) or to the change in the orientation of the I-A^k groove side-chains induced by a different binding mode of a mutated peptide (Fig. 10 B). However, the number of cases investigated to date is too small to identify a recognition pattern.

12.5 The MLP as an Additional Field in 3D QSAR

12.5.1 Limits of Standard CoMFA Approaches

Standard comparative molecular field analysis (CoMFA) has reduced the description of ligand-receptor complexes to intermolecular interactions described only by two classic potentials, namely a steric molecular field quantified with a Lennard-Jones function, and an electrostatic field quantified by a Coulombic potential. It is obvious that these two molecular fields cannot take into account all the complex intermolecular forces between ligands and receptors. Another important limitation of the current CoMFA methodology is due to the fact that its two molecular fields are purely
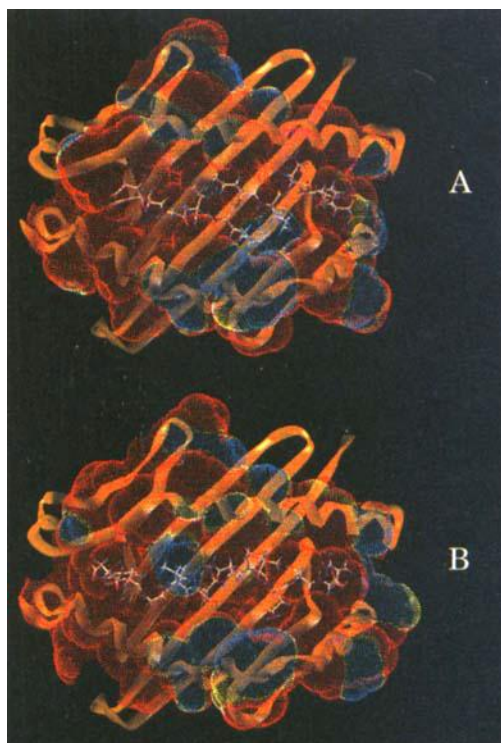
Figure 10. A, MLP of the complex between HEL(52–61) and the I-A^k MHC II protein seen from the top. For color coding, see Fig. 3; B, MLP of the complex between the modified peptide HEL(52-[Thr56]-61) and the I-A^k MHC II protein seen from the top. For color coding, see Fig. 3

enthalpic and fail to take into account the entropic component of the free energy of binding [29].

To overcome these limitations, several approaches were described including the use of additional parameters coming from traditional QSAR, e.g., log P [30], from theoretical approaches [31], from hydrophobic fields calculated by the HINT software [14, 32, 33] (see Chapter 13 for a detailed presentation of this approach), from hydrogen bonding fields [34], and from fields associated with a water probe [35–37] as calculated by the GRID software [38, 39]. Statistical pitfalls, especially the correlation between CoMFA results described below, limit the interest of these approaches.

12.5.2 The MLP, a Third Field in CoMFA

12.5.2.1 Theory

Due to the variety of intermolecular forces encoded in lipophilicity (see Chapter 4), the use of a molecular field of lipophilicity in CoMFA should enrich the modeling of ligand-receptor interactions. Since this is derived from experimental partition coeffi-

cients, it describes implicitly the entropy component of the binding free energy associated with solvation/desolvation of the ligand and binding site.

Thus, the molecular field of lipophilicity is calculated in the same 3D grid as the steric and electrostatic fields, using the MLP described by Eq. 2. It must be noted that this molecular field is not revealed by an atomic probe, the whole molecular environment being included in the fragmental values of lipophilicity. It should be noted that a prescaling of MLP values was defined (multiplication by a constant factor of 15) to render the lipophilic field numerically comparable with the steric and electrostatic fields.

12.5.2.2 Intercorrelations of CoMFA Results Obtained with Different Fields

The use of a molecular field of lipophilicity into CoMFA is not straightforward due to its complex nature and, particularly, to the well-recognized factorization of lipophilicity parameters into a hydrophobic component and polar terms [1]. This duality is series-dependent and tends to increase the statistical complexity of CoMFAs. As a result, wrong conclusions can be deduced from CoMFA models when the composition of lipophilicity in the investigated series is not carefully analyzed. The pernicious effects of neglecting the series-dependent components of lipophilicity become particularly clear when one analyzes ordinary physico-chemical parameters using CoMFA incorporating other molecular fields (hydrogen bond potentials, fields generated by a water probe) [34–37, 40].

While in classical QSAR the correlation between explanatory variables is easy to assess with a correlation matrix, the correlation between CoMFA signals associated with each molecular field is far from evident. PLS analysis is able to handle correlations between molecular fields, but to the best of our knowledge no satisfactory statistical tool exists to unmask a correlation between CoMFA results. Until such tools become available, we propose to perform all CoMFAs with single fields and with all possible combinations of fields. Comparing all statistical and graphical CoMFA results allows one to estimate correlations between CoMFA results obtained with different molecular fields. Correlations can be suspected when the two condition below are fulfilled simultaneously:

a) A first indication of correlation between results must be suspected when the statistical results of the different CoMFA models show a high similarity, in other words when the inclusion of an additional molecular field does not improve the statistical significance of the CoMFA models.
b) A second indication exists when the graphical results of the CoMFA models show a high similarity, that is to say when the signals of the statistical fields derived from CoMFA models are localized in the same regions of space.

At present, a systematic comparison of all CoMFA models generated by all possible combinations of molecular fields, plus the two above criteria, are the only means of detecting correlations between CoMFA results.

12.5.3 Applications

12.5.3.1 Binding to 5-HT$_{1A}$ Receptors [40a]

A set of 280 5-HT$_{1A}$ receptor ligands were selected from available literature data and subjected to 3D QSAR analysis using CoMFA. No model was obtained for serotonin analogs and aminotetralines, despite a variety of alignment hypotheses being tried. In contrast, the steric, electrostatic, and lipophilicity fields alone and/or in combination yielded informative models for arylpiperazines, aryloxypropanolamines, and tetrahydropyridylindoles taken separately. Arylpiperazines and aryloxypropanolamines were then combined successfully to yield reasonably good models for 131 compounds. In a last step, the three chemical classes were combined, again successfully.

This stepwise procedure not only ascertains self-consistency in alignments, it also allows statistical signals (i.e., favorable or unfavorable regions around molecules) to emerge which cannot exist in a single chemical class. The best model obtained using the steric, electrostatic, and lipophilic fields has the following statistical results (n = 185; $q^2 = 0.64$; $N = 4$; $r^2 = 0.82$; $s = 0.53$; $F = 205$). The relative contributions of each field were: 30 % steric, 39 % electrostatic, and 31 % lipophilic.

The graphical results of the best CoMFA model are given on Fig. 11, allowing a detailed picture of the pharmacophoric elements around the aromatic moiety and the basic nitrogen. The proposed pharmacophore can be summarized as shown in Fig. 12. This model is consistent with previous 3D QSAR models obtained for single classes of ligands [41–43], and with information on the 5-HT$_{1A}$ receptor obtained by homology modeling [44].

Whenever possible, the predictive power of any QSAR model should be assessed by its capacity to predict the 5-HT$_{1A}$ affinity of a test set as much as by its cross-validated correlation coefficient q^2. Here, a test set of 16 arylpiperazines and aryloxypropanolamines was set aside at the beginning of the study. The linear regression had a correlation coefficient of $r^2 = 0.72$ in agreement with the q^2 value of the model ($q^2 = 0.64$). The most deviant compounds had structural fragments not encountered in the training set. This illustrates the main limitation of 3D QSAR models, which may fail for structural motifs not included in the training set. It should be also noted that the use of the MLP increases the quality of prediction of binding affinities for the test set, presumably due to the implicit inclusion in MLP of entropic effects.

In summary, the use of the MLP in CoMFA allows a better description of the binding mode of 5-HT$_{1A}$ ligands, mainly by revealing new polar and lipophilic regions. Other CoMFA applications in our laboratory demonstrate that the use of the MLP as a third field in CoMFA improves the signal description, especially by associating lipophilic properties to steric regions. Indeed, a positive steric signal in CoMFA can be generated either by the presence of a lipophilic (apolar) fragment or by a polar fragment, an ambiguity cancelled by the MLP.
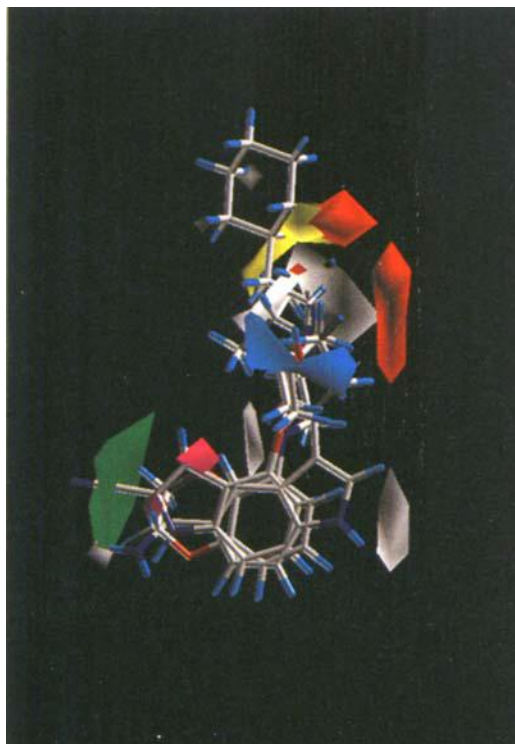
Figure 11. Graphical CoMFA results for the binding to the 5-HT$_{1A}$ receptor with the super-position of N-cyclohexylethyl N'-1,4-dioxotetralin piperazine, cyanopindolol and N-methyltetrahydropyridyl 5-carboxamidoindole displayed. The color coding is the following; green regions and red regions where steric interactions respectively increase and decrease 5-HT$_{1A}$ affinity; white and magenta regions where electrostatic interactions with a positive charge respectively increase and decrease 5-HT$_{1A}$ affinity: blue and yellow regions where polar interactions respectively increase and decrease 5-HT$_{1A}$ affinity. The interpretation of electrostatic and MLP signals is reversed if, respectively, a negative charge and hydrophobic interactions are considered
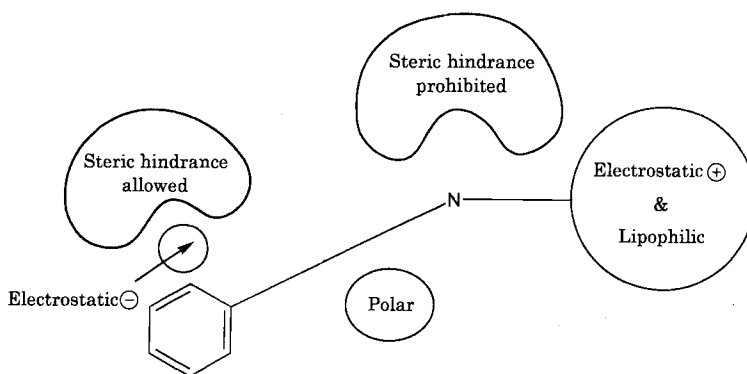
Figure 12. Schematic representation of the 5-HT$_{1A}$ binding mode as identified with CoMFA

12.5.3.2 CoMFA Models of Sweetness in Halogenated Sucroses

In the field of sucrochemistry the discovery of intensely sweet sugar derivatives with chloro substituents has led to the synthesis of numerous mono-, di-, tri- and tetradeoxy halogenated sugar derivatives. These compounds have revealed intense sweetness, up to several thousand times more than sucrose, leading in the selection of 4,1',6'-trichloro-4,1',6'-trideoxygalactosucrose (sucralose, see Appendix) as a dietary sweetener [45].

Compounds of this type appear particularly well suited as objects of structure-sweetness relationship studies meant to improve our understanding of sweet taste receptors [46]. Since lipophilicity has been recognized as an essential feature [47] of activators of sweet taste receptors, as confirmed by a principal component analysis of the 2D structural characteristics of sucralose derivatives [48] it was only logical to initiate a 3D QSAR analysis including the MLP.

The difficulty of 3D QSAR analyses was increased by the flexibility of the halogenated sucralose derivatives, high-level techniques being needed to explore their conformational hyperspace and identify their main conformers [45, 48]. Significant differences were observed between the conformation of sucralose in solution and in the crystal, rendering futile any 3D QSAR based on a single conformer. In the absence of any information on the active conformation, the following strategy was adopted:

- 18 pharmacophoric points were defined including the main features of previously published models of the sweet taste.
- The DISCO approach [49, 50] was used to select one conformer per compound, allowing a common superposition for the 18 pharmacophoric points. Several solutions were retained based on different conformers belonging to different regions of the conformational space.
- CoMFA models with the three fields were calculated following the strategy described in section 12.5.2.2.

In all models, inclusion of the MLP enhances significantly the CoMFA results, confirming the importance of lipophilicity in influencing sweetness. The best CoMFA models were not found with alignments of global minima, but with conformers belonging to a region within 8 kcal/mol of the global minimum. The best model obtained is based on the electrostatic and lipophilicity fields only (28 compounds, $q^2 = 0.79$, $r^2 = 0.94$, $N_{comp} = 3$, $s = 0.24$) with a respective weight of 43 % and 57 %.

The graphical results of this model confirm the importance of lipophilicity in the fructofuranose residue (Fig. 13). Indeed, the signal generated by the MLP is almost exclusively concentrated around this fragment, supporting the hypothesis of a hydrophobic pocket able to accommodate this ring [51]. The predictive power of the best model generated was determined with a test set of five compounds, giving a high concordance ($r^2 = 0.988$) between experimental and predicted sweetness.
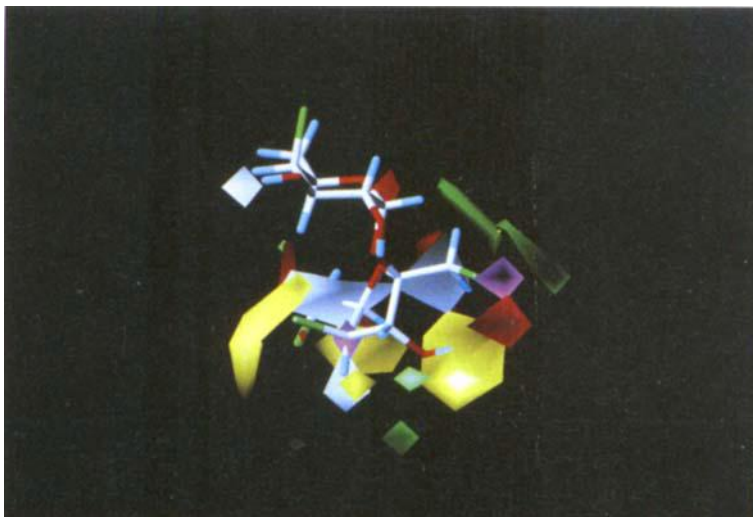
Figure 13. Graphical CoMFA results for the halogenated saccharose derivatives. The yellow signal underlines the importance of large hydrophobic interactions around the fructofuranose derivatives for their sweetness

12.6 Perspectives

In this review, we have presented the MLP as a recent and promising computational tool in drug design. As discussed in the various sections, the MLP in its present state of development finds applications in the following fields:

1. As a method for calculating partition coefficients, it has the unique characteristic of taking 3D conformational effects into account. Hence, the MLP can calculate virtual log P values of individual conformers, which when compared with experimental log P values can give indications on the predominant conformer of a given solute.
2. When comparing the lipophilicity fields of a ligand (the intrinsic MLP) and a binding site (the MLP emitted by the receptor and "perceived" by the ligand), significant information can be derived on modes of docking. Work is in progress to improve the quantitation of this comparison and to expand the MLP into an iterative docking tool.
3. Because of the richness of the information it encodes, the MLP has proven a most useful addition to the CoMFA method, allowing the lipophilicity field to be included in addition to the traditional steric and electrostatic fields. This approach is particularly useful in allowing hydrophobic interactions and H-bonds to be taken into account in 3D QSAR, not to mention the entropic component of recognition forces.

A characteristic of the MLP is that it is based on experimental data and not on calculations using an abstract probe. By encoding other experimental data into the MLP (e.g., log P values from alkane/water systems, Δlog P values), significant advances can be expected. Computational progresses can also be envisaged, e.g., a better definition

of atomic fragments and an improved inclusion of molecular topology. Many avenues are thus open to bring the MLP to age, and these are actively explored in our laboratory.

References

[1] El Tayar, N., Testa, B., and Carrupt, P. A., *J. Phys. Chem.* **96**, 1455–1459 (1992)
[2] Tsai, R. S., Testa, B., El Tayar, N., and Carrupt, P. A., *J. Chem. Soc., Perkin Trans.* **2**, 1797–1802 (1991)
[3] Leo, A. J., *Chem. Rev.* **93**, 1281–1306 (1993)
[4] Leo, A. J., *J. Pharm. Sci.* **76**, 166–168 (1987)
[5] Rekker, R. F., and De Kort, H. M., *Eur. J. Med. Chem.* **14**, 479–488 (1979)
[6] Hansch, C., and Leo, A., *Substituent Constants for Correlation Analysis in Chemistry and Biology.* Wiley: New York, 1979
[7] Broto, P., Moreau, G., and Vandycke, C., *Eur. J. Med. Chem.,* **19**, 61–65 (1984)
[8] Ghose, A. K., and Crippen, G. M., *J. Comput. Chem.* **7**, 565–577 (1986)
[9] Kantola, A., Villar, H. O., and Loew, G. H., *J. Comput. Chem.* **12**, 681–689 (1991)
[10] Fauchère, J. L., Quarendon, P., and Kaetterer, L., *J. Mol. Graphics* **6**, 203–206 (1988)
[11] Audry, E., Dubost, J. P., Colleter, J. C., and Dallet, P., *Eur. J. Med. Chem.* **21**, 71–72 (1986)
[12] Audry, E., Dallet, P., Langlois, M. H., Colleter, J. C., and Dubost, J. P., *Prog. Clin. Biol. Res.* **291**, 63–66 (1989)
[13] Furet, P., Sele, A., and Cohen, N. C., *J. Mol. Graphics* **6**, 182–189 (1988)
[14] Kellogg, G. E., Semus, S. F., and Abraham, D. J., *J. Comput.-Aided Mol. Des.* **5**, 545–552 (1991)
[15] Hirono, S., Liu, Q., and Moriguchi, I., *Chem. Pharm. Bull.* **39**, 3106–3109 (1991)
[16] Camilleri, P., Watts, S. A., and Boraston, J. A., *J. Chem. Soc., Perkin Trans.,* **2**, 1699–1707 (1988)
[17] Gaillard, P., Carrupt, P. A., Testa, B., and Boudon, A., *J. Comput.-Aided Mol. Des.* **8**, 83–96 (1994)
[18] Gaillard, P., Carrupt, P. A., and Testa, B., *Bioorg. Med. Chem. Lett.* **4**, 737–742 (1994)
[19] Altomare, C., Cellamare, S., Carotti, A., Casini, G., Ferappi, M., Gavuzzo, E., Mazza, F., Carrupt, P. A., Gaillard, P., and Testa, B., *J. Med. Chem.* **38**, 170–179 (1995)
[20] N'Goka, V., Schlewer, G., Linget, J. M., Chambon, J.-P., and Wermuth, C. G., *J. Med. Chem.* **34**, 2547–2557 (1991)
[21] Carrupt, P. A., Tsai, R. S., El Tayar, N., Testa, B., de Paulis, T., and Högberg, T., *Helv. Chim. Acta* **74**, 956–968 (1991)
[22] Tsai, R. S., Carrupt, P. A., Testa, B., Gaillard, P., El Tayar, N., and Högberg, T., *J. Med. Chem.* **36**, 196–204 (1993)
[23] Dewar, M. J. S., *The PMO Theory of Organic Chemistry.* Plenum Press: New York, 1975

[23a] Brunner-Guenat M., Carrupt P. A., Lisa G., Testa B., Rose S., Thomas K., Jenner P. and Ventura P., J. Pharm. Pharmacol. **47**, 861–869 (1995)

[24]   Dourish, C. T., *Prog. Neuro-Psychopharmacol. & Biol. Psychiatr.* **7**, 3–27 (1983)

[25]   Evrard, Y., *Actualités Thérapeutiques,* Special issue, 16–20 (1991)

[26]   Livingstone, C. D., Strange, P. G., and Naylor, L. H., *Biochem. J.* **287**, 277–282 (1992)

[27]   Hernandez, J. F., Cretin, F., Lombard-Platet, S., Salvi, J. P., Walchshofer, N., Gerlier, D., Paris, J., and Rabourdin-Combe, C., *Peptides* **15**, 583–590 (1994)

[28]   Stern, L. J., Brown, J. H., Jardetzky, T., Gorga, J. C., Urban, R. G., Strominger, J. L., and Wiley, D. C., *Nature* **368**, 215–221 (1994)

[29]   Klebe, G., and Abraham, U., *J. Med. Chem.* **36**, 70–80 (1993)

[30]   Altomare, C., Carrupt, P. A., Gaillard, P., El Tayar, N., Testa, B., and Carotti, A., *Chem. Res. Toxicol.* **5**, 366–375 (1992)

[31]   Waller, C. L., and Marshall, G. R., *J. Med. Chem.* **36**, 2390–2403 (1993)

[32]   Horenstein, B. A., and Schramm, V. L., *Biochemistry* **32**, 9917–9925 (1993)

[33]   Abraham, D. J., and Kellogg, G. E., Hydrophobic fields. In: *3D QSAR in Drug Design. Theory Methods and Applications*, Kubinyi, H. (Ed.). ESCOM: Leiden; 506–522 (1993)

[34]   Kim, K. H., Greco, G., Novellino, E., Silipo, C., and Vittoria, A., *J. Comput.-Aided Mol. Des.* **7**, 263–280 (1993)

[35]   Kim, K. H., *Quant. Struct.-Act. Relat.* **11**, 309–317 (1992)

[36]   Kim, K. H., *Med. Chem. Res.* **1**, 259–264 (1991)

[37]   Kim, K. H., *Quant. Struct.-Act. Relat.* **14**, 8–18 (1995)

[38]   Wade, R. C., Clark, K. J., and Goodford, P. J., *J. Med. Chem.* **36**, 140–147 (1993)

[39]   Wade, R. C., and Goodford, P. J., *J. Med. Chem.* **36**, 148–156 (1993)

[40]   Steinmetz, W. E., *Quant. Struct.-Act. Relat.* **14**, 19–23 (1995)

[40a] Gaillard P., Carrupt P. A., Testa B. and Chambel P., J. Med. Chem., in press (1996)

[41]   Agarwal, A., Pearson, P. P., Taylor, E. E., Li, H. B., Dahlgren, T., Herslöf, M., Yang, Y., Lambert, G., Nelson, D. L., Regan, J. W., and Martin, A. R., *J. Med. Chem.* **36**, 4006–4014 (1993)

[42]   El-Bermawy, M. A., Lotter, H., and Glennon, R. A., *Med. Chem. Res.* **2**, 290–297 (1992)

[43]   Langlois, M., Brémont, B., Rouselle, D., and Gaudy, F., *Eur. J. Pharmacol., Mol. Pharmacol. Sect.* **244**, 77–87 (1993)

[44]   Kuipers, W., van Wijngaarden, I., and Ijzerman, A. P., *Drug. Design & Discovery* **11**, 231–249 (1994)

[45]   Meyer, C., Pérez, S., Hervé du Penhoat, C., and Michon, V., *J. Am. Chem. Soc.* **115**, 10300–10310 (1993)

[46]   Yamazaki, T., Zhu, Y. F., Benedetti, E., and Goodman, M. An approach to a three-dimensional model for sweet taste. In: *Peptide Chemisty, 1992, Proc. Jpn. Symp., 2nd,* Yanaihara, N. (Ed.). ESCOM: Leiden; 343–346 (1993)

[47]   Lichtenthaler, F. W., and Immel, S., Sucrose, sucralose and fructose: correlations between hydrophobicity potential profiles and AH-B-X assignements. In: *Sweet-Taste Chemoreceptors.* Mathlouthi, M., Kanters, J. A., and Birch, G. G. (eds.). Elsevier: London; 21–53 (1993)

[48]   Meyer, C., *Bases moléculaires du goût sucré: flexibilité conformationnelle et relations structure-activité d'édulcorants intenses dérivés du saccharose.* Ph. D. Thesis, Université de Nantes, Nantes, 1994

[49]   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)

[50]   Myers, A. M., Charifson, P. S., Owens, C. E., Kula, N. S., McPhail, A. T., Baldessarini, R. J., Booth, R. G., and Wyrick, S. D., *J. Med. Chem.* **37**, 4109–4117 (1994)

[51]   Faurion, A., and MacLeod, P. Sweet taste receptor mechanisms. In: *Nutritive Sweeteners.* Birch, G. G., and Parker, E. D. (Eds.). Applied Science Publishers: London; 247–273 (1982)

Appendix

Structural formulae
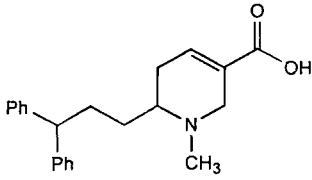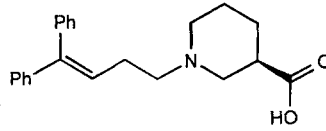
6-(3,3-diphenylpropyl)guavacine   SKF 89976-A   N-methylnipecotic acid
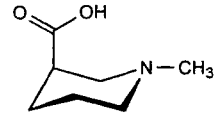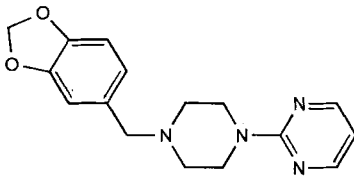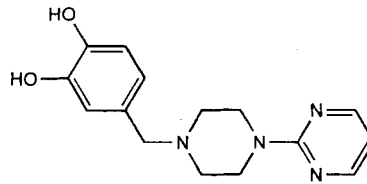
Piribedil   S 584

Asp - Tyr - Gly - Ile - Leu - Gln - Ile - Asn - Ser - Arg

HEL(52-61)
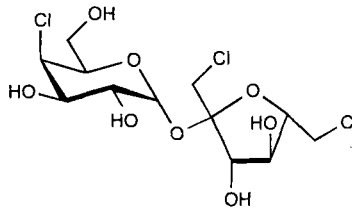
Sucralose

13 Hydrophobic Fields in Quantitative Structure-Activity Relationships

Gerd Folkers und Alfred Merz

Abbreviations

CADD	Computer-aided drug design
CoMFA	Comparative molecular field analysis
DLVO	Derjaguin-Landau-Verwey-Overbeek (potential)
DNA	Deoxyribonucleic acid
HSV1-Tk	Herpes-simplex-virus thymidin kinase
MHC	Major histocompatibility complex
MEP	Molecular electrostatic potential
MLP	Molecular lipophilic potential
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser enhancement (effect)
QSAR	Quantitative structure-activity relationship
3D QSAR	Three-dimensional quantitative structure-activity relationship
RNA	Ribonucleic acid
WGA	Wheat-germ agglutinin

13.1 Introduction

Hydrophobic effects play a key role in the architecture of biopolymers and their interactions with small molecules. Both the folding and the ligand interaction of proteins is predominantly governed by hydrophobicity. The same is true for building and stabilization of membranes, for instance by steroid-phospholipid interaction.

Very recently, our attention has been drawn to a further, hitherto underestimated importance of hydrophobicity in protein-ligand interactions. DNA and RNA interaction with proteins might be more influenced by hydrophobic binding than it has been expected [1]. Characterization of the TATA box recognition by its protein revealed that the nucleosides are nearly exclusively bound by hydrophobic interaction, and that this process is sufficient for a specific recognition [2, 3]. A more recent observation concerns the valine-binding part of RNA [4]. The binding of the amino acid is quite specific, though it is merely hydrophobic. Even stereoisomers can be discriminated showing a 10- to 100-fold difference in affinity. In spite of the knowledge of many structural details, at least for cytosolic protein-ligand complexes, the nature of the hydrophobic effect has not yet been assessed experimentally, nor formulated theoretically in a satisfactory manner. Measurements and definitions of global molecular lipophilic properties like log P are valuable for quantitative structure-activity relationships (QSAR)

and drug formulation, but none of the approaches yields information about conformation dependency and molecular surface distribution of hydrophobicity. The latter might be one of the most important points in study of ligand-protein interactions and for the investigation of biological activity of drugs on the molecular level. Therefore, an appropriate description of hydrophobicity at an atomic level, would be an important progress for drug design. The most common tool in drug design are force fields. Most descriptions of properties of bioactive ligands are done on the force field level. Thus, it might be desirable also to have a "hydrophobic force field".

Geometry optimization of prefolded cytosolic proteins very often yield plausible structures close to the X-ray or NMR results. The empirical force field methods are even used to optimize experimentally derived structures. This raises the question whether hydrophobicity could at least be defined in empiric terms or deconvoluted on the force field level.

Is it permitted to talk about hydrophobicity as a field property, although it has no explicit formulation? We agree with the arguments of Abraham and Kellogg [5] that, since a field in physics is by definition "a region in space characterized by a physical property, such as gravitational or electromagnetic force or fluid pressure, having a determinable value at every point in the region", a field in general may be defined in terms of empiric force fields. Empiric force fields are artificial descriptions of physical reality assuming atoms to have masses, to behave like very small particles and to be connected by bonds, both masses and bonds following Newton's laws. Many of the ideas presented here are already summarized in the paper of Abraham and Kellogg, cited above.

13.2 Definition

In principle, two different approaches can be chosen to define an empirical hydrophobic force field, leading to two different but complementary tools in drug design.

The first approach is the addition of atomic property contributions in order to describe a molecular field. *In praxi* the single atom-atom interactions between protein and ligand or solute and ligand are summed up to an energy, related to a distance function and empiric microenvironment parameters. Cut-off values that neglect any effect beyond a certain distance are frequently introduced as a tool to save computer time. The contributions to atomic properties are derived from the experiment. Because the hydrophobic effect is closely related to partitioning phenomena, the partition coefficient can be used for deductions concerning size and magnitude of the hydrophobic contribution.

In the second approach, a standard force field is used and "trained" by adjusting the parameters in reference to crystal structures of hydrophobic molecules, such as membrane segments like phosphatidylcholines [6]. By this operation microenvironment effects might not be explicitly treated, but are implicitly contained in the parameters by the empirical adjustment of the force constants of the force field. In contrary to the first case, the parametrization of a hydrophobic force field is aimed to study interaction of hydrophobic molecules or parts of it with the membrane, e.g., the membrane interaction of drugs like verapamil or the dihydrophyridines [7]. On the one hand, this

seems to be a valuable approach because the Lennard-Jones potential, representing the van der Waals interactions in the force fields, has an almost identical energy to distance plot as the DLVO potential (Derjaguin, Landau, Verwey, Overbeek), which is empirically accessible by coagulation experiments in colloids [8]. On the other hand, the term *hydrophobic* is somewhat misleading in the force field context, because it describes rather a phenomenon, than a single force to be defined by a function in force field. Several other single forces contribute to the phenomenon hydrophobicity. Strong dispersion and van der Waals forces exist between water molecules and nonpolar solutes [9]. They are considered to be substituted by hydrophobic interactions (i.e. interactions between nonpolar groups) in the folded conformational states of the solutes (cf. theories on protein folding and stabilization). Hydrogen bonding is also of extreme importance for the hydrophobic effects since depending on the solvent, the pattern of hydrogen bonds in a solute is quite different. Optimization of this pattern, influenced by the surrounding solvent molecules might therefore be one approach to simulation of hydrophobic effects. Hydrogen bonds can be defined by a potential in force fields, such fields being widely used in dynamic simulations of peptidic ligands and proteins. However, it remains an open question whether geometrical optimization of hydrogen bond patterns simulates hydrophobic effects. This has been the standard interpretation, which, however, has sometimes been contradicted in recent communications [8]. Desolvation and structural rearrangment of water molecules may, at least in infinite dilution, not be the reason for association of nonpolar groups because the gain of entropy by destroying the water molecular network is not large enough to compensate the loss of entropy by association of the nonpolar groups. Therefore, force-field modeling of hydrophobic interactions seems to be strongly case-dependent. The situation in anisotropic, high-density structures, such as membranes and vesicles or near receptor proteins may be quite different from the situation in solution.

The first definition of the hydrophobic field, is mostly used to describe a quantifiable molecular shape of a drug molecule or is binding site at its biological target. Mapping, quantification and visualization are the aims of hydrophobic fields of the first definition. These data are further used as the basis of (three-dimensional) QSAR studies.

Therefore, in contrast to the force fields, these mapping data allow for the design of active ligands at a stage at which the receptor or active sites are largely unknown.

13.3 Fragmental Property Contributions

The term comprises those fractional contributions within a molecule than sum up to the molecular property-lipophilicity in this instance. They may include solvent-accessible areas, atomic hydropathic constants, fragment constants, partition coefficient-based substituent constants, or atomic parameters. Use of the term is at last based on the assumption of an additivity of group contributions to values of biological activity (for a comprehensive discussion, see [10]). Actually, fragment contributions from a basis of all approaches that define hydrophobic fields. A short summary of these approaches is given in this chapter.

Partitioning between polar and nonpolar phases was the first experimentally accessible property of molecules of biomedicinal interest. Based on observations which were first made during the last century, Hansch and coworkers were the first who derived, some 40 years ago, constants for fragments of molecules (the substituents) that could be used in additive manner for description of a molecular property correlated to the partition coefficient between octanol and water (P). Fragment values for a number of atoms and functional groups were summarized by Rekker [11] and Hansch and Leo [12].

A next step forward was setting up an algorithm that enabled one to describe atomic solvent-accessible surfaces [13]; these have since been used for a number of approaches. Lipophilicity parameters for different atom types derived from water octanol partition coefficients have been introduced by Ghose and Crippen [14].

A number of combinations of these fundamental approaches with new experimental techniques as protein crystallography and microcalorimetry have led to very efficient and considerably stable algorithms for the expression of molecular hydrophobic fields. In the following we shall focus on the most widely used and most transparent techniques.

13.4 Algorithms for Calculation of Hydrophobic Fields

13.4.1 GRID

The GRID program is one of the most widely used computational procedures for the mapping of molecular surfaces of biomedicinal interest (drug molecules and binding sites) [15]. By moving probe atoms or probe molecules like water, CH or nitrogen across the surface of the target molecule and calculating the interaction energies from simple force field potentials, one yields a property distribution of attractive and repulsive forces projected on the solvent accessible surface of the target molecule. There are sev-
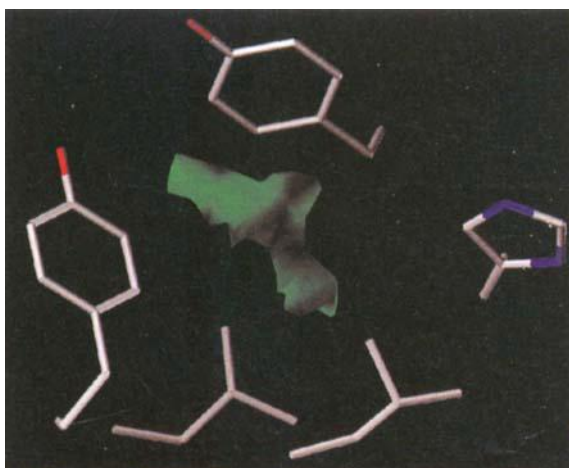
Figure 1. A region of optimal hydrophobic interaction (green) in a MHC pocket [18] calculated by the GRID method. Similar regions result from the HINT approach

eral procedures to obtain quantitative data. Inspection of the interaction tables [16] or sampling algorithms [e.g., 17] provides sizes and magnitudes of the hydrophobic effects on the ligand's surface (Fig. 1).

However, intelligent use of probes offers a very useful tool to collect data of molecular property fields that can be used in QSAR studies. In a comparative study on MHC-peptide interactions, we were able to show that GRID in combination with 3D QSAR tools, led to quite the same interpretation of ligand interaction as the procedures which are based on hydropathic fragment constants [19].

13.4.2 Molecular Lipophilicity Potential (MLP)

The procedure of Dubost [20, 21] resulted in a molecular lipophilicity potential (MLP), which resembles in the set up the MEP, the Molecular Electrostatic Potential.

The MLP is based on atomic hydrophobicity constants, which originate from application of the autocorrelation function on molecular lipophilicity [22]. Surrounding a molecule with lipophilic particles or solvent molecules intuitively creates the idea of a force that is repulsive or attractive for the solvent particles at different parts of the solute molecule, depending on the lipophilicity of its single atoms in that parts. This picture immediately results in a force field expression, associating a distant-dependent energy function to each of the solvent particles in relation to each of the atoms of the solute. This causes a distribution pattern of lipophilic solvent molecules around the solute molecule that depends on their atom-to-atom interaction energies. Connecting all identical interaction energies in 3D- space creates a form of isolipophilicity contours, that provide a new insight in the molecular shape. Application within 3D QSAR studies has been reported (see examples below).

13.4.3 Hydrophobic Interaction Potential (HINT)

The first use of HINT visualization of hydrophobic binding interactions. The importance of such representation has been exhaustively discussed in the literature [e.g., 13, 23]. HINT may be characterized as a force field, because it combines calculations of an interaction energy with a distance function. The resulting field, originating from multiplication of atomic octanol/water partition coefficient of the ligand atoms with those of the interaction site, is called a hydropathic field. Solvent accessible surface and van der Waals interaction is taken into account in order to monitor close contacts. Since it is based on hydrophobic fragment constant by Hansch and Leo [12], HINT differs from other hydrophobic fields in taking into account neighboring structural informations, such as branching, side-chain contacts, etc. Details of the algorithm are given in [5]. The use of HINT is very convenient, since it can be very easily integrated in all the leading modeling software. Salient features are the representation of maps that originate from the receptor site and define an optimal hydrophobic interaction area for putative ligands or the display of the hydrophobic interaction surfaces of bioactive ligands and their binding sites.

Meanwhile, there are numerous examples of applications of hydrophobic fields in studies on structure-activity relationships. We will focus on some recent reports that demonstrate a multidisciplinary approach using molecular field analyses.

13.5 Combination of Hydrophobic Fields with 3D QSAR Techniques

Derivatives of the plant alkaloid ryanodine have been used to identify the pharmacophore of this calcium blocker [24]. Ryanodine is a natural insecticide which exhibits action on the vertebrate skeletal and cardiac muscles, by inhibition of calcium flux. It was demonstrated to bind specifically to a protein, now commonly referred to as the ryanodine receptor, that acts by regulation of a calcium channel. The binding to this receptor protein is complex, displaying multiple affinities and a cooperative action. From a training set of 19 ryanodine derivatives, the common molecular properties of the derivatives have been calculated. The relative rigidity of the ryanodine molecules is of considerable advantage in modeling, facilitating the choice of appropriate minimum energy conformations.

Whereas the calculation of general properties as overall hydrophobicity or log P is important for the biokinetics of the drug and its potential development, the study of the detailed interaction phenomena associated to the pharmacophore must be made by identification of specific areas in 3D space around the molecule. CoMFA is one prominent tool that provides such facilities [25] and the use of HINT within the CoMFA algorithm allows areas of hydrophobic interactions to be displayed and quantified.

It should be emphasized, that the present results once more confirm the view of the molecular shape being the important information for interpretation of structure-activity relationships. Covalent modification of the ligand does not necessarily result in a different mode of action or alternate binding. In most cases modification of the molecular shape results in a change of binding mode, or even mode of action, possibly due to considerable changes occurring in hydrophobic and/or electrostatic fields, or covalent and conformational changes.

Several new ryanodine derivatives could be predicted using the molecular fields, relocating critical substituents or replacing them. The field-based predictions, using CoMFA, were very close to the experimental dissociation constants of these new compounds.

13.6 Mechanistic Interpretation of Protein-Ligand Crystal Data

Wheat-germ agglutinin (WGA) is a plant protein, binding among others to glycophorin A in erythrocytes. Binding is both triggered and established by tetrasaccharides, which yields a asymmetric interlinkage of dimers. Protein crystallography of a complex of WGA with a bivalent sialoglycopeptide [26] yielded new insight in the binding mode. Hydrophobic field description of the different binding sites in the dimers were used to differentiate and to quantify the different types of noncovalent binding interactions. It became clear, however, that the hydrophobic interactions are predominant for one of the binding sites.

13.7 YAK

Pseudoreceptor modeling is half-way between 3D QSAR and structure-based modeling techniques for protein-ligand interaction. The program YAK provides a tool for creating a ligand-binding site around the ligand in question [27]. In contrast to the underlying principles in building by homology, it is not necessary that the pseudoreceptor model resembles structurally the true biological receptor. The main requirement is that it binds its ligands by sufficient by specific noncovalent interactions so as to accommodate them in a manner similar to a biological receptor. Given a pharmacophore, vectors are generated from ligand functional groups based upon the directionality of molecular interactions. Hydrogen-extension vectors originate at H-bond donors and lone-pair vectors arise from H-bond acceptors. Their end-point marks an ideal position for a H-bond acceptor and a H-bond donor, relative to the donor and acceptor atoms, respectively.

Hydrophobicity vectors come from apolar hydrogen atoms. Their end-points mark the approximate position for a hydrophobic moiety relative to the apolar hydrogen atom. Amino acid residues selected from a preference data base of the most frequently occurring interactions between a functional group and an amino acid are then automatically docked and oriented. The hydrophobic contributions are taken implicitly from atomic fragments, based on the MLP extension, as proposed by Furet et al. [28]. A distance-dependent function is used in YAK to calculate the lipophilicity of points in 3D space at the end-point of hydrophobic vectors, where the potential receptor might interact with the ligand molecules (Fig. 2). So far, YAK is very near a structure-based modeling technique, quantification of the free energy of binding being used for valida-
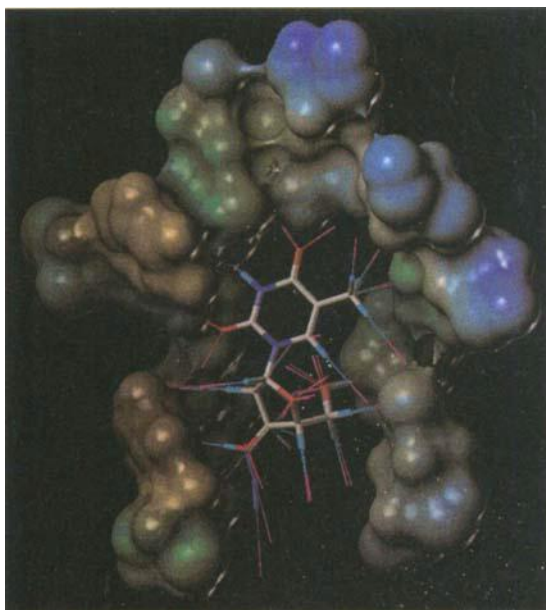
Figure 2. The YAK model of human thymidine kinase [29] is shown, containing the property vectors of a ligand which is surrounded by the pseudoreceptor. The pseudoreceptor is represented by its van der Waals surface and color-coded by its hydrophobic potential, based on Ghose and Crippen parameters [14]

tion of the model. Experimental binding data are compared with Gibbs energy for the ligands of the training set that are based in part on desolvation energies. A correlation in the $\Delta\Delta G$ plot indicates that the model is able to describe at least parts of the biophysical behavior of the ligand-protein interaction in terms of the fields mentioned.

YAK provided new insight into the subtype specificity in a class of viral and human thymidine kinases [29] (Fig. 2). This subtype specificity could not be found by sequence alignment techniques or modeling of the active site [30].

Two positions within the set of the ligands have been identified as important for the differentiation between human and viral protein. The binding partner in the viral protein is a hydrophilic residue, most likely an aspartate, whereas in the human protein this role is clearly taken over by a hydrophobic binding residue, such as a phenylalanine.

13.8 Experiments and Caveats

The use of hydrophobic fields in 3D QSAR, which has been described in Chapter 12, is faced with at least two caveats: these are the multiple or alternate binding mode, and the conformation dependency of the molecular fields.

It is evident that the nature of the molecular hydrophobic field depends strongly on both molecular geometry and conformation. This has been shown experimentally by spectroscopic tracing of the conformational behavior of calcium antagonists interacting with membranes [9]. When passing into the membrane, the molecules adopt a maximum hydrophobic surface. The process may be similar to the interaction with hydrophobic areas of a binding site. Hydrogen-bond interactions with the surrounding water molecules are disrupted, while Polar interactions are established within the molecule
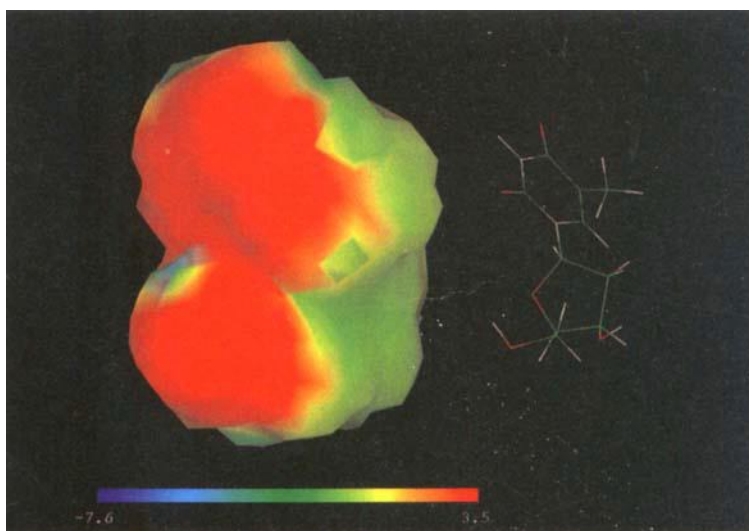
Figure 3a.
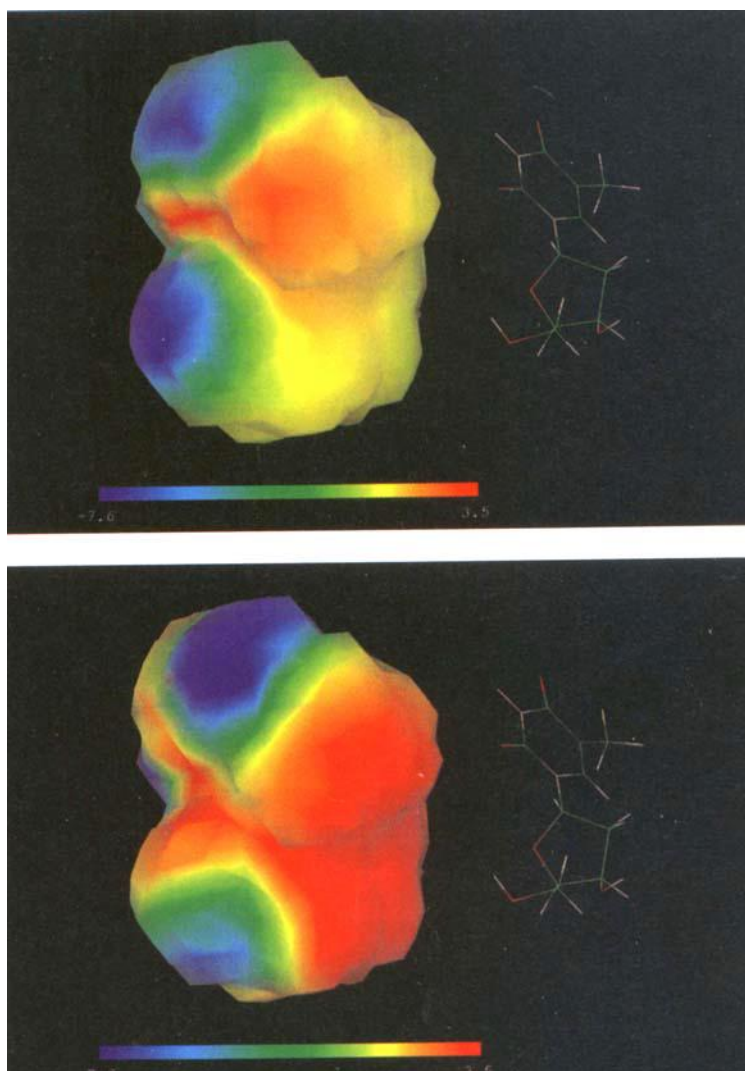
Figure 3. Thymidine hydrophobic surfaces, coming from (a) GRID, (b) HINT and (c) SYBYL algorithms. Projection on the optimum steric interaction surface (minimum of the corresponding Lennard-Jones surface) yields more or less the same pattern for all algorithms. Each of the surfaces can be used in 3D QSAR

itself and to the binding site, respectively. Desolvation effects occur both at the ligand and the interaction site. These circumstances may be not comparable with an infinite dilution, since compartmentation at or near biological membranes and/or receptor proteins may cause aggregation of molecules. The structures of the membrane provide a dense packing. All the above steps are associated with a gain in entropy; therefore, the effects should be accessible experimentally by calorimetric measurements.
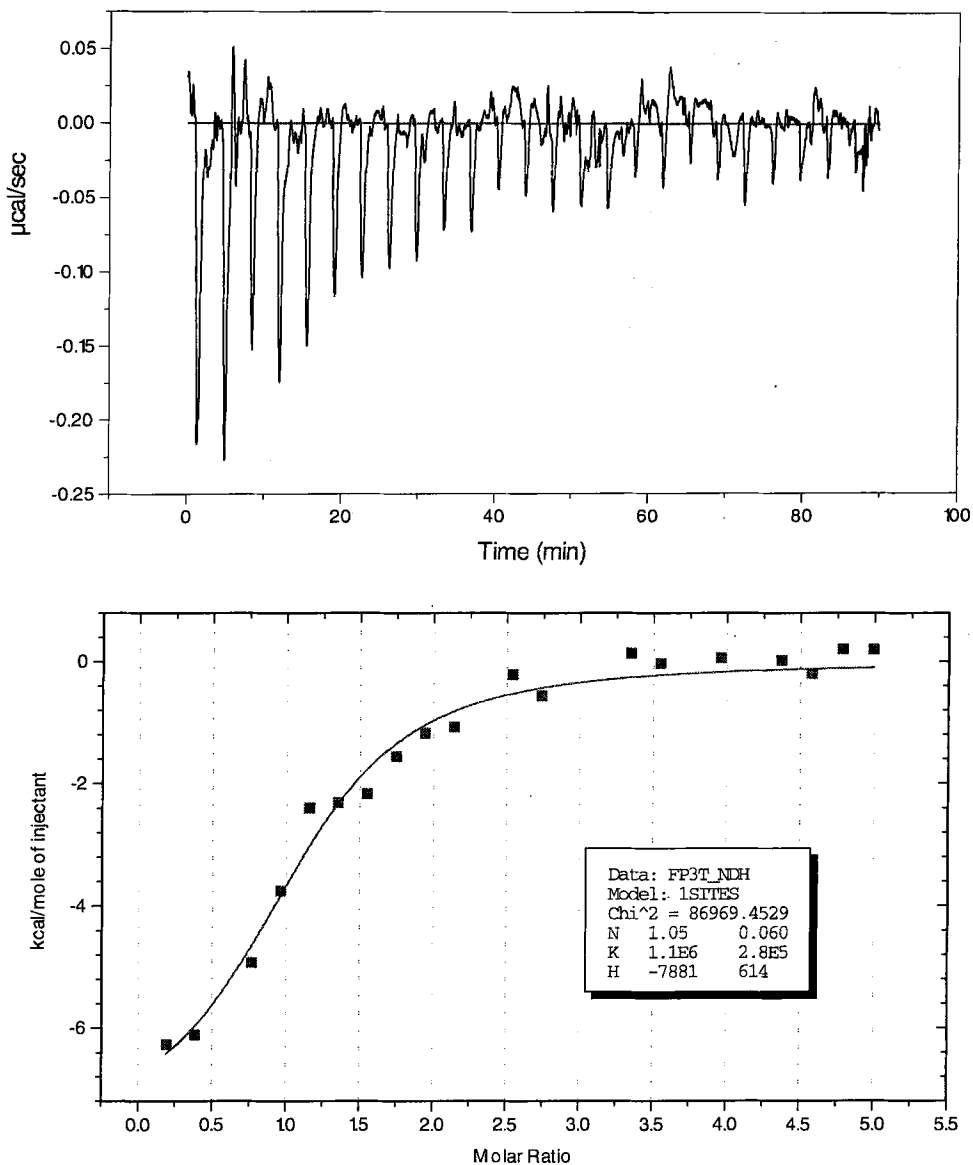
Figure 4.   Isothermal microcalorimetry. Stepwise addition of the substrate thymidine to the recombinant thymidine kinase in the microcalorimeter results in the binding curve and yields thermodynamic parameters by iterative fitting.

Calorimetric measurements on nucleic acids and their derivatives interacting with a catabolic salvage pathway enzyme thymidine kinase from herpes simplex virus (HSV1-Tk) were recently carried out in our own laboratory. We were especially interested to compare thermodynamic measurements with theoretical data from CADD (computer-
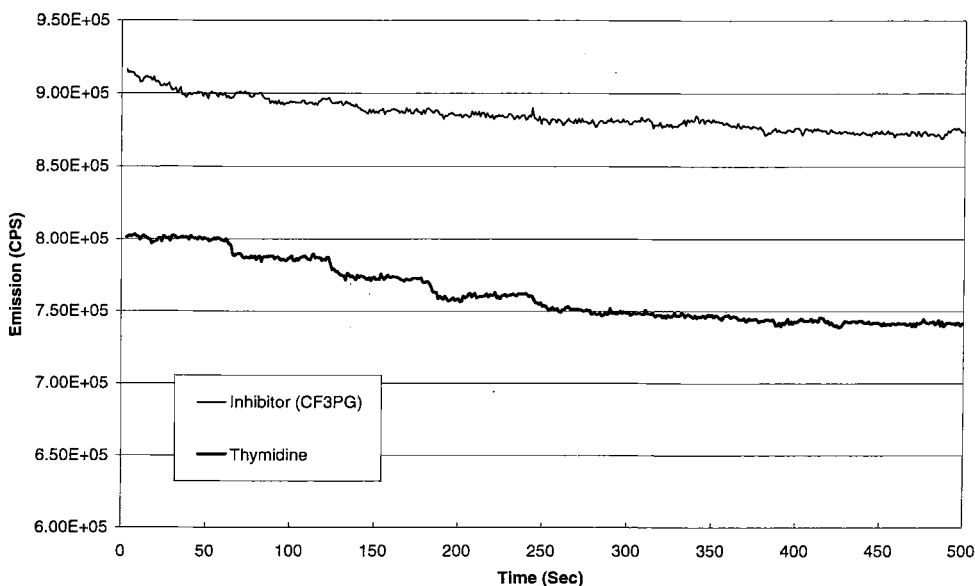
Figure 5. Microcalorimetry. Temperature-dependent fluorimetry (only binding assay) reveals that the inhibitor (upper curve) does not allow for tryptophan quenching, which is in turn observed for the natural substrate. Inhibitor binding seems therefore to be an alternative.

aided drug design) techniques. Hydrophobic fields for instance are used to describe surfaces of drug molecules (Fig. 3). These surface descriptions should correlate with enthalpic or entropic contributions to the binding of an inhibitor, a substrate or an antagonist and agonist respectively.

In principle, this technique should provide information about the size of the optimal hydrophobic surface and hence the associated conformation. Subsequently, an idea might be derived of the appropriate interaction geometry of the ligand.

Thymidine – the natural substrate of HSV1-Tk – was chosen for the experiment and found to substitute the guanines that were shown to inhibit the enzyme activity. The viral protein has been cloned and expressed in large amounts in our laboratory [31].

Isothermal microcalorimetry resulted in a clear-cut differentiation between the substrate and inhibitor. Whereas the thymidine yielded a clearly S-shaped binding isotherm, the guanine inhibitor produced only a horizontal line (Fig. 4). In the experiment, the heat of formation of the interaction complex was directly measured. From such data it can be concluded, that the inhibitor binding is a purely entropy-driven process, the binding of both substrate and inhibitor – determined independently by standard methods – showing high-affinity constants.

In order to assess the entropic contribution in a less indirect manner, we used fluorimetry with a step-by-step increase of temperature. HSV1-Tk contains four tryptophane residues that are sensitive to fluorescence excitation. Whereas the binding reaction of the natural substrate could be followed with saturation at eight different temperatures, the binding of the inhibitor again produced a horizontal line (Fig. 5).
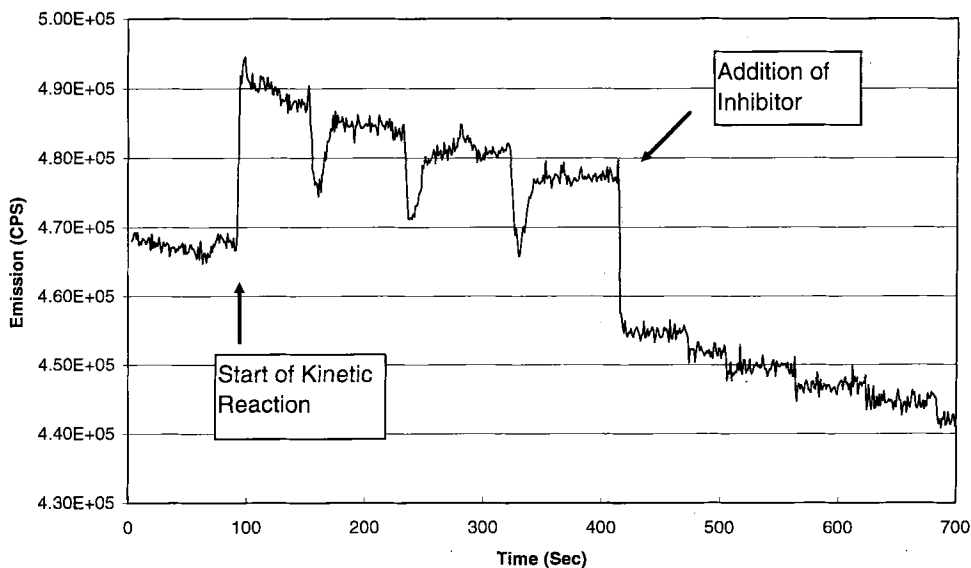
Figure 6. Fluorimetric assay. In the functional assay, the reaction starts with the addition of Mg^{2+}, which obviously provides a type of induced fit by microfolding of the active site. The quenching is markedly increased. Adding the inhibitor stops both the reaction and the quenching

On this occasion the horizontal line indicated a lack of fluorescence quenching; however, since the binding parameters of the inhibitor are known from kinetic experiments, this behavior must be interpreted as multiple, or alternate, binding mode. Repeating the same experiment in the presence of magnesium – which starts the enzymatic reaction by complexing the cosubstrate and inducing microfolding of the active site – clearly showed breakdown of the reaction at the precise moment that the inhibitor is added to the mixture (Fig. 6). The graph shows a lack of reversible fluorescence quenching, which can only be interpreted by an alternate interaction of the inhibitor with the active site. Such interpretation is strongly supported by the X-ray structure of the HSV1 Tk co-crystallized with its substrates and co-substrates [32]. One of the tryptophane residues is located in the active site, just perpendicular to the thymidine molecule. There is obviously no further space to bind a larger molecule such as guanine without changing the geometry of the binding site and the binding geometry of the guanine itself with respect to the thymidine.

13.9 Outlook

Hydrophobic fields have proven to be valuable tools in the interpretation of biochemical and structural data. They are, therefore, of great importance for the modeling of the protein-ligand complexes, that unquestionably form the basis for the rational design of drugs in the future.

So far, however, many effects of hydrophobic forces are not fully understood. New features of hydrophobic interactions are appearing as an increasing number of complex macromolecular X-ray and NMR structures. Examples of the latter, which are of great scientific interest, are macromolecular oligomerization or organization on membranes, protein-DNA/RNA interaction and cellular contacts. Beyond such molecular complexes, the biopharmaceutical and material sciences are awaiting new insights into hydrophobic effects.

References

[1] Fedor, M. J., *Nature Structural Biology*, **1**, 267–269 (1994)
[2] Kim, J. L., Nikolov, D. B., and Burley, S. K., *Nature* **365**, 520–527 (1993)
[3] Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B., *Nature* **365**, 512–520 (1993)
[4] Majerfeld, I., and Yarus, M., *Nature Structural Biology* **1**, 287–292 (1994)
[5] Abraham, D. J., and Kellogg, G. E., Hydrophobic Fields. In: *3D-QSAR in Drug Design.* Kubinyi, H. (Ed.). Escom: Leiden 506–522 (1993)
[6] Williams, D. E., and Stouch, T. R., *J. Comp. Chem.* **14**, 1066–1076 (1993)
[7] Herbette, L. A structural model for drug interactions with biological membranes: Hydrophobicity, hydrophilicity and amphiphilicity in drug structures. In: *Trends in QSAR and Molecular Modelling 92*. Wermuth, C. G. (Ed.). Escom: Leiden 76–85 (1993)
[8] Israelachvili, J., *Intermolecular and Surface Forces*, 2nd ed. Academic Press, London, 1994
[9] Blokzijl, W., and Engberts, J. B. F. N., *Angew. Chem.* **105**, 1610–1648 (1993)
[10] Kubinyi, H., *QSAR: Hansch Analysis and Related Approaches*. VCH: Weinheim, 1993
[11] Rekker, R. F., *The Hydrophobic Fragmental Constant*. Elsevier: Amsterdam, 1977
[12] Hansch, C., and Leo, A. J., *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Wiley: New York, 1979
[13] Lee, B., and Richards, F. M., *J. Mol. Biol.* **55**, 379–400 (1971)
[14] Ghose, A. K., and Crippen, G. M., *J. Comput. Chem.* **7**, 565–577 (1986)
[15] Goodford, P. J., *J. Med. Chem.* **28**, 849–857 (1985)
[16] Zimmermann, N., Rötzschke, O., Falk, K., Rognan, D., Folkers, G., Rammensee, H. G., and Jung, G., *Angew. Chem.* **104**, 929–931 (1992)
[17] Mason, K. A., Katz, A. H., and Shen, C. F., Grid-assisted similarity perception (GRASP): a new method of overlapping molecular structures. In: *Trends in QSAR and Molecular Modelling 92*. Wermuth, C. G. (Ed.). Escom: Leiden 394–395 (1993)
[18] Rognan, D., Scapozza, L., Folkers, G. and Daser, A., *Proc. Natl. Acad. Sci. USA* **92**, 753–757 (1995)
[19] Folkers, G., Trumpp-Kallmeyer, S., Gutbrod, O., Krickl, S. Fetzer, J. and Keil, G., *J. Comput.-Aided Mol. Design* **5**, 382–404 (1991)
[20] Audry, E., Dubost, J. P., Colleter J. C., and Dallet, Ph., *Eur. J. Med. Chem.* **21**, 71–72 (1984)
[21] Dubost, J. P., 2D and 3D lipophilicity parameters in QSAR. In: *Trends in QSAR and Molecular Modelling 92*. Wermuth, C. G. (Ed.). Escom: Leiden; 93–100 (1993)
[22] Broto, P., Moreau, G., and Vandycke, C., *Eur. J. Med. Chem.* **19**, 66–70 (1984)
[23] Abraham, D. J., and Leo, A. J., *Proteins* **2**, 130–152 (1987)
[24] Welch, W., Ahmad, S., Airey, J. A., Gerzon, K., Humerickhouse, R. A., Besch, H. R., Ruest, L., Deslongchamps, P., and Sutko, J. L., *Biochemistry* **33**, 6074–6085 (1994)
[25] Cramer, R. D. III, Patterson, D. E., and Bunce, J. D., *J. Am. Chem. Soc.* **110**, 5959–5967 (1988)
[26] Schubert-Wright, Ch., and Jaeger, J., *J. Mol. Biol.* **232**, 620–638 (1993)

[27] Vedani, A., Zbinden, P., Snyder, J. P., and Greenidge, P., *J. Am. Chem. Soc.* **117**, 4987–4994 (1995)

[28] Furet, P., Sele, A., and Cohen, N. C., *J. Mol. Graphics* **6**, 182–189 (1988)

[29] Greenidge, P. A., Merz, A., and Folkers, G., *J. Comput.-Aided Mol. Design* (in press)

[30] Folkers, G., Merz, A., and Rognan, D., CoMFA as a tool for active site modelling. In: *Trends in QSAR and Molecular Modelling 92*. Wermuth, C. G. (Ed.). Escom: Leiden 233–244 (1993)

[31] Fetzer, J., Michael, M., Bohner, M., Hofbauer, R., and Folkers, G., *Prot. Expr. Purification* **5**, 432–441 (1994)

[32] Wild, K., Bohner, T., Aubry, A., Folkers, G., and Schulz, G. E., *FEBS Lett.* **368,** 289–292 (1995)

14 Physico-chemical and Biological Factors that Influence a Drug's Cellular Permeability by Passive Diffusion

Robert A. Conradi, Philip S. Burton and Ronald T. Borchardt

Abbreviations

AP	Apical plasma membrane
BL	Basolateral plasma membrane
BB	Brain/blood concentration ratio
CNS	Central nervous system
P-gp	P-glycoprotein

Symbols

D_j — Membrane diffusion coefficient for solute j
P_j — Membrane partition coefficient for solute j
$P_{alk/w}$ — Partition coefficient between alkane and water
$P_{oct/w}$ — Partition coefficient between octanol and water
$P_{oct/w, app}$ — Apparent partition coefficient between octanol and water for an ionizable compound at the pH of the study
$\Delta \log P$ — $\log P_{oct/w}$ minus $\log P_{alk/w}$
$P_{H/G}$ — Partition coefficient between heptane and ethylene glycol
p_e — Effective permeability coefficient for a biological barrier
p_{mono} — Permeability coefficient for a cell monolayer corrected to exclude the contribution of the aqueous boundary layer
p_j — Membrane permeability coefficient of solute j
r_j', r_j'' — Interfacial transfer resistances
x — Membrane thickness

14.1 Introduction

Successful drug development requires not only optimization of specific and potent pharmacological activity at the target site, but also efficient delivery to that site. Due to advances in rational drug design, many promising pharmacologically active agents are being synthesized, yet clinical development is often stymied by delivery problems [1, 2]. Overcoming these problems will require a thorough understanding of the physico-chemical and biological barriers which stand between the point of administration and the pharmacological target in the body. Such knowledge will enable efficient delivery characteristics to be factored into the rational design equation.

14.1.1 Cellular Barriers to Drug Transport

To illustrate some of the relevant physiological barriers, consider oral administration of a central nervous system (CNS) active agent (Fig. 1). If formulated in a solid dosage form, the drug molecule must first break free from the strong attractions of its neighbors in the process of dissolution. The molecule must then avoid being metabolized or sequestered while it finds its way to the gut wall. To reach the circulation, the drug must traverse the mucus layer, the intestinal epithelia, the basal lamina and the endothelial cells making up the intestinal capillaries, again avoiding potential metabolism. Once in the blood, the drug may undergo metabolism, protein binding, and/or clearance by the liver and kidneys. If it succeeds in reaching the cerebral circulation, the drug must be able to desorb from blood proteins and cross the formidable blood-brain barrier. Considering this complex array of obstacles, it is not surprising that delivery to the target site can be challenging. In this chapter we set aside problems of dissolution, metabolism, and clearance in order to focus specifically on those factors limiting drug permeation across cellular barriers.

The important general features of a cellular barrier may be seen in the intestinal epithelium pictured in Fig. 2. The cell itself is a complex structure consisting of a nucleus, various organelles, and soluble protein within an aqueous cytoplasmic compartment bounded by the plasma membrane. The plasma membrane is comprised of a complex mixture of neutral and charged lipids and both integral and peripheral protein which serve both structural and metabolic functions for the maintenance of the cell [3]. In intestinal epithelial cells, the apical (AP) plasma membrane, also known as the brush-border membrane, is further organized into a lattice of microvilli which functionally increase the interfacial area. These cells are linked to each other by a continuous junctional complex referred to collectively as the *zonula occludens* or tight junction. This interconnected monolayer of cells is the principle permeation barrier for oral absorption of drugs. Similarly, a special class of capillary endothelial cells interlinked by exceptionally tight junctions constitute the principle barrier for drug transport from blood to brain.
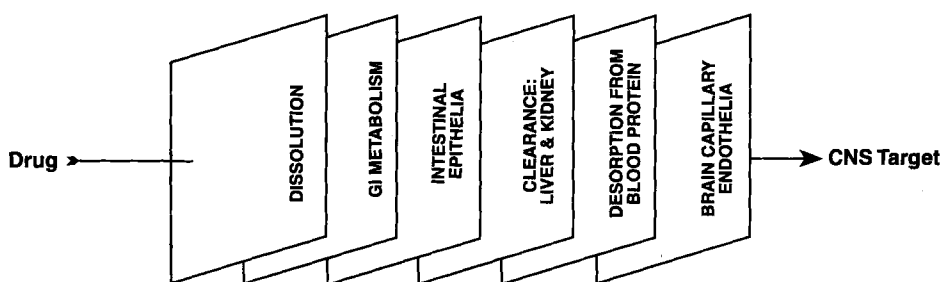
Figure 1. Schematic diagram depicting various barriers that a CNS drug must circumvent in order to elicit the desired pharmacological response

Figure 2. Intestinal epithelium of the villus from a starved rat. Several complete absorptive cells and a portion of a goblet cell are shown. The internal complexity and polarity of the epithelial cells is evident. The apical surface is composed of closely packed microvilli (Mv); the adjacent cytoplasm, which is relatively free of organelles, is the region of the terminal web (TW). Below the TW, the cytoplasm contains smooth endoplasmic reticulum (SER), while rough endoplasmic reticulum (RER) is found somewhat deeper. The Golgi complex (G) is found immediately above the nucleus. Mitochondria are widely distributed with a greater concentration in the infranuclear cytoplasm. The closely apposed lateral membranes are sometimes folded ($\rightarrow$), and below the nucleus may form interdigitating processes (P). The intercellular space is often wider in these lower regions (*). The basal lamina (BL) serves as the base for the confluent absorptive cells and separates them from the lamina propria (LP). Magnification 2700. Reproduced from *The Journal of Cell Biology* (1967) **34**, 123, by copyright permission of the Rockefeller University Press and the principal author, R. R. Cardell)

14.1.2 Transport Pathways

14.1.2.1 Paracellular Transport

As shown schematically in Fig. 3, drug molecules may cross cellular barriers either by paracellular diffusion between the cells, or transcellular diffusion across the cells. In the case of paracellular diffusion, the ability of a solute to move through this space is
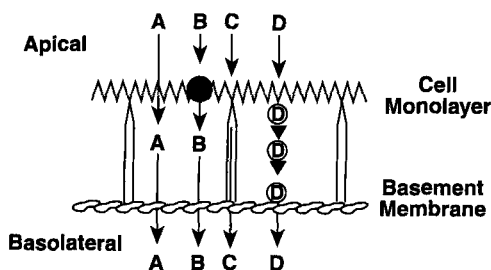
Figure 3.   Pathways by which a drug may cross a confluent monolayer. A, transcellular passive diffusion through the plasma membranes and the cytoplasmic compartment; B, carrier-mediated uptake and passive diffusion; C, paracellular passive diffusion through the intercellular space; D, transcytosis: receptor-mediated or adsorptive endocytosis, followed by vesicle migration, and exocytosis at the basolateral membrane

limited by the presence of the tight junctions. The tight junction is a region where the outer leaflets of the lipid bilayer comprising the plasma membrane of neighboring cells are fused [4, 5]. The contacts between the cells is composed of anastomosing "strands" which are continuous chains of membrane protein whose composition is presently unknown [6]. It is hypothesized that this protein complex contains narrow "gates" or "pores" which open or close, thus creating a dynamic seive which regulates molecular diffusion [7]. Further, these transient pores are thought to contain fixed negative charge thus conveying electrical selectivity [8]. The number and complexity of the strands, the number of "gates" and the fraction open at any moment all dictate the path length which a molecule must travel. These barrier properties vary with cell type, being very restrictive in brain capillary endothelia and less restrictive in gut epithelia or kidney proximal tubule, for example. Transport across tight junctions has been successfully modeled in terms of molecular size resistricted diffusion within an electrostatic field of force [9].

It has been shown that perturbation of the tight junction may be achieved in response to depletion of extracellular calcium, addition of cytochalasin D, or addition of various so-called drug absorption promoters [10]. The paracellular route may be the primary pathway by which relatively low molecular weight hydrophilic molecules cross epithelial barriers. Transport of larger molecules may be enhanced through modulation of the junctional pores. However, highly lipophilic molecules generally do not show enhanced transport upon opening the junctions because their affinity for cell membranes precludes significant diffusion through the aqueous intercellular space [11].

14.1.2.2 Transcellular Transport

While the paracellular route may contribute significantly to the transport of some molecules, the majority of common drugs traverse cellular barriers by the transcellular pathway. This route involves movement of a molecule across or through the cell. For such a process to occur, the solute must interact with some component of the plasma

cell membrane. In some cases, the integral membrane proteins present may serve as specific recognition sites for carrier-mediated transport [12]. Alternatively, after binding to a cellular component, the plasma membrane may invaginate to form a vesicle which may diffuse across the cell and fuse with the basolateral membrane in a process called transcytosis [13]. However, for the vast majority of drugs no such specific mechanisms exist, rather, transport is mediated by passive diffusion of the drug through the apical plasma membrane, across the cell proper and across the basolateral membrane. How molecules negotiate the complex cytoplasmic milieu is unknown and probably involves multiple pathways, perhaps including lateral diffusion within the membrane. Nonetheless, whatever may be the route across the cytoplasm, the drug molecule must at least cross the AP and BL membranes. The possibility of a drug simply adsorbing to the exoplasmic leaflet of the plasma membrane and diffusing laterally around the outside of the cell is precluded by the tight junctions which block lateral diffusion in the outer leaflet of the bilayer [4]. On the other hand, lateral diffusion across a cell is possible in the inner, cytoplasmic leaflet [14, 15]. Even this pathway, however, requires an initial passage of substrate to the inner leaflet of the AP membrane and a subsequent return to the outer leaflet of the BL membrane. It is these transmembrane events which are believed to be the rate-limiting steps in transcellular diffusion of most drugs.

14.2 Physico-chemical Factors Influencing Transcellular Passive Diffusion

From the preceding discussion, it is clear that transcellular passive diffusion is the result of a number of different processes. However, this complicated series of events is frequently modeled as transport across a single membrane as the relevant transport barrier. Such a convention will also be followed here, with the explicit understanding of the limitations imposed by these unphysiological assumptions.

One of the first models of solute membrane transport goes back to work by Overton around the turn of the century. Overton's rules suggested that membranes present an oil-like barrier to solutes [16]. Simply stated, the ability of a molecule to permeate a membrane should be related to its capacity to partition into the membrane phase and its diffusion coefficient within that phase. These concepts may be summarized by the simple formula:

$$p_j = P_j D_j / x \tag{1}$$

where p_j is the permeability coefficient for solute j, P_j and D_j are the membrane partition coefficient and diffusion coefficient, respectively, for solute j, and x is the thickness of the membrane [17].

In fact, membranes are not homogenous oil phases; rather, the phospholipids and related amphiphilic molecules making up the bilayer are organized such that their polar "head" groups are on the exterior bordering the aqueous phase and the lipid "tails" extended toward the center of the bilayer. The polar head groups have varied functionalities including charged groups and complex carbohydrates. The lipid tail region

Water Phase

Polar Region
Highly-Ordered
Lipid Region

Relatively-Disordered
Lipid Region

Highly-Ordered
Lipid Region
Polar Region
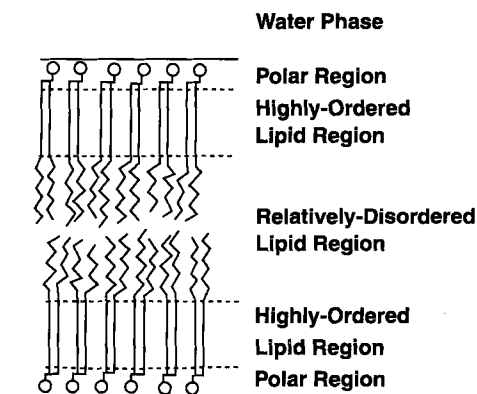
Figure 4. A schematic illustration depicting three distinct regions within a bilayer membrane. (Adapted from [20])

includes saturated and unsaturated fatty acyl chains and variable amounts of cholesterol. Importantly, biological membranes are not symmetrical, the different phospholipid classes being unequally represented on the two leaflets of the bilayer [18, 19]. Interfacial forces constrain the membrane lipid tails ordering them in a way that varies with the depth from the surface. While the innermost region of the bilayer is relatively disordered, the region closer to the interface is highly ordered [20–22]. This arrangement is illustrated in Fig. 4. A consequence of this bilayer architecture is that both solvent properties and diffusion coefficients vary as a function of depth in the membrane. When measuring a membrane/water partition coefficient, most of the membrane-associated solute will concentrate in the domain with lowest potential energy. However, it is the domain in which the solute has the highest potential energy – the domain from which solute is most excluded – which will most impede transflux across the membrane [17]. Thus, membrane partition coefficients disproportionately represent partitioning into the region of the membrane of least relevance to solute transport.

In order to identify the solute properties which are conducive to facile transport we must then determine the properties of the rate-controlling region of the membrane. There are two approaches to this problem, the first of which is based on the partition-diffusion model simplistically formulated in Eq. 1. Since we are now aware that both the solvent properties and the diffusion coefficient vary with depth, the equation becomes:

$$p_j = \int_o^{x_o} \frac{P_j(x)D_j(x)}{dx} \tag{2}$$

where $P_j(x)$ and $D_j(x)$ represent the partition and diffusion coefficient for solute j which varies as a function of x [17]. The membrane environment will be more polar in the vicinity of the head groups, while polarity will be extremely low in the center of the bilayer. Using the "like dissolves like" rule, lipophilic molecules would be expected to prefer an interior environment while more polar solutes would be found in a more peripheral location. The diffusion coefficient will be lowest in the highly ordered regions illustrated in Fig. 4 and will be higher in the less-ordered center. The product of P_j and D_j will have a minima with respect to depth which depends upon the properties of the
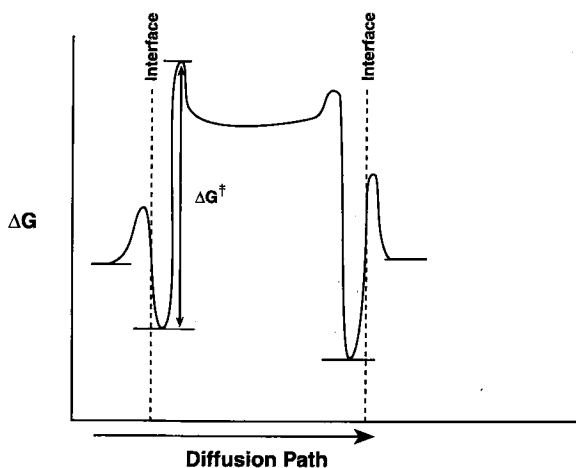
Figure 5. A free energy plot illustrating the transport energy profile for a hypothetical amphiphilic solute crossing a bilayer (barriers to solute transport)

solute and the particular membrane under consideration. To a first approximation, the region of the minima can be considered the rate-determining barrier domain [20]. This then is the microenvironment which must be emulated if a partition coefficient is to be used to predict permeability.

An alternative approach to defining the rate-limiting barrier is based on transition state theory as elaborated by Eyring and coworkers [23, 24]. A diffusing solute is pictured as moving along a path in a series of jumps from one equilibrium state to the next. A certain activation energy is required for each jump. At phase boundaries the solute must move through a higher energy transition state and these more difficult jumps constitute the greatest barriers to diffusion through a membrane. The concept of phase boundaries may be expanded to include sharp transitions within the membrane such as the boundary between the interfacial domain and the interior [25–27]. A hypothetical energy profile is pictured in Fig. 5. The essential distinction between the partition-diffusion model and the transition state model is that the former focuses on diffusion through barrier domains while the latter emphasizes movement across interfaces.

Diamond and Katz [17] proposed a unifying treatment which includes contributions from both models as barriers in series. Thus Eq. 2 is modified to yield:

$$\frac{1}{p_j} = r'_j + r''_j + \int_0^{x_o} \frac{dx}{P_j(x)D_j(x)} \tag{3}$$

where r'_j and r''_j are interfacial transfer resistances. In this formulation it is acknowledged that either phase transitions or retarded diffusion across an unfavorable domain could be rate limiting for a given solute. The equation suggests that the initial adsorption at the membrane/water interface, diffusion to the opposite interface, or desorption from the opposite interface could control the rate of transflux across a membrane. Again, the physico-chemical properties of the solute would determine which is most important.

Since it is not usually practicable to measure directly the balance of solute-membrane microdomain forces which may be important in the overall transport process, partition coefficients in some artificial system are usually employed as a surrogate measurement. As is outlined in more depth elsewhere in this book, partitioning behavior is determined by a number of forces including solvent-solvent, solute-solvent, and sometimes solute-solute interactions. Briefly, since cohesive solvent-solvent interactions in the polar phase (usually water) must be disrupted to make a solute-accommodating cavity, solute is driven into the less polar phase as a function of its molecular volume [28]. This cohesive energy, augmented by the increase in entropy resulting from reduction of the contact area between water and the hydrophobic surfaces of the solute, constitute the hydrophobic effect. Opposing these forces are any interactions between solute and the polar phase (such as hydrogen bonding or dipole-dipole attractions) which are lost upon transfer to the more lipidic phase. The relative contributions of these various interactions to the observed partition coefficient depends on the nature of the solvent pair.

14.2.1 Predictive Partition Coefficients

One method for appraising this balance of forces is the solvatochromic comparison method pioneered by Kamlet et al. [29, 30]. Multiple regression is used to quantitate the influence of independently determined solute properties on an experimentally determined partition coefficient. Thus, the log of the permeability coefficient is set equal to the sum of a collection of physical parameters, each multiplied by a weighting coefficient computed to optimize the fit. Solute properties included in this analysis are molecular volume, polarity/polarizability, hydrogen bond accepting ability, and hydrogen bond donating ability. Table 1 presents some applications of this method [27, 31]. In addition to the log of the partition coefficients between octanol/water and alkane/water, a $\Delta \log P$ parameter is listed which is the difference between these terms. Seiler first introduced this parameter as a crude measure of the hydrogen bonding capacity of a solute [32]. This $\Delta \log P$ is equivalent to the hypothetical partition coefficient between water-saturated octanol and alkane. The apparent usefulness of this parameter encouraged us to measure partitioning in the heptane/ethylene glycol system which we assumed would represent a similar balance of forces and yet not require the determination of two separate partition coefficients [27]. As can be seen from Table 1, the relative importance of the solvatochromic parameters for each partitioning system is different. Log $P_{oct/w}$ is seen to be most influenced by volume and thus by the hydrophobic effect while $\Delta \log P$ is primarily a function of the solute's hydrogen bond donor ability [31]. A more complex mix of forces govern the other partition coefficients. How well the balance of forces represented by a particular partition coefficient compares with the forces controlling transport through membranes will determine the utility of that coefficient for predicting permeability.

Table 14.1. Regression coefficients with 95 % confidence limits showing relative contribution of the solvatochromic parameters to partition coefficients in various systems

$\log P = a\alpha + b\beta + d\pi^* + c(V_I/100) + \log P_0$				
Solvent system	a	b	d	c
Octanol/water ($\log P_{O/W}$)	-0.15 ± 0.23	-3.51 ± 0.38	-0.74 ± 0.31	5.83 ± 0.53
Heptane/water ($\log P_{H/W}$)	-3.54 ± 0.30	-5.35 ± 0.5	-1.02 ± 0.39	6.78 ± 0.69
Heptane/octanol ($\Delta\log P$)	-3.40 ± 0.25	-1.96 ± 0.42	-0.12 ± 0.30	Negligible
Heptane/glycol ($\log P_{H/G}$)	-4.41 ± 0.40	-1.69 ± 0.58	-1.53 ± 0.76	2.79 ± 0.52

Note: V_I, the molecular volume in $cm^3 \ mol^{-1}$, is divided by 100 to roughly normalize this parameter with respect to the other parameters. The parameters α, β, and π^* represent a solute's hydrogen bond-donating ability, hydrogen bond-accepting ability, and polarity/polarizability respectively.

14.2.2 Relationship to a Drug's Lipophilicity

The most commonly used measure of lipophilicity as a predictor of solute membrane partitioning and hence transport is the log of the octanol/water partition coefficient ($P_{oct/w}$). The relationship of solute absorption or permeability with $P_{oct/w}$ has been studied in a large number of systems including *in vivo* intestinal absorption models [33, 34], blood-brain barrier models [35] and cell culture models [36, 37] to name a few. Plots of effective permeability coefficient (p_e) vs log $P_{oct/w}$, where permeability is determined by solute disappearance from the donor solution, have a sigmoid appearance. Mechanistically this is interpreted in terms of a model where very polar molecules, with small or negative log $P_{oct/w}$, are relatively membrane-impermeable and presumably cross the absorptive barrier by the relatively inefficient paracellular pathway [38]. As lipophilicity, log $P_{oct/w}$, increases, membrane permeability increases until a plateau region is reached beyond which permeability becomes independent of log $P_{oct/w}$. The plateau region corresponds to the situation where membrane transport has become so rapid that convective diffusion through the unstirred aqueous layer immediately adjacent to the membrane becomes rate limiting [38]. Studies demonstrating these sigmoid curves generally involve a solute set with rather homologous structures and/or minimal functionality as illustrated in Fig. 6a, b [39]. However, when data for unrelated sets are plotted together on the same graph as in Fig. 6c an overall sigmoid relationship is not obvious. Consequently, compounds with similar permeability coefficients may have partition coefficients differing by 5–6 orders of magnitude. Thus, while log $P_{oct/w}$ may be a useful parameter for predicting permeability within a carefully circumscribed set of substrates, it is not a universal indicator of a solute's transport potential.
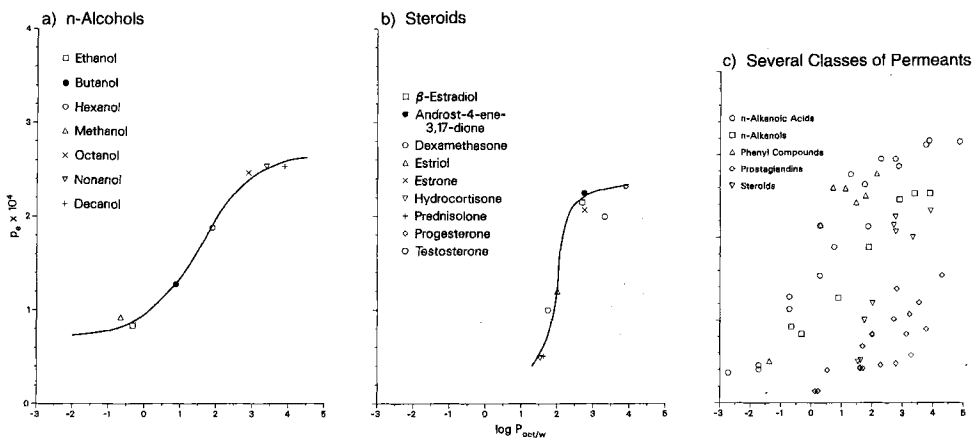
Figure 6. Plots showing the relationship between solute permeability coefficients in the rat jejunum and their octanol/water partition coefficients. (a) Data for a set of alcohols; (b) data for a set of steroids; (c) combined data for various classes of solutes.

14.2.3 Relationship to a Drug's Hydrogen Bonding Potential

14.2.3.1 Intestinal Mucosal Cell Transport

Fig. 7a shows the relationship of permeability, in Caco-2 cells as a model of the human intestinal mucosa, with $P_{oct/w}$ for a series of model peptide oligomers prepared from D-phenylalanine [40, 41]. For the zwitterionic series, p_{mono} (the permeability coefficient for the cell monolayer alone after correcting for the contribution of the aqueous boundary layer) was very low for all of the compounds and was inversely correlated with log $P_{oct/w}$. When the N and C termini of these zwitterionic oligomers were capped to eliminate charge, permeability increased substantially but the inverse relationship with log $P_{oct/w}$ remained. These results demonstrated that charge has an adverse effect on transport as expected from the pH partition model for drug transport [42]. In fact, the transport of the charged compounds was consistent with a paracellular pathway in which the size of the molecules controls passage through the narrow pores of the tight junctions [9].

More surprising was the decrease in permeability of the non-electrolyte peptides with increasing log $P_{oct/w}$. This trend suggested that some property other than lipophilicity was controlling permeability. We postulated that this permeability decrease was either due to strongly size-dependent diffusion or to the greater energy required for desolvation of the extra amide bonds in the larger oligomers upon passage into a less polar phase. To distinguish between these two possibilities a series of tetrapeptide analogs was prepared from Ac-$(DPhe)_3$-NH_2 in which the amide nitrogens were sequentially alkylated with from 1–4 methyl groups. This effectively maintained a constant chain length while the solvation potential was incrementally decreased with each additional methyl group introduced. Consistent with these expectations, fairly small changes in molecular weight and log $P_{oct/w}$ (Fig. 7b), but more substantial changes in
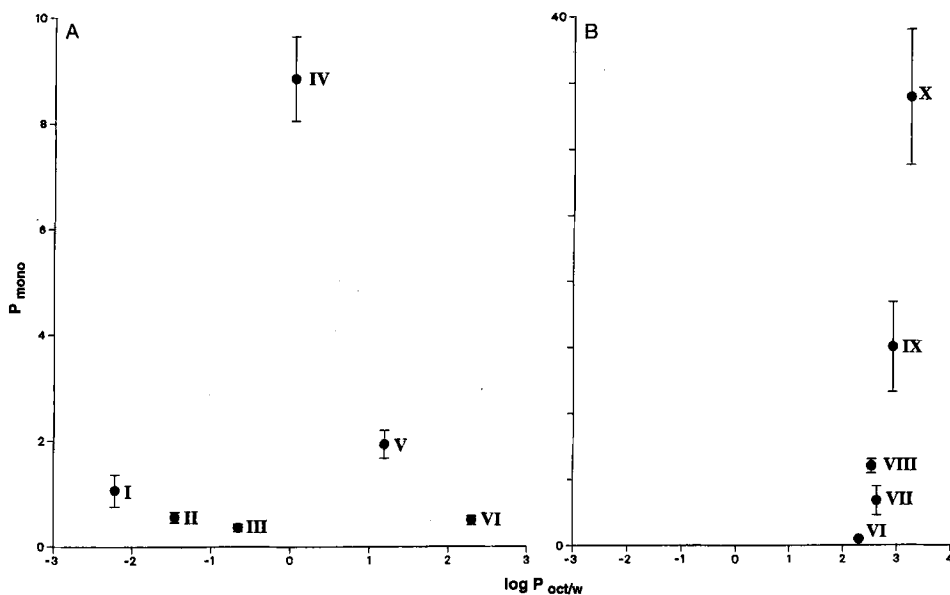
Figure 7. The relationship of model peptide permeability coefficients in Caco-2 cell monolayers to their octanol/water partition coefficients. (a) Zwitterionic PheGly (I), Phe_2Gly (II), Phe_3Gly (III) and neutral $AcPheNH_2$ (IV), $AcPhe_2NH_2$ (V), $AcPhe_3NH_2$ (VI) peptides containing from 1–3 D-phenylalanine residues. (b) Tetrapeptide mimetics containing from 1–4 N-methyl groups: $AcPhe_2(NMePhe)NH_2$ (VII), $AcPhe(NMePhe)_2NH_2$ (VIII), $Ac(NMePhe)_3NH_2$ (IX) and $Ac(NMePhe)_3NHMe$ (X)

Δ log P, a measure of H-bonding capacity were found. The p_{mono} values of these peptides increased with the number of N-methyl groups, suggesting that desolvation energy rather than size was controlling transport. As seen in Fig. 8, a plot of p_{mono} vs Δ log P demonstrated a reasonable correlation for all of the peptide non-electrolytes [43]. A similar correlation has been found with log $P_{H/G}$, the heptane/ethylene glycol partition coefficient [27]. These data are consistent with classical work by Stein and others showing a correlation between membrane permeability and the number of hydrogen bonds a solute can form [44, 45]. Further, Roseman has shown that the free energy associated with the transfer of a phenylalanine side chain from water to alkane is –13.6 kJ/mol while the energy cost to transfer an amide bond is 25.5 kJ/mol [46, 47]. It is reasonable, therefore, that the driving force associated with the hydrophobic effect can be overwhelmed by the energy required to desolvate an amide bond. Thus, the negative effect of hydrogen-bonding groups on permeability suggest that the barrier domain of membranes is an alkane-like environment.

The hydrogen bond effect on permeability is not a unique property of the amide moiety. Similar transport studies using peptide mimetics and peptides possessing other hydrogen-bonding functional groups display similar correlations [43, 48]. It should be emphasized, however, that what matters is not the sum of all hydrogen bonds which can form between solute and water, but the number of those bonds which are lost upon
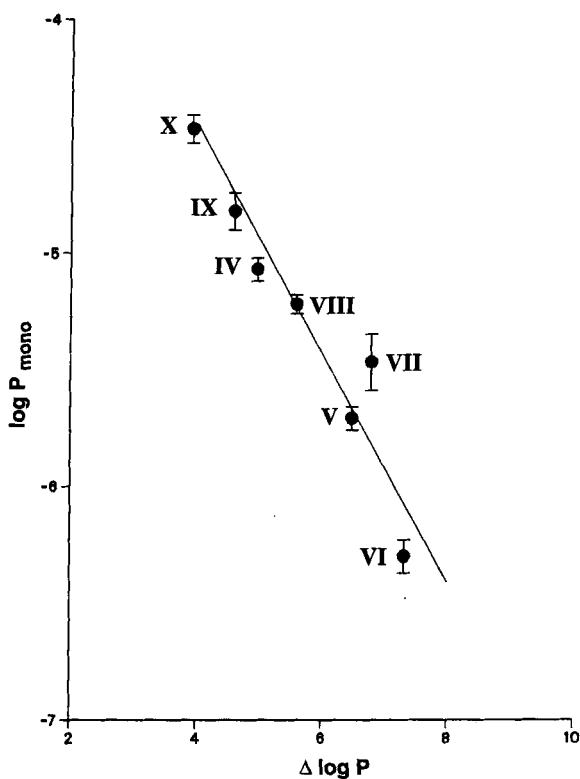
Figure 8. The linear relationship between log permeability coefficient in Caco-2 cell monolayers and $\Delta \log P$ for a set of model peptides

phase transfer. Where intramolecular hydrogen bonds can form, the resistance to transport is reduced [48]. Designing in such intramolecular interactions has been used as a strategy to improve drug transport [49]. With this caveat in mind, good correlation is seen between permeability and hydrogen-bond capacity for the more diverse solute sets.

The importance of hydrogen bonding in transport is also not limited to the Caco-2 cell model. Transport studies with *in situ* perfused rat ileum using the same model peptides as in the Caco-2 work have shown, qualitatively similar results [50]. Indeed, linear correlation of p_e with partition coefficients yielded correlation coefficients (r) of 0.86 and 0.96 for $\Delta \log P$ and $\log P_{H/G}$ respectively, but only 0.60 for $\log P_{oct/w}$. Even in a whole-rat model, intestinal absorption was highly correlated with permeability in Caco-2 cells, if the two smallest of the 14 peptides and peptide mimetics were excluded [51, 52]. These animal studies support the conclusion that the epithelial cells themselves constitute the principal barrier to peptide oral absorption *in vivo* and that Caco-2 cells represent a reasonable model of this barrier.

14.2.3.2 Blood-Brain Barrier Transport

Pardridge was among the first to suggest that drug uptake into the brain is related to the drug's potential to participate in hydrogen bonds [53]. More recently, Mitchel et al. measured the brain/blood concentration ratio (BB), a crude measure of brain uptake, for a series of potential CNS agents. They found that BB is poorly correlated with $\log P_{oct/w}$, but well correlated with $\Delta \log P$ [49, 54]. Combining these data with new determinations, Abraham et al. [55] used multiple regression to delineate the contributions of various forces to BB. Solute size was found to effect an increase in BB, while polarity/polarizability, hydrogen-bond acidity, and hydrogen-bond basicity all caused a decrease. Interestingly, a significant correlation with $\Delta \log P$ was still observed ($r = 0.885$, $n = 32$) for combined data from the two sets [55]. Thus, these data accord with an overriding influence of hydrogen bonding on uptake and a rather low molecular volume dependence.

Consistent results have been obtained for both *in vivo* and *in vitro* studies of the permeability of the phenylalanine oligomers discussed previously. Using a bovine brain microvessel endothelial cell monolayer model and an *in situ* rat brain perfusion technique, a high correlation was found between permeability and various indicators of hydrogen-bonding potential. In both models correlations with $\log P_{oct/w}$ were poor while r values > 0.94 were found in correlations with either $\Delta \log P$ or $\log P_{H/G}$ [56]. Collectively, these data suggest that the permeability barrier properties of brain capillary endothelial cells are qualitatively quite similar to those of the intestinal epithelial cells.

14.2.3.3 Mechanistic Considerations

All of the drug transport studies cited above may imply that hydrogen-bonding capacity alone is adequate to predict transport potential. However, if the barrier domain of membranes is alkane-like, one might well ask why a correlation with an alkane/water partition coefficient ($P_{alk/w}$) would not be more appropriate. In fact, correlations of permeability coefficients with $\log P_{alk/w}$ are better than with $\log P_{oct/w}$, but correlations with their difference, $\Delta \log P$, are superior to either parameter alone. One possible explanation for this is based on the Eyring transition state energy model. Jacobs and White [57] showed through neutron scattering studies that membrane-associated small peptides reside in the interfacial region of the bilayer where much of the driving force of the hydrophobic effect is satisfied without concomitant desolvation. In order to cross the membrane a peptide, or any other solute, must undergo "flip-flop" to the opposite interface [58]. Consequently, the rate-limiting phase transfer step may be between the favored interfacial domain and the very restrictive interior domain. The energetics of flip-flop may then be better modeled as a partitioning phenomenon between the *interfacial* and interior domains of the bilayer rather than between the exterior and interior domains [27]. This type of behavior is also exemplified by the transport of hydrophobic cations across bilayer membranes [25, 26].

It must also be emphasized that the relationship between permeability and hydrogen bonding does not apply to all compounds. One very important exception is the case of highly membrane-interactive compounds. Membrane transflux studies with highly

lipophilic molecules frequently display much faster disappearance from the donor compartment than appearance in the receiver. In such cases desorption from the cellular barrier becomes the rate-limiting step in transflux [59]. In cell culture models, protein constituents in the receiver solution may accelerate desorption by binding the drug [60, 61]. Thus, to predict the *in vivo* transflux of such drugs from an *in vitro* model one would need to emulate accurately the receiver compartment as well as the rate limiting membrane. Possibly the prostaglandin transport data in Fig. 6c fails to display the expected sigmoidal relationship due to rate-limiting desorption.

14.2.4 Relationship to a Drug's Solution Conformation

If hydrogen bonding, particularly in the amide functionality of peptides, plays such an important role in preventing the insertion and translocation of such molecules across cell membranes, how do native proteins overcome this barrier? Clearly, integral membrane and secreted proteins have devised mechanisms to do this. In a majority of cases, this is accomplished by means of helix formation which results in intramolecular hydrogen bonding with amide groups along the chain. This configuration results in a net decrease in the free energy of transfer from 25.9 kJ/mol for solvated amide bonds to 2.3 kJ/mol if the amide is involved in an intramolecular bond [47]. Conformational promotion of intramolecular hydrogen bonds as a mechanism for improving membrane permeability is not restricted only to proteins. Delta-sleep-inducing peptide (DSIP, MW 849) shows the unusual ability to passively diffuse across the blood-brain barrier both *in vivo* [62] and *in vitro* [63]. The solution structure of this peptide was shown to contain several intramolecular hydrogen bonds, resulting in an overall amphiphilic structure which may account for its unexpectedly high permeability [64]. Similar results were found with a homologous series of synthetic pyridylcarboxamide HIV protease inhibitors which differed only in the position of the nitrogen in the pyridine ring. The 2-pyridyl isomer was much more permeable across Caco-2 cell monolayers than either the 3- or 4-congeners. In this case, the 2-isomer also showed significantly smaller log $P_{H/G}$ than the other isomers, consistent with its permeability, and supporting the presence of a conformationally promoted intramolecular hydrogen bond which is not possible in the other isomers [48].

While the foregoing examples focus on solution conformation of a molecule which helps to mask polar groups in order to improve permeability, the idea of membrane-induced conformational changes in a molecule which favor a transient, permeable structure has also been discussed. Carrupt et al. [65] found that morphine glucuronides can exist in at least two different conformations which differ significantly in their apparent lipophilicity. They theorized that this molecule shows unexpectedly high blood-brain barrier permeability resulting from formation of the more lipophilic conformer within the biological membrane. Similarly, the hydrogen-bonding configuration and lipophilicity of cyclosporin A were shown to change in going from a polar to an apolar solvent [66]. Such a conformational change might be expected to occur at a membrane interface, where similar "solvent" property changes occur, and may help to explain the relatively high absorption of cyclosporin compared with other peptide drugs. The idea of the membrane interface as a catalyst for promoting conformational

structure which is not necessarily seen in aqueous solution has been discussed in some detail previously [67, 68]. These considerations suggest that, when exploring the relationship of physico-chemical properties with transport, the potential contribution from different conformers to these processes be considered.

14.3 Biological Factors Influencing Transcellular Passive Permeability: Polarized Efflux Systems

Another transport-limiting factor which is only recently becoming appreciated is active efflux. While the contribution of active intestinal elimination pathways to the systemic elimination of drugs has been known for some time [69, 70] the actual mechanism involved and the implications of such systems to drug absorption had not been considered in any significant way. With the identification in pleiotropic drug-resistant cancer cells of an actual transporter molecule, a potential mechanistic explanation for these phenomena was obtained. Briefly, one of the mechanisms by which previously sensitive cancer cells develop resistance to a cytotoxic agent is by increased expression of a membrane glycoprotein called P-glycoprotein or P-gp [71]. P-gp seems to work by actively reducing intracellular accumulation of the cytotoxic agent and thus effectively detoxifying the cell. It was soon found that P-gp, though exaggerated in drug-resistant cancer cells, is in fact present in many normal cell types. Among the tissues with higher constitutive expression of this transporter are intestinal epithelia [72] and brain capillary endothelia [73]. The role of P-gp in these barrier cells, although presently unknown, has been speculated to be protective as well. Due to P-gp expression, systemic or brain exposure to potentially harmful xenobiotics would be reduced. In support of this hypothesis, a recent study using mice in which the gene for P-gp has been deleted showed these animals to be about 100-fold more sensitive to the neurotoxic agent ivermectin than wild-type [74]. Although P-gp has been the focus of these recent investigations, many such transport proteins may exist which play similar protective roles.
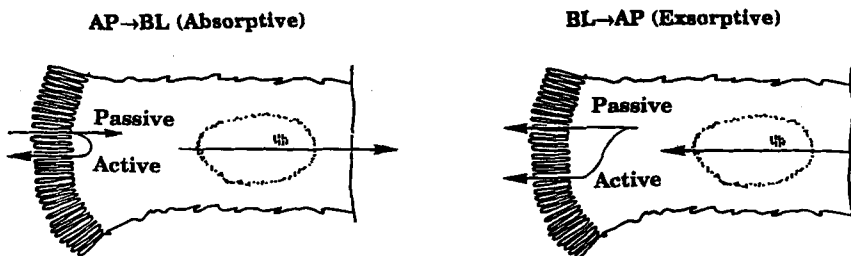
Figure 9. Influence of a polarized efflux pathway on solute flux. In this schematic representation polarized efflux (PE) will impede absorptive passive flux and will augment passive flux in the opposite direction. Since PE is a saturable system, its contribution to overall flux will depend upon intracellular solute concentration
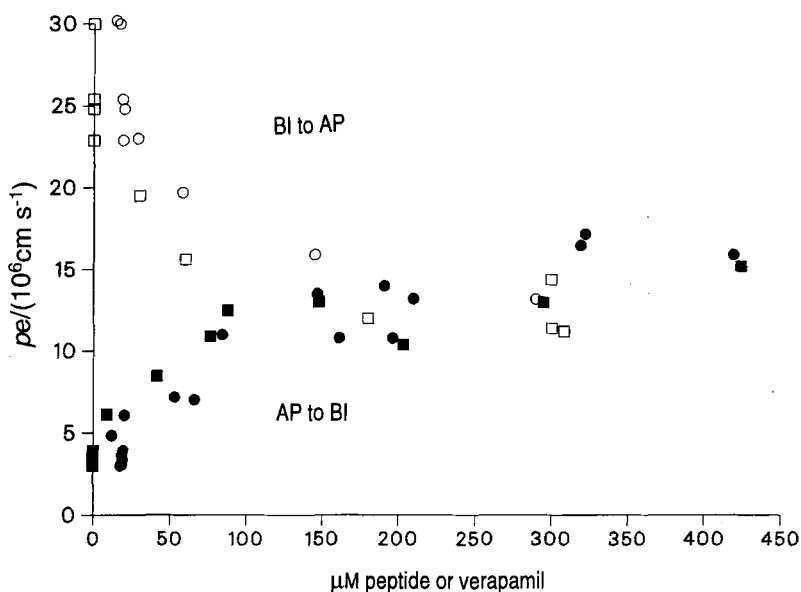
Figure 10. Plots showing the effect of substrate and verapamil concentrations on observed permeability coefficients for transport of Ac-DPhe(N-Me-DPhe)$_2$NH$_2$ across Caco-2 monolayers in the AP→BL and BL→AP direction. Circles represent varied peptide concentration with no verapamil; squares represent varied verapamil concentration with constant peptide concentration (19 μM). Open symbols are for BL→AP transport, closed symbols are for AP→BL transport

It seems clear that the presence of such active efflux systems may be advantageous from a detoxification perspective. However, with regard to drug absorption, these efflux pathways will serve as additional impediments to the transport of a solute which is a substrate, by returning a portion of the absorbed drug back to the lumen before it is able to cross the cellular barrier (Fig. 9). The characteristics of such concurrent passive and polarized active transport are illustrated for the case of peptide permeability in the Caco-2 cell model [75]. As shown in Figure 10, the permeability, p_e, for apical to basolateral (AP→BL) transflux was less than the p_e for basolateral to apical (BL→AP) transflux. As peptide concentration increased, the AP→BL p_e increased and the BL→AP p_e decreased until the two values converged at an intermediate plateau value [76]. Further, other related peptides did not display this concentration-dependent or polarized transport. These results are indicative of a saturable transport system in Caco-2 cell which accelerates solute flux in the BL→AP direction, reduces net flux in the AP→BL direction and shows substrate specificity. The transport system was also found to be inhibited by verapamil, a recognized inhibitor of P-gp. Other groups have provided evidence of similar behavior in Caco-2 monolayers using vinblastine [77], cyclosporin A [78], and taxol [79] as substrates.

Evidence for the presence of this active efflux system for peptides in rat blood-brain barrier has also been found using the *in situ* rat brain perfusion model [80]. As with Caco-2 monolayers, verapamil increased peptide uptake from the apical side of the

barrier without increasing paracellular leakage. Having shown with one peptide that the p_e reaches a plateau value when verapamil is present at 500 µM uptake of the other peptides was examined at the same concentration of verapamil. These "intrinsic" p_e values, reflecting permeability unaffected by transporter, still displayed a linear correlation with various indicators of hydrogen-bonding capacity [80]. Thus the transporter appeared to change the intercept and to a lesser degree the slope of the correlation but had little effect on r. Similar results have been found in the Caco-2 model (R. A. Conradi et al., unpublished results). This finding may be consistent with the suggestion that the "active site" for the transporter lies within the lipid bilayer and acts as a flippase [81]. In this model the efficiency of the transporter could be more dependent on the concentration of substrate in the vicinity of the transporter than on affinity. That concentration may parallel the intrinsic permeability. Whatever the mechanism, the presence of such systems will effectively serve as additional barriers to drug absorption for those molecules which are substrates.

14.4 Rationally Designing Drugs with Enhanced Cellular Permeability

Our present understanding for drug transport mechanisms offers guidance in how we approach the delivery aspects of rational drug design. First, we must recognize that the nebulous property called "lipophilicity" is inadequate in itself to predict permeability. Instead, we need to appreciate the relative contributions of individual forces such as hydrogen bonding and hydrophobic effect which together govern a solute's ability to permeate a membrane barrier. Studies such as those of Abraham and Mitchell on blood/brain partition ratios (cited above) should serve to define the relative importance of these forces. Second, we should use this knowledge to show us what part of our drug molecules should be modified to effect the greatest gain in transport. For example, adding hydrophobicity may be less effective than reducing hydrogen-bonding capacity. Third, we should recognize that factors such as conformational variability, excess lipophilicity, protein binding, or affinity for an efflux transporter may influence drug permeation in ways which simple models fail to predict. Appropriate experiments to assess the importance of such mechanisms may lead to different optimization strategies. Finally, the reality that modifications to improve transport often adversely affect target receptor binding may necessitate a compromise or the use of bioreversible derivatives, i.e., prodrugs.

References

[1] Lee, V. H. L., *Peptide and Protein Delivery*, Dekker, New York, 1991

[2] Audus, K. L., and Raub, T. J., *Biological Barriers to Protein Delivery*, Plenum Press, New York, 1993

[3] Klausner, R. D., Kempf, C., and van Renswoude, J. (Eds.). Membrane Structure and Function. In: *Current Topics in Membranes and Transport*, Vol. **29**. Academic Press: Orlando, FL (1987)

[4] Diamond, J. M., *Physiologist* **20**, 10–18 (1977)

[5] Gumbiner, B., *Am. J. Physiol.* **253**, C749–C758 (1987)

[6] Hirsch, M., and Noske, W., *Micron* **24**, 325–352 (1993)

[7] Schneeberger, E. E., and Lynch, R. D., *Am. J. Physiol.* **262**, L647–L661 (1992)

[8] Madara, J. L., Barenberger, D., and Carlsson, S., *J. Cell. Biol.* **102**, 2125–2136 (1986)

[9] Adson, A., Raub, T. J., Burton, P. S., Barsuhn, C. L., Hilgers, A. R., Audus, K. L., and Ho, N. F. H., *J. Pharm. Sci.* **83**, 1529–1536 (1994)

[10] Madara, J. L., *J. Clin. Invest.* **83**, 1089–1094 (1989)

[11] Artursson, P., and Magnusson, C., *J. Pharm. Sci.* **79**, 595–600 (1990)

[12] Stein, W. D., *Channels, Carriers and Pumps. An Introduction to Membrane Transport*, Academic Press, San Diego, CA, 1990

[13] Rodman, J. S., Mercer, R. W., and Stahl, P. D., *Curr. Opin. Cell Biol.* **2**, 664–672 (1990)

[14] Simons, K., and van Meer, G., *EMBO J.* **5**, 1455–1464 (1986)

[15] Banks, W. A., Kastin, A. J., and Barrera, C. M., *Pharm. Res.* **8**, 1345–1350 (1991)

[16] Overton, E., *Vierteljahrsschr. Naturforsch. Ges. Zürich* **44**, 88–135 (1899)

[17] Diamond, J. M., and Katz, Y., *J. Membrane Biol.* **17**, 121–154 (1974)

[18] Rothman, J. E., and Lenard, J., *Science* **195**, 743–753 (1977)

[19] Simons, K., and van Meer, G. Lipid sorting in epithelial cells. In: *Perspectives in Biochemistry*, Vol. **1**. Neurath, H. (Ed.). American Chemical Society: Washington, DC; 232–237 (1989)

[20] Xian, T.-X., and Anderson, B. D., *J. Membrane Biol.* **140**, 111–122 (1994)

[21] Dill, K. A., and Flory, P. J., *Proc. Natl. Acad. Sci. USA* **78**, 676–680 (1981)

[22] Seelig, A., and Seelig, J., *Biochemistry* **13**, 4839–4845 (1974)

[23] Zwolinski, B. J., Eyring, H., and Reese, D. E., *J. Phys. Colloid Chem.* **53**, 1426–1452 (1949)

[24] Davson, H., and Danielli, J. F., *The Permeability of Natural Membranes*. Cambridge Press: London, 1943

[25] Ketterer, B., Neumcke, B., and Läuger, P., *J. Membrane Biol.* **5**, 225–245 (1971)

[26] Miyauchi, S., Ono, A., Yoshimoto, M., and Kamo, N., *J. Pharm. Sci.* **82**, 27–31 (1993)

[27] Paterson, D. A., Conradi, R. A., Hilgers, A. R., Vidmar, T. J., and Burton, P. S., *Quant. Struct.-Act. Relat.* **13**, 4–10 (1994)

[28] Sinanoglu, O., and Abdulnur, S., *Fed. Proc.* **24**, Suppl. 15, S12–S23 (1965)

[29] Kamlet, M. J., Abraham, M. H., Doherty, R. M., and Taft, R. W., *J. Am. Chem. Soc.* **106**, 464–466 (1984)

[30] Taft, R. W., Abraham, M. H., Famini, G. R., Doherty, R. M., Abboud, J.-L. M., and Kamlet, M. J., *J. Pharm. Sci.* **74**, 807–814 (1985)

[31] El Tayer, N., Tsai, R.-S., Testa, B., Carrupt, P.-A., and Leo, A., *J. Pharm. Sci.* **80**, 590–598 (1991)

[32] Seiler, P., *Eur. J. Med. Chem.* **9**, 473–479 (1974)

[33] Komiya, I., Park, J. Y., Kamani, A., Ho, N. F. H., and Higuchi, W. I., *Int. J. Pharm.* **4**, 249–262 (1980)

[34] Dressman, J. B., Amidon, G. L., and Fleisher, D., *J. Pharm. Sci.* **74**, 588–589 (1985)

[35] Levin, V. A., *J. Med. Chem.* **23**, 682–684 (1980)

[36] Hilger, A. R., Conradi, R. A., and Burton, P. S., *Pharmaceut. Res.* **9**, 902–910 (1990)

[37] Artursson, P., *J. Pharm. Sci.* **79**, 476–482 (1990)

[38] Ho, N. F. H., Park, J. Y., Ni, P. F., and Higuchi, W. I., Advancing quantitative and mechanistic approaches in interfacing gastrointestinal drug absorption studies in animals and man. In: *Animal Models for Oral Drug Delivery in Man: In Situ and In Vivo Approaches.*, Crouthamel, W. G. and Sarapu, A. (Eds.). APhA/APS: Washington, D.C.; 27–106 (1983)

[39] Ho, N. F. H., Park, J. Y., Morozowich, W., and Higuchi, W. I., Physical model approach to the design of drugs with improved intestinal absorption. In: *Design of Biopharmaceutical Properties through Prodrugs and Analogs*. Roche, E. B. (Ed.). APhA/APS: Washington, D.C. 136–227 (1977)

[40] Conradi, R. A., Hilgers, A. R., Ho, N. F. H., and Burton, P. S., *Pharmaceut Res.* **8**, 1453–1460 (1991)

[41] Conradi, R. A., Hilgers, A. R., Ho, N. F. H., and Burton, P. S., *Pharmaceut Res.* **9**, 435–439 (1992)

[42] Hogben, C. A. M., Tocco, D. J., Brodie, B. B., and Schanker, L. S., *J. Pharmacol. Exp. Ther.* **125**, 275–282 (1959)

[43] Burton, P. S., Conradi, R. A., Hilgers, A. R., Ho, N. F. H., and Maggiora, L. M., *J. Controlled Release* **19**, 87–98 (1992)

[44] Stein, W. D., The molecular basis of diffusion across cell membranes. In: *The Movement of Molecules Across Cell Membranes*. Academic Press: New York; 65–215 (1967)

[45] Diamond, J. M., and Wright, E. M., *Proc. R. Soc. Lond. (Biol.)* **172**, 273–316 (1972)

[46] Roseman, M. A., *J. Mol. Biol.* **200**, 513–522 (1988)

[47] Roseman, M. A., *J. Mol. Biol.* **201**, 621–623 (1988)

[48] Conradi, R. A., Hilgers, A. R., Burton, P. S., and Hester, J., *J. Drug Targeting* **2**, 167–171 (1994)

[49] Young, R. C., Mitchell, R. C., Brown, T. H., Ganellin, 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)

[50] Kim, D.-C., Burton, P. S., and Borchardt, R. T., *Pharmaceut Res.* **10**, 1710–1714 (1993)

[51] Karls, M. S., Rush, B. D., Wilkinson, K. F., Vidmar, T. J., Burton, P. S. and Ruwart, M. J., *Pharmaceut Res.* **8**, 1477–1481 (1991)

[52] Conradi, R. A., Wilkinson, K. F., Rush, B. D., Hilgers, A. R., Ruwart, M. J., and Burton, P. S., *Pharmaceut Res.* **10**, 1790–1792 (1993)

[53] Pardridge, W. M., and Mietus, L. J., *J. Clin. Invest.* **64**, 145–154 (1979)

[54] 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., Use of partition coefficients as a model for brain penetration applied to the design of H_2-receptor histamine antagonists. In: *QSAR. Rational Approaches to the Design of Bioactive Compounds*. Silipo, C., and Vittoria, A. (Eds.). Elsevier Science: Amsterdam; 103–110 (1993)

[55] Abraham, M. H., Chadha, H. S., and Mitchell, R. C., *J. Pharm. Sci* **83**, 1257–1268 (1994)

[56] Chikhale, E. G., Ng, K.-Y., Burton, P. S., and Borchardt, R. T., *Pharmaceut Res.* **11**, 412–419 (1994)

[57] Jacobs, R. E., and White, S. H., *Biochemistry* **25**, 2605–2612 (1986)

[58] Higgins, C. F., *Cell* **79**, 393–395 (1994)

[59] Bernards, C. M., and Hill, H. F., *Anesthesiology* **77**, 750–756 (1992)

[60] Raub, T. J., Barsuhn, C. L., Williams, L. R., Decker, D. E., Sawada, G. A., and Ho, N. F. H., *J. Drug Targeting* **1**, 269–286 (1993)

[61] Sawada, G. A., Ho, N. F. H., Williams, L. R., Barsuhn, C. L., and Raub, T. J., *Pharm. Res.* **11**, 665–673 (1994)

[62] Kasten, A. J., Banks, W. A., Castellanos, P. F., Nissen, C., and Coy, D. H., *Pharmacol. Biochem. Behav.* **17**, 1187–1191 (1982)

[63] Raeissi, S., and Audus, K. L., *J. Pharm. Pharmacol.* **41**, 848–852 (1989)

[64] Gray, R. A., Van der Velde, D. G., Burke, C. J., Manning, M. C., Middaugh, C. R., and Borchardt, R. T., *Biochemistry* **33**, 1323–1331 (1994)

[65] Carrupt, P.-A., Testa, B., Bechalany, A., El Tayar, N., Descas, P., and Perrissoud, D., *J. Med. Chem.* **34**, 1272–1275 (1991)

[66] El Tayar, N., Mark, A. E., Vallat, P., Brunne, R. M., Testa, B., and van Gunsteren, W. F., *J. Med. Chem.* **36**, 3757–3764 (1993)

[67] Kaiser, E. T., and Kézdy, F. J., *Science* **223**, 249–255 (1984)

[68] Sargent, D. F., and Schwyzer, R., *Proc. Natl. Acad. Sci. USA* **83**, 5774–5778 (1986)

[69] Lauterbach, F., Intestinal secretion of organic ions and drugs. In: *Intestinal Permeation*, Kramer, M., and Lauterbach, F. (Eds.). Excerpta Medica: Amsterdam; 173–195 (1977)

[70] Israili, Z. H., and Dayton, P. G., *Drug Metab. Rev.* **15**, 1123–1159 (1984)

[71] Gottesman, M. M., and Pastan, I., *Annu. Rev. Biochem.* **62**, 385–427 (1993)

[72] Peters, W. H. M., Boon, C. E. W., Roelofs, H. M. J., Wobbes, T., Nagengast, F. M., and Kremers, P. G., *Gastroenterology* **103**, 448–455 (1992)

[73] Tatsuta, T., Naito, M., Oh-hara, T., Sugawara, I., and Tsuruo, T., *J. Biol. Chem.* **267**, 20383–20391 (1992)

[74] Schinkel, A. H., Smit, J. J. M., van Tellingen, O., Beijnen, J. H., Wagennar, E., van Deemter, L., Mol, C. A. A. M., van der Valk, M. A., Robanus-Maandag, E. C., the Riele, H. P. J., Berns, A. J. M., and Borst, P., *Cell* **77**, 491–502 (1994)

[75] Burton, P. S., Conradi, R. A., Hilgers, A. R., and Ho, N. F. H., *Biochem. Biophys. Res. Comm.* **190**, 760–766 (1993)

[76] Ho, N. F. H., Burton, P. S., Conradi, R. A., and Barsuhn, C. L., *J. Pharm. Sci* **84**, 21–27 (1995)

[77] Hunter, J., Hirst, B. H., and Simmons, N. L., *Pharmaceut Res.* **10**, 743–749 (1993)

[78] Augustijns, P. F., Bradshaw, T. P., Gan, L.-S. L., Hendren, R. W., and Thakker, D. R., *Biochem. Biophys. Res. Commun.* **197**, 360–365 (1993)

[79] Wils, P., Phung-Ba, V., Warnery, A., Lechardeur, D., Raeissi, S., Hidalgo, I. J., and Scherman, D., *Biochem. Pharmacol.* **48**, 1528–1530 (1994)

[80] Chikhale, E. G., Burton, P. S., and Borchardt, R. T., *J. Pharmacol. Exp. Ther.* **273**, 298–303 (1995)

[81] Higgins, C. F., and Gottesman, M. M., *Trends Biochem. Sci.* **17**, 18–21 (1992)

15 Lipophilicity of Metabolites and Its Role in Biotransformation

Bernard Walther, Peter Vis and Albert Taylor

Abbreviations

CL_r renal clearance
CoA Coenzyme A

Symbols

$D_{7.4}$ Distribution coefficient (apparent octanol/water partition coefficient at pH 7.4)
P_{oct} Octanol/water partition coefficient
pK_a Negative logarithm of a dissociation constant K_a (acidic groups)

15.1 Introduction

Relationships between chemical or physico-chemical properties and metabolism [1] or disposition [2] have been the subject of a number of reviews. However, most of the attention has been focused on the parent compound and its metabolic routes while the effects of metabolic transformation, particularly on lipophilicity, have been very rarely addressed [3, 4].

The sequence of metabolism of drugs is classically described as a two-step process [5]: Phase 1, in which functional groups are added to the molecule by oxidation, reduction, or made available after hydrolysis, and phase 2, in which the functional groups can be conjugated to endogenous compounds such as amino acids, fatty acids, glucuronic acid, sulfate, glutathion, acetyl and methyl groups, etc. [5].

Traditionally, it has been considered that the effect of metabolism via both phase 1 and/or phase 2 reactions is to increase the water solubility of the xenobiotics in order to facilitate their elimination. However, a number of exceptions have been published over the past decade. In these instances, common metabolic reactions lead to an unexpected lipophilicity and even an increase in lipophilicity of the metabolic product.

These metabolic reactions can be divided into two types, those which introduce a lipophilic group, thus increasing the lipophilicity of the product [6], and those which introduce a polar group, changing the physico-chemical properties of the molecule and leading to unexpected higher lipophilicities.

This chapter reviews these metabolic reactions and discusses the possible role of changes in lipophilicity in biotransformation and their pharmacokinetic and pharmacodynamic consequences.

15.2 Introduction of a Lipophilic Group into a Drug

The metabolic processes leading to the formation of highly lipophilic metabolites have been well studied and documented in the past decade [6]. Because these metabolites are formed after conjugation with lipophilic endogenous components, the metabolic products are expected to be more lipophilic than the parent drug. These lipophilic conjugates are obtained after esterification of drugs with fatty acids or acylation with cholesterol derivatives.

The most common example for the esterification with fatty acids is the fatty ester of 7-hydroxytetrahydrocannabinol (Fig. 1) [7]. Further examples have been reported for the nonsteroidal antiinflammatory drug, etofenamate (Fig. 2), in dogs [8], and for antiarrhythmic drugs such as dipyridamole and mopidamol (Fig. 3) in rats and in humans [9]. The fatty acids involved are saturated and unsaturated fatty acids including oleic, palmitic, linoleic, stearic, palmitoleic, myristic, and lauric acids. This acylation process occurs between drugs containing a hydroxyl group and high-energy fatty acyl thioesters (acyl CoA thioesters).

Formation of cholesterol esters can either be obtained via a similar mechanism involving acyl CoA thioesters of cholesterol reacting with a drug, as documented with prednimustine (Fig. 4) [10], or are formed by acylation of cholesterol by a drug containing a carboxylic acid group as in the case of the hypolipidemic drug, BRL 24139 (Fig. 5) [11].

In an similar manner, incorporation of carboxylic acid containing drugs into triglycerides have been identified for 2-arylpropionic acid antiinflammatory drugs like ibuprofen, ketoprofen, and fenoprofen [12].
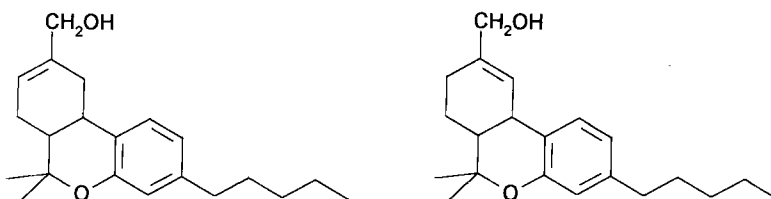
Figure 1. 7-Hydroxytetrahydrocannabinols.
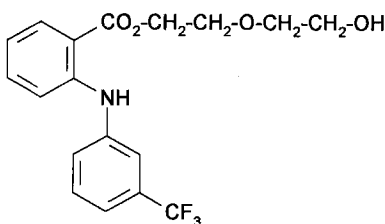
Figure 2. Etofenamate.

(a)

(b)
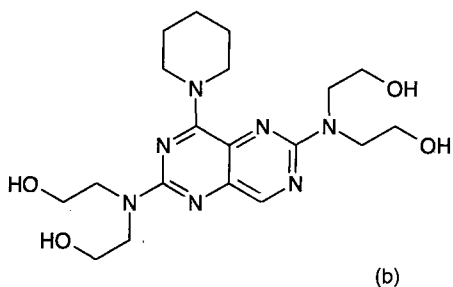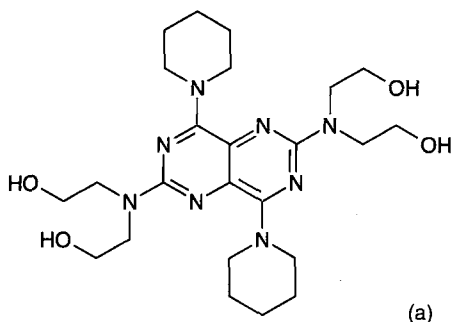
Figure 3. (a) Dipyridamole;
(b) mopidamol.
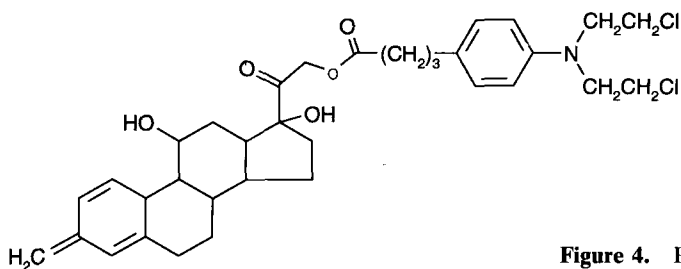
Figure 4.   Prednimustine.
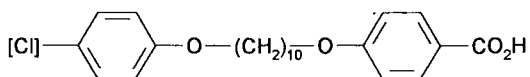
Figure 5.   BRL 24139.

In addition, certain xenobiotic acids can also be involved in elongation reactions of fatty acid biosynthesis which results in an increase of lipophilicity due to the addition of two carbon units. This type of reaction has been reported for furfural (Fig. 6) [13] and benzoic acid (Fig. 7) [14].
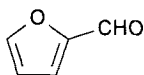
Figure 6. Furfural.
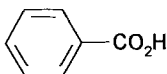
Figure 7. Benzoic acid.

15.3 Introduction of a Polar Group into a Drug

Formation of a more polar metabolite after oxidation or reduction and/or conjugation to an endogenous component is the more classical situation of drug metabolism with which chemists are confronted. However, several examples have been reported where the introduction of such polar groups has not resulted in the expected increase in hydrophilicity.

Of particular interest are certain hydroxylation and N-oxidation products (phase 1), glucuronide and sulfate conjugates (phase 2), normally regarded as hydrophilic substances, which, as a consequence of the changes to their physico-chemical properties, are more lipophilic than expected.

15.3.1 Increase of Lipophilicity Following a Phase 1 Reaction

First-phase reactions are known to cause only minor changes in structure with little and in most cases easily quantifiable changes in lipophilicity. Most of the unexpected lipophilicities reported as consequences of these reactions occur when the structural modification takes place near the site of ionization, thus altering the degree of ionization. Such a shift in pK_a leads to metabolites which are more lipophilic than expected at physiological pH.

One example has been described after N-oxidation of the basic drug tiaramide [3] (see Fig. 8). Tiaramide has a pK_a of 6.6 and a log $D_{7.4}$ of 0.87, the basicity of the compound depending on the piperazine ring nitrogen carrying the hydroxyethyl substituent. This pK_a is decreased by N-oxidation with the corresponding metabolite being less ionized and hence more lipophilic (log $D_{7.4} = 0.7$) than expected at physiological pH. In the case of tiaramide the N-oxide was still less lipophilic than the parent but with more basic drugs this type of shift in pK_a would result in a higher lipophilicity for the metabolite than for the parent compound. A further example has shown an increase of lipophilicity after oxidative metabolism. A cholinesterase inhibitor SM-10888 (9-amino-8-fluoro-1, 2, 3, 4-tetrahydro-2-4-methanoacridine; Fig. 9) is metabolized in rats after a first oxidation at the C1 position and a subsequent oxidation to a ketone [15]. The latter metabolite is more lipophilic at physiological pH (log $P_{oct} = 2.66$) than the secondary alcohol formed after the first oxidative reaction (log $P_{oct} = 1.59$) and the unchanged drug (log $P_{oct} = 2.23$). This has been explained by the carbonyl group withdrawing electrons from the aromatic ring more potently than the hydroxyl group, with a direct effect on the pK_a of the nitrogen. This is probably compounded by the forma-
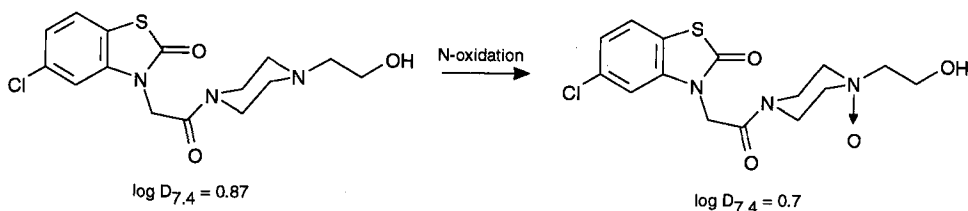
Figure 8. *N*-Oxidation pathway of tiaramide.
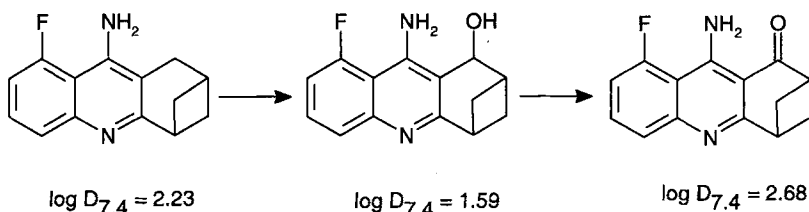
Figure 9. Phase 1 metabolism of SM-10888 in the rat.

tion of a hydrogen bond between the carbonyl and the amino groups forming a reso-
nant six-membered ring further contributing to a change in pK_a. The more lipophilic
metabolite showed a different pharmacokinetic behavior with higher protein binding
and lower renal clearance (CL$_r$) than the unchanged drug.

15.3.2 Increase of Lipophilicity Following a Phase 2 Reaction

In comparison with phase 1 reactions, conjugation reactions result in major changes in
structure and physico-chemical properties. It has always been assumed that these com-
pounds are markedly less lipophilic then the unchanged drug, and are readily elimi-
nated in the urine and/or faeces via biliary excretion.

A number of cases have been reported in which phase 2 metabolism, for two major
types of metabolic reaction – formation of sulfate and glucuronide conjugates – has
either not reduced or even increased the lipophilicity of the metabolic product.

For strong bases, conjugation to sulfate occurs via formation of zwitterionic com-
plexes. In the case of propranolol (Fig. 10) [4] (log $D_{7.4}$ = 1.18), sulfation on the side-
chain hydroxyl group has been found to have very little effect on lipophilicity of the
propranolol-*O*-sulfate (log $D_{7.4}$ = 1.12) at physiological pH. In contrast, the pro-
pranol 4'-*O*-sulfate conjugate was much more hydrophilic (log $D_{7.4}$ = –1.79) than the
propranolol 4'-OH (log $D_{7.4}$ = 0.49). The log D–pH profile for all four components is
shown in Fig. 11. These differences in lipophilicity after conjugation with a sulfate have
been ascribed to the proximity of the opposite charges (the anionic ionized sulfate and
cationic amino functions) separated by only two carbons in the case of propranol-*O*-
sulfate, with a consequent change in the solvation properties.

propranolol

log D$_{7.4}$ = 1.18

propranolol-O-sulfate

log D$_{7.4}$ = 1.12 (Δlog D = − 0.06)

4'-hydroxy propranolol

log D$_{7.4}$ = 0.49

propranolol-4'-O-sulfate conjugate

log D$_{7.4}$ = − 1.77 (Δlog D = − 1.28)
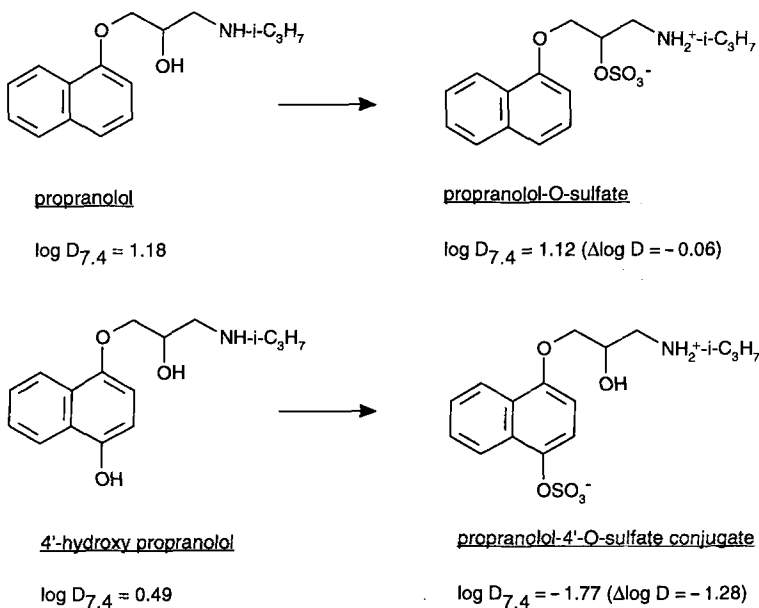
Figure 10. Metabolism of propranolol: changes in lipophilicity after formation of zwitterionic complexes.
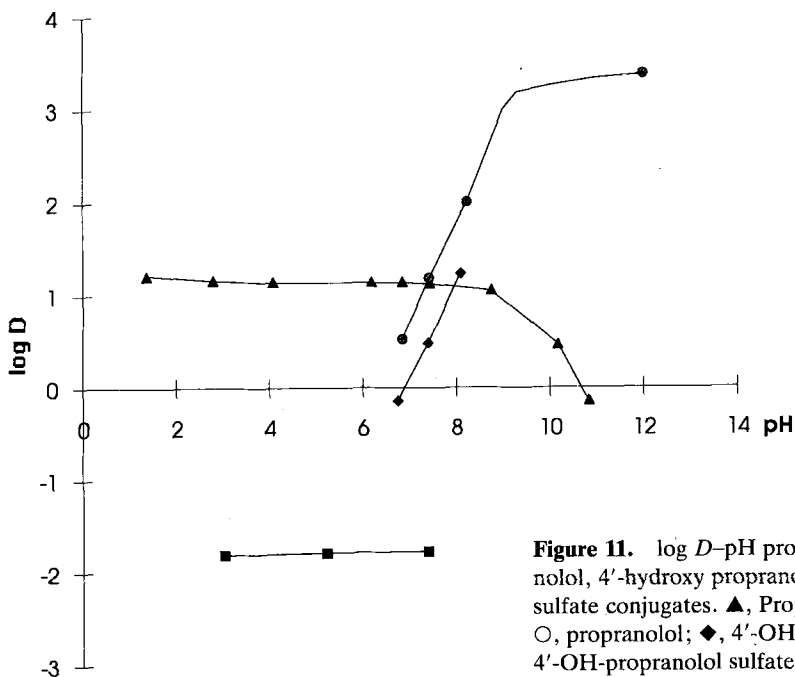
Figure 11. log D–pH profile for propranolol, 4'-hydroxy propranolol and their sulfate conjugates. ▲, Propranolol sulfate; ○, propranolol; ◆, 4'-OH-propranolol; ■, 4'-OH-propranolol sulfate.

morphine-3-
glucuronide

morphine-6-
glucuronide
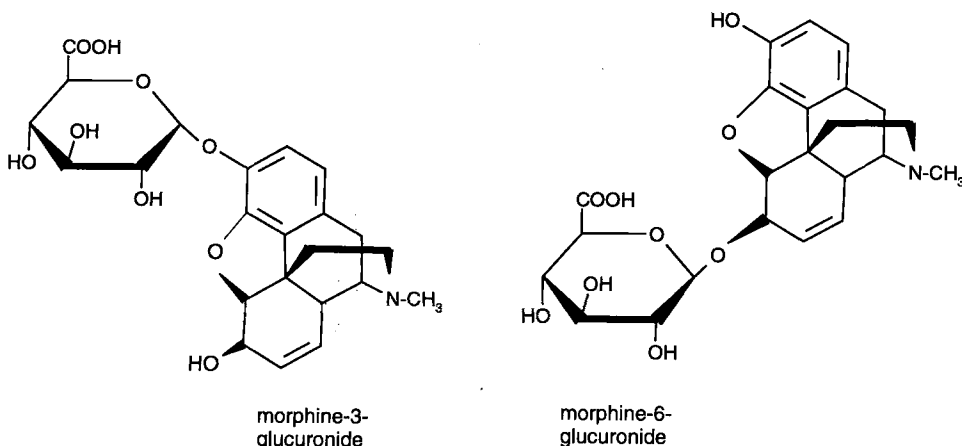
Figure 12.   Morphine glucuronides.

A lipophilicity of morphine glucuronides larger than expected has been documented [16]. Morphine undergoes glucuronidation of the 3-OH phenolic group and the 6-OH alcoholic group (Fig. 12). At physiological pH, morphine 6-glucuronide – and to a lesser extent morphine 3-glucuronide – has been found to be far more lipophilic than predicted. Conformational studies have shown that the two glucuronides exist as extended (and highly hydrophilic) conformers in water and as folded (and more lipophilic) forms in media of low polarity such as biological membranes. This is an important example that provides physico-chemical explanations for the passive crossing of biological membranes – in this case the blood–brain barrier – and, more importantly, shows that glucuronide conjugation can be an activation step, as opposed to a termination, of the biological activity of a drug.

15.4 Pharmacokinetic and Pharmacodynamic Consequences

In drug development there is an increasing need for a better understanding of the pharmacological and toxicological role of metabolites.

It is possible for many phase 1 metabolites to retain a certain degree of pharmacological activity, because these reactions result only in minor changes in structure and lipophilicity. In contrast, for phase 2 reactions, where relatively large changes of structure occur, it has been assumed traditionally that these molecules are directly eliminated and are totally devoid of activity. This is no longer a valid generalization: for instance, analgesic activity of morphine 6-glucuronide has been demonstrated after morphine administration [16]. Moreover, this example also illustrates a need to identify more clearly the factors that influence the role and fate of, particularly, those metabolites with more than expected lipophilicity. This is also important in the light of another glucuronidation process, namely the formation of acyl glucuronide conjugates which

have been associated with irreversible protein binding followed by potential immuno-toxicity [17].

In the past decade most attention has been devoted to the correlation of lipophilicity and pharmacokinetics [2]. The main concern with lipophilic metabolites is that the metabolite may show an extended elimination half-life caused by elimination rate-limited kinetics and an increased volume of distribution. This can result in chronic accumulation of a metabolite in the body, with its subsequent redistribution [18].

Retention in the body of lipophilic metabolites and, in fact, any lipophilic compound, is mostly attributed to the so-called *deep compartment*, from which redistribution into the plasma proceeds slowly, with a possible prolongation of action [18]. This is particularly true with the use of fatty acid esters as long-acting prodrugs of psychotropic agents and steroids [19]. In toxicological terms, fatty acid or cholesterol esters of xenobiotics may provide long-lasting depots with a slow release of subtoxic levels of the metabolites and possible consequences on endogenous processes [9].

However, these simplified views have little predictive value and do not always hold true. *In vivo*, for instance, uptake into adipose tissue of compounds appears to correlate with lipophilicity for acidic but not for basic drugs [18]. This lack of correlation is caused by other factors such as plasma protein binding, tissue protein binding, and perfusion limitation of the adipose tissue. Clearly, the lipophilicity of a drug only enables us to assess the affinity to adipose tissue without any other factors present.

Pharmacokinetic models have been applied to very lipophilic compounds, mainly in the environmental field [20, 21]. This type of model is particularly useful in extrapolation from animals to humans, for which a robust model is required. The pharmacokinetic model does not make use of a set of hypothetical rate constants, but uses transport and uptake processes. The models can vary in complexity, taking blood flow, organ size, tissue binding, metabolic rates, active and passive transport into account. These parameters can be determined separately and the model can be based on this information. In this way, it can be understood in a fundamental way why a lipophilic metabolite shows unexpected activity or pharmacokinetic behavior.

In order to understand more fully why metabolites are pharmacologically or toxicologically active one has to look, whenever possible, at the pharmacodynamic consequences of the presence of these lipophilic metabolites. Plasma levels do not always reflect concentrations in the target organ; however, correlations between plasma concentrations and pharmacological/toxicological data will allow us to assess the concentration of these compounds at the site of action more accurately.

The approach of combining physiological pharmacokinetics with pharmacodynamics will bring more understanding of the processes involved in the behavior of lipophilic metabolites.

References

[1] Testa, B., and Jenner, B., A structural approach to selectivity in drug metabolism and disposition In: *Concepts in Drug Metabolism, Vol. A*. Jenner, P., and Testa, B. (Eds.). Marcel Dekker: New York; 53–176 (1980)

[2] Seydel, J. K., and Schaper, K. J., *Pharmacol. Ther.* **15**, 131–182 (1981)

[3] Manners, C. N., Payling, D. W., and Smith, D. A., *Xenobiotica* **18**, 331–350 (1988)

[4] Manners, C. N., Payling, D. W., and Smith, D. A., *Xenobiotica* **19**, 1387–1397 (1989)

[5] Gibson, G., and Skett, P., *Introduction to Drug Metabolism*. Chapman & Hall: London, 1994

[6] Caldwell, J., and Parkash, M. K., Lipid conjugates of drugs and other xenobiotics: formation and possible significance. In: *Perspectives in Medicinal Chemistry*. Testa, B., Kybutz, E., Fuhrer, W., and Giger, R. (Eds.). VCH: Weinheim; 595–609 (1993)

[7] Leighty, E. G., Fentiman, A. F., Jr., and Foltz, R. L., *Res. Commun. Chem. Pathol. Pharmacol.* **14**, 13–28 (1976)

[8] Dell, H. D., Fiedler, J., Kamp, R., Gau, W., Kurz, J., Weber, B., and Wuensche, C., *Drug Metab. Dispos.* **10**, 55–60 (1982)

[9] Caldwell, J., and Marsh, M. V., *Biochem. Pharmacol.* **32**, 1667–1672 (1983)

[10] Gunnarson, P. O., Johansson, S. A., and Svensson, L., *Xenobiotica* **24**, 569–574 (1984)

[11] Fears, R., Baggaley, K. H., Walker, P., and Hindley, R. M., *Xenobiotica* **12**, 427–433 (1982)

[12] Fears, R., Baggaley, K. H., Alexander, R., Morgan, B., and Hindlex, R. M., *J. Lipid Res.* **19**, 3–11 (1978)

[13] Flek, J., and Šedivec, V., *Int. Arch. Occup. Environ. Health* **41**, 159–168 (1978)

[14] Caldwell, J., and Nutlea, B. P., *Br. J. Cancer* **52**, 457 (1985)

[15] Yabuki, M., Mine, T., Iba, K., Nakatsuka, I. and Yoshitake, A., *Drug Metab. Dispos.* **22**, 294–297 (1994)

[16] Carrupt, P. A., Testa, B., Bechalany, A., Tayer, N. E., Descas, P., and Perissoud, D., *J. Med. Chem.* **34**, 1272–1275 (1991)

[17] Spahn-Langguth, H., and Benet, L. Z., *Drug Metab. Rev.* **24**, 5–48 (1992)

[18] Bickel, M. H., *Prog. Drug Res.* **24**, 273–303 (1984)

[19] Sinkula, A. A., and Yalkowsky, S. H., *J. Pharm. Sci.* **64**, 181–210 (1975)

[20] Matthews, H. B., Tuey, D. B., and Anderson, M. W., *Environ. Health Perspect.* **20**, 257–262 (1977)

[21] Gerlowski, L. E., and Jain, R. K., *J. Pharm. Sci.* **72** 1103–1127 (1983)

16 The Role of Lipophilicity in Biological Response to Drugs and Endogenous Ligands

Vladimir Pliška

Abbreviations

B	"Bulk" compartment (water environment around a cell surface)
DADLE	(D-Ala^2,D-Leu$^{5)}$enkephalin
DNA	Deoxyribonucleic acid
GABA	γ-Aminobutyric acid
HS plot	Relationship between $\Delta S°$ and $\Delta H°$
L	Ligand
LR, LR$'$	Ligand-receptor complexes (Eq. (17))
R	Receptor
R*	Activated receptor (Eq. (17))
S	Membrane surface compartment
SRC	Stimulus-response coupling
TMA	Trimethylammonium

Symbols

a_0	Constant in the function $\varphi(p)$ (Eqs. (24, 25))
a_i	Pre-exponential constants (Eq. (2))
c_b	Concentration of a bioactive substance in the bulk compartment (Eq. (3))
c_s	Concentration of a bioactive substance in the vicinity of the membrane surface (Eq. (3))
c_r	Substance concentration in the receptor compartment (Eq. (1))
C_b, C_s	Equilibrium values of c_b, c_s
C_m	Equilibrium substance concentration in the membrane (Eq. (9))
C_ϑ	Affinity constant in the response function (Eq. (13))
D_2	Dose eliciting the half-maximal response, $E_{max}/2$ (Eq. (15))
e	Cellular response, *e(t)* denotes its time-course (Eq. (1))
$\exp(x)$	Exponential function, $\exp(x) \equiv e^x$
E	Response of a biological system (Eq. (13))
E_{max}	Maximal (asymptotic) response of a biological system (Eq. (13))
EC_{50}	Empiric D$_2$ value
F	Faraday constant
$G°$	Standard Gibbs free energy (Eq. (6))
G_{trs}	Gibbs free energy for transport of a molecule to the Helmholtz surface
h_0, h_1	Constants in linear equation describing entropy-enthalpy compensation (Eq. (34))

H°	Standard enthalpy
k_{tr}	Rate constant of transport between two compartments
K_p	Affinity constant for the access of a substance to the receptor environr (Eq. (23))
K_x, K_y	Dissociation constants of ligand-receptor complexes LR, LR' defined by (17)
K_Θ	Equilibrium constant for ligand-receptor interaction causing a membrane turbation (Eq. (28))
m	Substance mass transported between two compartments (Eq. (3))
n	Number of C-atoms in an alkyl group
p	Substance lipophilicity defined by Eq. (22)
$p0$, p_{max}	Constants in the function $\varphi(p)$ (Eq. (23))
$<p>$	Set of parameters in the response function Ψ
pA_2	Negative decadic logarithm of A_2
pD_2	Negative decadic logarith of D_2
$P_{m/s}$	Partition coefficient for partitioning between membrane interior and H holz surface
r_0	Total receptor concentration in a biological system (Eq. (14))
r_D	Shielded (Debye) distance (Eq. (5))
R	Gas constant
S°	Standard entropy
V_s	Fictive volume of the compartment S (membrane vicinity) (Eq. (3))
x	Perpendicular distance from a membrane
z	Number of net charges per molecule
α	Intrinsic activity of a bioactive substance
α_{max}	Maximal intrinsic activity in a group of bioactive substances (Eq. (26))
Δ	Symbol denoting a change of a thermodynamic function
Φ_s	Potential of the Helmholtz surface (Eq. (5))
Φ_x	Gouy-Chapman potential in the distance x from the Helmholtz surface ((5))
γ	Accumulation factor (Eq. (4))
Γ	Surface access of a substance (Eq. (10))
Γ_{max}	Maximal surface access (Eq. (10))
$\varphi(p)$	Function describing effects of lipophilicity p on $\varkappa$ (Eq. (21))
$\vartheta(C)$	Function of the dose C in the response function (Eq. (13))
$\varkappa$	Ratio of the dissociation constants K_x/K_y (Eq. (20))
λ_i	Exponential constants (Eq. (2))
μ	Empirical power coefficient in Eq. (28)
ν	Empirical power coefficient in Eqs. (24–26)
π	Hansch hydrophobic constant
$\Theta(C)$	Function describing a membrane perturbation (Eq. (28))
ω	Empirical power coefficient in Eq. (29)
ξ	Response potential (Eq. (16))
Ψ	Response function (Eq. (1))

16.1 Introductory Comments and Definitions

The subject of this chapter is the final step of interaction between a chemical substance and a biological system – a living organism or its isolated integral part. Such a system is, by definition, a set of elements and dynamic connections between them which exercises basic life functions: metabolism and growth, self-reproduction with optional mutations of genetic information, and self-regulation of internal processes including morphogenesis. The substance entering into the interaction with a biological system will be assigned here as *bioactive substance* (or, in relation to its interaction with a receptor, as a *ligand*); it is basically immaterial whether it is a natural agent like a hormone, neurotransmitter, internal ligand, etc., or a xenobiotic (drug, toxic agent, etc.).

A biological system represents a black box. Its structure and its internal functions are known only insufficiently, and its investigation proceeds solely via input/output analysis. Moreover, an equilibrium state in which all variable parameters reach invariant levels, does not exist in biological systems: it would assume an infinitely long constancy of internal and environmental conditions which, *sensu stricto*, never occurs. A steady state represents a temporary equilibrium when time changes of all variables are equal to zero (usually in the time point of inflection, when directions of their changes are reversed). Also this state is uncommon and mostly restricted only to a subset of variables. Very frequently, a biological system is investigated in the transient state, after a change of external or internal conditions assigned here as *stimulus*. In general, time derivatives of individual variables in such a state are nonzero, in contrast to the other two states mentioned above. Mostly, however, the state of the system cannot be determined. Thus, a firm basis for a response analysis is still lacking and any structure-
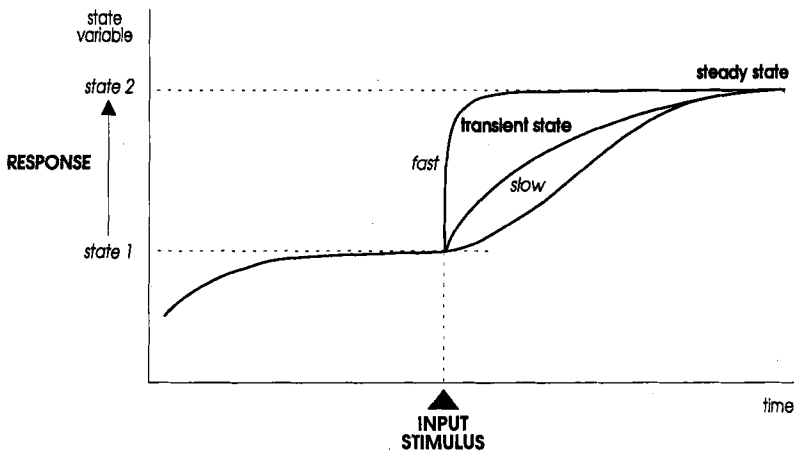
Figure 1. Biological response as a state change of a biological system. An input stimulus (e.g., addition of a bioactive substance, but also a rapid change of physical conditions, change of composition of the surrounding medium, etc.) causes a transfer from one state to another. The difference between two levels of the measured state variable (e.g., muscle tension, ionic flux, rate of hormone secretion, etc.) is assigned as biological response.

activity study is, at best, a semiempiric one. Now as before, the "structure" of a molecule pertinent to the structure-activity relationship studyseems to be far more precisely characterized than the other end of the chain, the biological "activity". This has some epistemological consequences, in particular, with regard to the reliability of conclusions, and to the test procedures employed in this area.

The state of all system variables is assigned as the *state of the system*. The state change of a biological system reflects change of variable parameters like chemical composition, rates of processes, energy content, temperature, steric structure, etc. A state change caused by an interacting substance is conventionally assigned as *"biological response"*; it is usually assessed as the difference of a single parameter in the final and initial state. This notion is visualized in Fig. 1.

16.2 Phases of a Biological Response

Despite the unrevealed structure of a biological system responding to a bioactive substance,certain phases of its response can be determined [1, 2] (Fig. 2). In general, substance transporting and responding subsystems can be delimited in any such system. The stimulating substance enters the system at various gates, in *in-vivo* conditions usually via blood circulation. In the intravascular space, the substance exists in a free form and in forms bound to plasma proteins. These forms passes by diffusion through fenestration of the vessel walls (passive transport), most of the forms bound to plasma proteins remain in the circulating fluid. Rate of the free passage depends upon physicochemical features of the substance; lipophilicity is frequently rate determining. The same applies for transport and partitioning between compartments within the interstitial space (extracellular fluid). For purposes of pharmacological analysis, it appears useful to differentiate between compartments within the *nontarget* tissue, i.e., the tissue without responding cells to the substance in question, and the *target* tissue. The interaction with the receptor, assumed to follow general rules of chemical kinetics, is under the existing conditions controlled by the substance concentration in its close vicinity. This must be particularly emphasized since concentration gradients may exist even very close to the target cell membrane, owing to the hydrodynamic forces within the capillary structure of the intercellular space. The *receptor compartment* [3, 4] is therefore defined only operationally, as a distribution space whose substance concentration determines the rate of ligand-receptor interaction. Potential receptor internalization, causing a nonconstancy of the number ("concentration") of receptors, does not change the rate-determining role of the bioactive substance. Kinetics of the internalization process itself depends upon the bioactive substance concentration in the receptor compartment.

Ligand-receptor interaction initiates a number of cellular events that are virtually autonomous and largely independent of the substance properties. The substance molecule, as an information transmitter, finishes its role after entering into interaction with its receptor. *Biological response* is then reflected as a state change of a specified cellular subsystem, e.g., rearrangement of actin/myosin in muscle contraction, closing/opening of ionic channels, or enhanced gene transcription rate on a DNA sequence, etc.

ENTRY OF
STIMULATING AGENT

```
  ┌──────────────┐                    binding      ┌──────────┐
  │ inactive form of │◄──────  ┌──────────────┐  ─────────►│ plasma   │
  │  stimulating    │          │ blood plasma │◄──────────│ proteins │
  │  substance      │ inactivation └──────────────┘          └──────────┘
  └──────────────┘
```

passive
transport

NON-TARGET
TISSUE

TARGET
TISSUE

```
  ┌──────────────┐        ┌──────────────┐        ┌──────────────┐        ┌──────────────┐
  │ inactive form of │◄───│ extracellular │◄──────►│  receptor    │──────►│ inactive form of │
  │  stimulating    │    │ compartments  │        │ compartment  │       │  stimulating    │
  │  substance      │ inactivation └──────────────┘   inactivation │  substance      │
  └──────────────┘                                    └──────────────┘
```

elimination
excretion

binding

**TRANSPORTING
SUBSYSTEM**

**RESPONDING
SUBSYSTEM**

```
                                ┌──────────────┐
                                │  receptors   │
                                └──────────────┘
```

STIMULUS-RESPONSE COUPLING

```
  ┌──────────────┐        ┌──────────────┐
  │ effector in  │───────►│ effector in  │    ➡ RESPONSE
  │ inactive state│◄──────│ active state │
  └──────────────┘        └──────────────┘
```

Figure 2. Diagram of a responding system. The figure shows individual compartments of a biological system responding to a stimulating agent. Several entries of this agent can be considered: blood plasma or extracellular compartments. The compartment in close contact to cellular responding sites (receptors, etc.) is called the receptor compartment. Processes of stimulus-response coupling (cell signaling) are largely autonomous (independent of the substance concentration in the receptor compartment). Stimulus-response coupling leads to a state change of a cellular unit, the effector. This change is measured as biological response.

16.3 Stimulus-Response Profiles

16.3.1 Characteristic Types of Response Profiles

In the focus of interest are relationships between physico-chemical properties of the bioactive substance and its effects on the investigated biological system. These relationships reflect transient or equilibrium behavior of the system after stimulation. System can be investigated by means of analysis of their transfer functions, the dynamic input/output relationships. Such an approach has been used many times in biology. However, it is less powerful for a detailed analysis of biological response systems, and mostly yields solely formal analytical solutions, without clearly interpretable parameters. Therefore, biological responses are mostly assessed in form of arbitrary defined descriptors. This makes the analysis uncertain and strongly dependent on stimulation and other conditions.

Stimulation of a system proceeds in several ways. Decisive for the input/output analysis is the profile of the applied stimulus. The analysis of physical or technical systems
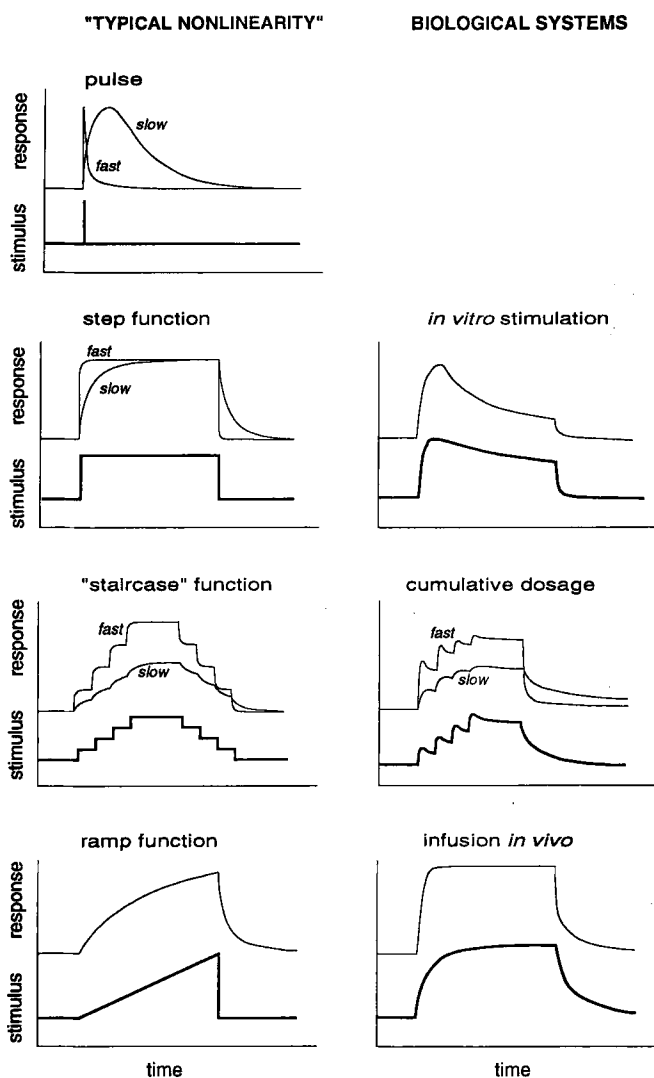
Figure 3. Possible forms of biological response. The input stimulus usually follows a discontinuous function called **"typical nonlinearity"**: a pulse, a step function or a ramp function. The expression **"nonlinearity"** concerns the discontinuity in the time point of stimulation. Left-hand panels: typical nonlinearities in a physical system; right-hand panels: their possible equivalents in a biological system (*in-vitro* or *in-vivo* responses).

can usually employ stimuli which have a form of a "typical" nonlinearity like the unit pulse, unit step and its superposition (called sometimes "staircase signals"), and the unit ramp. They are shown in the left-hand panels Fig. 3. Ideal profiles of this kind neither occur within a biological responding sytem *in vivo*, nor can they be perfectly simulated in conditions of an experiment. A stimulus deformation depicted in the

righthand panels of Fig. 3 usually occurs already during the initial phase, since the stimulating substance undergoes partial inactivation, elimination, andtransport within the system. Thus, in an *in-vitro* experiment with an isolated tissue, the substance concentration in the organ medium is not an ideal step function ("square wave") but rather an exponential; the same is valid for *in-vivo* infusion experiments or analogous *in-vitro* experiments where the stimulation should ideally follow a ramp function. All this complicates the application of the common systems analysis to biological systems responding to bioactive substances.

16.3.2 Time and Intensity Components of a Response

A question now arises as to how individual phases of action of the bioactive substance influence the final response, and which phase is the most relevant one. The response is composed by two vectors [3]: 1) response time, expressed on a scale relative to the moment of stimulation; and 2) time-dependent intensity. Obviously, the intensity is determined both by properties of the cell-signaling system (affinity of the substance to its receptor) and transport processes that control the concentration in the receptor compartment. Rates of transport processes and the dynamics of the stimulus-response coupling, on the other hand, are determining for the time component [4]. The autonomous nature of the stimulus-response coupling (SRC), i.e., the sequence of cellular processes between receptor and the final responding unit [5], may lead to an option known in pharmacology as the "hit-and-run" response: the bioactive substance has already been eliminated from the receptor compartment but the response lasts until SRC reaches the nonstimulated state. Effects of this kind may *not* display correlations with the physico-chemical properties of the bioactive substance. Although they are virtually not independent of them, the time-determining rate of SRC conceals their role in partitioning and cell signaling.

Rates of inactivation and elimination are, as a rule, considerably lower compared with that of the transport processes. This can be inferred from modeling of compartmental biological systems. The rate of inactivation is rate determining for both components of the response. The time component seems to be affected mainly by the inactivation rate in the receptor compartment, and less by that in the other compartments [4]. Therefore, pharmacokinetic data obtained for circulating plasma and other body fluids may be less relevant for conclusions about the time-course of a response itself.

16.4 Bioactive Substance in the Receptor Compartment: Response Function

16.4.1 General Formula of the Response Function

As mentioned, the intensity of a response is determined by the concentration of the bioactive substance in the vicinity of its cellular receptors, i.e., in the receptor

compartment. The concept was originally employed for a transient state response: it evolved that an *in-vivo* system could frequently be reduced to three compartments [4] and the substance can be transferred between any two of them. A superimposed exponential function then describes time changes of substance concentration within each compartment [4, 6]. The exponential and pre-exponential constants are functions of initial concentrations, capacities (volumes) of individual compartments, and rate constants of participating processes like transport, inactivation, etc. The time-course of cellular response (*e*), assigned here as the response function (*Ψ*), is then

$$e(\text{t}) \;=\; \Psi(c_r(t), <p>) \tag{1}$$

c_r is the substance concentration in the receptor compartment, $<p>$ a set of substance-dependent and substance-independent constant parameters specifying the response system (ligand-receptor interaction, stimulus-response coupling, etc.). The form of Ψ was discussed many times in the literature. Let us just mention in passing that it always has an empirical nature; most suitable functions are the power function of the Hill type [7–9], its simplified form – the rectangular hyperbola [10–12], and a rational fraction [6].

16.4.2 Transport and Partitioning

Time-course of the substance concentration in the receptor compartment, $c_r(t)$, is in common cases described by

$$c_r(t) \;=\; \sum_i a_i \exp(-\lambda t) \tag{2}$$

with pre-exponential constants a_i (proportional to the initial substance concentrations in individual compartments) and exponential constants λ_i; *i* relates to an individual exponential term, the summation is extended over all terms. In a simple case, the total number of terms equals to the number of compartments. A stable steady state, with a long time period of a quasi-constant nonzero concentration C_r, ban be reached only when at least one of the λ constants is zero or, at least, much smaller than the others.

Such circumstances exist only in closed systems, in which no chemical modification of the substance and no exchange of substances with their environment occurs. There is scarcely a biological system which would meet this condition; to a certain degree, it can be mimicked by a permanent infusion (*in-vivo*), by inhibition of substance inactivation, etc. Our further deductions are related to an ideal closed system in which the bioactive substance is transported between two compartments: the close vicinity of the membrane surface S (which may likely be the receptor compartment) and the "bulk" compartment B composed of the water environment of the membrane (subscripts s and b, respectively). The transport of the substance from B to S follows the rate equation

$$dm/dt \;=\; V_s k_{tr} \, (c_b - c_s/\gamma) \tag{3}$$

m is the transported mass, V_s the volume of the compartment S, k_{tr} is the rate constant (dimension time^{-1}), γ is a numeric factor that expresses the effect of c_b upon the transport rate. An active transport from B into S occurs when $\gamma > 1$, in the opposite direc-

tion when $\gamma < 1$; passive transport along the concentration gradient takes place when $\gamma = 1$. The concentrations c_b and c_s reach at $dm/dt = 0$ their equilibrium values C_s and C_b, respectively; then

$$\gamma = C_s/C_b \tag{4}$$

stands for a "partition coefficient", assigned in this case as *accumulation factor.*

16.4.3 Compartmentation in the Vicinity of a Membrane

Essential for compartmentation of a substance are conditions along the membrane. A bilayer biological membrane possesses a highly hydrophobic core and carries a negative gross charge on its outer surface (Fig. 4), indicating a fixed layer of negatively charged particles in the membrane-aqueous phase interphase. The distribution of the electric potential in the vicinity of a membrane is supposed to follow the Gouy-Chapman model of a diffuse double layer [13]. The Gouy-Chapman potential Φ in the distance x from the upper margin of the negatively charged layer (Helmholtz surface) follows as exponential

$$\Phi_x = \Phi_s \, exp(- x/r_D) \tag{5}$$

where r_D, called shielded (Debye) length, determines the steepness of the potential gradient, i.e., a deviation of the "pure" Coulomb value Φ_s of the Helmholtz surface. Using the Boltzmann distribution

$$C_s = C_b \, exp(- \Delta G_{trs}/RT) \tag{6}$$

for a free energy difference between molecules of the substance on or close to the membrane surface (concentration C_s) and in the distant layer of the aqueous phase (C_b), and assuming that the free-energy change ΔG_{trs} of this process is defined by
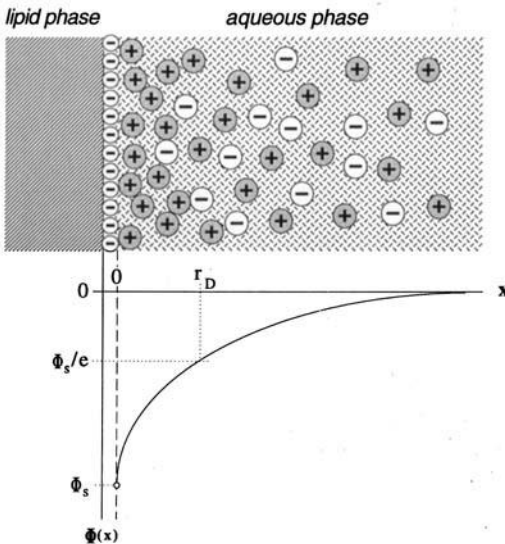
Figure 4. Guoy-Chapman potential Φ in a distance x from a positively charged biological membrane. Φ_s is the potential of the Helmholtz surface, r_D is the shielded length (cf. Eq. (5) and the corresponding text).

$$\Delta G_{trs} = - z \, F \, \Phi_s \tag{7}$$

(z is the number of net charges per molecule, F is the Faraday constant, R is the gas constant, T is the absolute temperature), one obtains for the accumulation factor (called here Boltzmann accumulation factor)

$$\gamma = \exp(z\Phi_s F/RT) \tag{8}$$

defining the ratio of the two concentrations.

This formula shows the effect of the diffuse double-layer potential on the relationship between the substance concentration in the distant aqueous medium and the nearest vicinity of a receptor [13]. This relationship is dependent on physico-chemical conditions of the system and can by no means be considered as invariable. The effect of, for instance, ionic strength of the aqueous environment may quite considerably affect the compartmentation, as it has already been demonstrated in the literature [14–17]. Under these circumstances, measurement and computation of ligand-receptor affinities becomes highly problematic: the "true" equilibrium constant can be obtained only by division of the assessed apparent dissociation constant by γ which is usually unknown [18].

16.4.4 Partitioning in the Aqueous/Lipid Interphase on Cell Surface

Biological membranes do not possess an ideal structure assumed by the Gouy-Chapman model. The inserted protein molecules may carry a significant positive charge and cause perturbations in the diffuse double layer. Therefore, the membrane surface does not display any particular selectivity for positively charged particles. Its hydrophilic nature and its ability to attract polar molecules is decisive for the partitioning described by Eqs. (4), (5) and (8). On the other hand, fixation of a molecule in the aqueous-lipid interphase of the membrane, and therefore the persistence of a substance on the cell surface (in the vicinity of receptors), is implemented by hydrophobic interactions between the bioactive substance and the lipid core. The hypothetical equilibrium concentration in the membrane (C_m) is determined by the partition coefficient

$$P_{m/s} = C_m/C_s. \tag{9}$$

Definitions of both C_m and C_s are, however, vague. They are related to the amount of the substance in the compartments, the volumes of which are not sharply delimited. This restriction holds true particularly for the membrane compartment, in which the maximal access for a solute is restricted to the thin layer adjacent to the membrane interphase. Rather, one should consider the surface access of the substance. In a monomolecular layer, the access Γ follows the Langmuir isotherm

$$\Gamma = \Gamma_{max} \frac{P_{m/s}C_s}{P_{m/s}C_s + 1} \tag{10}$$

where Γ_{max} is the saturability of the membrane interphase with the solute [13]. $P_{m/s}$ follows from

$$\Delta G_{trs} = -RT \ln P_{m/s} \tag{11}$$

ΔG_{trs} is the free energy of transport from the membrane to the aqueous receptor compartment. Cumulation of the substance in the membrane is then described as an apparent partitioning between the bulk and membrane; its partition coefficient $P_{m/b}$ is

$$P_{m/b} = \gamma P_{m/s} \tag{12}$$

It is, however, *a priori* not possible to determine which of the two domains S and M fulfill the operational definition of the receptor compartment. The problem is not irrelevant since it touches the question of optimal hydrophobicity of a ligand interacting with a particular binding domain of a membrane-bound receptor. From the viewpoint of statistical thermodynamics, ligands moving freely by diffusion in the membrane proximity can enter much more easily into the interaction with receptors than those fixed in the membrane. Moreover, highly hydrophobic molecules have less chance to pass the polar barrier represented by the Helmholtz surface; their transfer from the lipid to aqueous phase is associated with a high free-energy change. On the other hand, binding domains of membrane receptors can be potentially reached from the hydrophobic core. The experience also shows that high lipophilicity is not fully prohibitive to the fixation of the substance in the aqueous-lipid interphase, since some fraction of the molecules determined by the Boltzmann distribution law may reach the interphase despite of the existing barrier.

Another important detail should be emphasized. Most of the bioactive substances are amphiphilic and contain both hydrophobic and hydrophilic domains in their molecules. Peptides are particularly interesting from this point of view since their flexible backbone facilitates a fixation in the water-lipid interphase without further diffusion into the interior of the membrane. According to their secondary structure and to the distribution of polar and nonpolar sites in the molecule, they may attain any position varying between two extreme types, a parallel one and a perpendicular one, to the
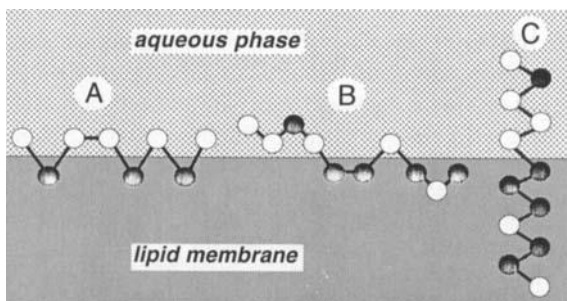
Figure 5. Behavior of amphiphilic molecules possessing high degree of structural flexibility (e.g., peptides) in the water-lipid interphase of a biological membrane (according to Schwyzer [13]). Closed and open circles denote hydrophobic and hydrophilic sites of the molecule, respectively. A and B, horizontal positioning; C, vertical positioning (perpendicular to the interphase). Note that the location in individual phases is determined by the overall lipophilicity of corresponding segments (a hydrophilic group may occur in the hydrophobic site of the interphase and vice versa).

interphase (also called *secondary* and *primary,* respectively) [13, 19]. They are shown schematically in Fig. 5. Thus, the membrane may participate on the formation of the molecular configuration, and the "membrane-induced conformation" may be more suitable, or selective for some particular receptors, than the conformation in polar solvents. Opioid peptides, as shown recently by Schwyzer and coworkers [13–16], are an interesting paradigm for these effects of biological membranes.

16.5 Ligand-Receptor Interaction

A collision between a ligand molecule and a target area of the receptor results in the formation of a relatively weak ("intermediary") complex which is supposed to follow regular stoichiometric rules (Dalton's law of multiple and constant proportions between reacting compounds). Furthermore, it is assumed that the ligand-receptor interaction is a diffusion-controlled process [20, 21]. Whereas the former rule could not have been disproved in any instance as yet, the latter assumption is obviously subjected to several restrictions. To begin with, diffusion rates of the two components show substantial differences. Whereas ligand molecules can be transported virtually by free
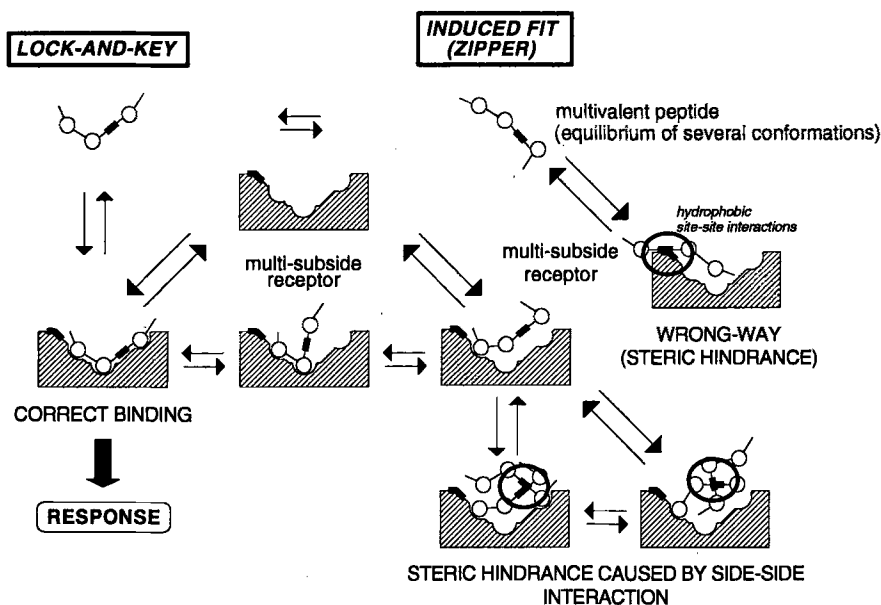
Figure 6. Lock-and-key and induced fit: two extreme models of a ligand-binding site interaction. The lock-and-key model assumes a full complementarity of ligand and binding site conformations and thus, a "perfect" fit. In the case of an induced fit, the ligand conformation matches in a stepwise process the one of the binding site (apparently, both conformations may be adapted in this process). Interactions (encircled) between hydrophobic molecular segments (bold sections) of both ligand and binding site, or its closest environment, may cause obstacles to the "correct" binding needed for the initiation of a biological response.
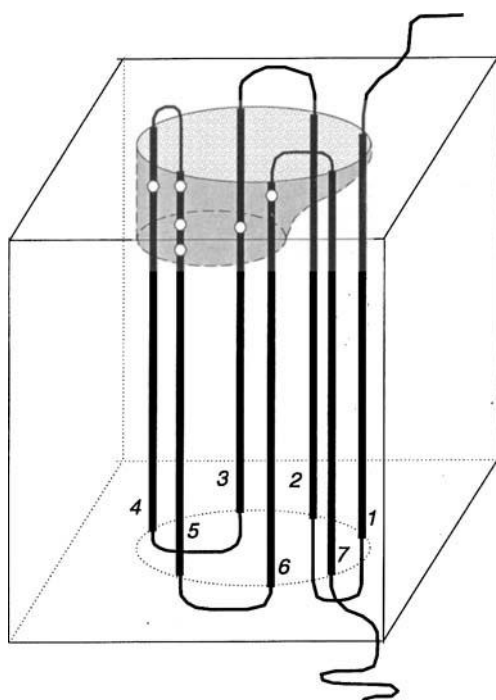
Figure 7. Scheme of the space organization of a heptahelical membrane receptor (G protein-coupled): β_2-adrenergic receptor. The diagram represents a hypothetical cubic section of a receptor-carrying cell membrane. Bold lines are transmembranal helices (numbers indicate their sequence in the primary structure of the receptor), open circles putative attachment sites for adrenaline. The shaded segment shows a space for ligand binding within the outer part of the membrane.

diffusion (their barriers in the receptor compartment are mentioned above), membrane-bound receptors are fixed in a rigid environment of the membrane which allows for only a slow diffusional movement. Furthermore, several attractive and/or repulsive forces, among them effects of hydrophobic interactions, may be operative when a ligand molecule approaches the receptor. And finally, a collision does not in each instance, and immediately, lead to the formation of a "productive" complex, i.e., one which initiates the cellular response. Structural complementarity of ligand molecules and the binding site region on receptors, the crucial feature of the lock-and-key model, occurs probably only very rarely [21]. Rather, the molecular segments of the ligand and the receptor that are involved in the formation of intermediary complexes undergo an *induced fit*. Virtually, both moieties adjust their conformation to that of the partner, although the ligand molecule is likely to accommodate more flexibly. This "zipper-like" adjustment of the ligand-receptor bonds is subjected to perturbances in which hydrophobic forces may play an important role (Fig. 6). A wrong-way binding owing to an interaction of two hydrophobic regions not involved in regular binding may result in a steric hindrance of the receptor, not allowing further correct-way interaction [22]. Also, they may cause a mutual interaction between such hydrophobic regions of two ligand molecules which also may prevent a further correct binding process. These phenomena may have at least two pharmacological consequences: irregularities of dose-response relationships at higher ligand concentrations ("downhill" phase), and the so-called *partial agonism* (see below).

A

B

helix   **5**   **4 3**   **6**
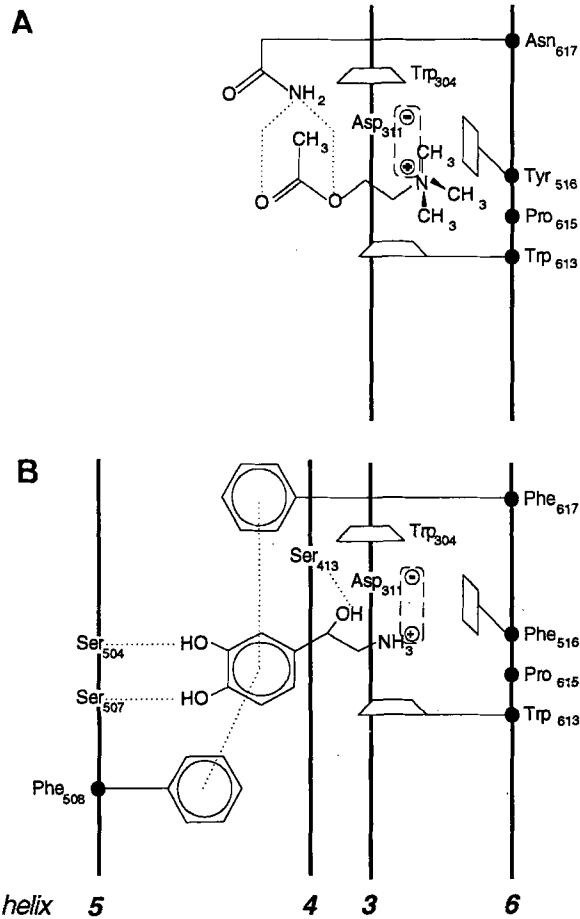
Figure 8. Binding domain maps of β_2-adrenergic (panel A) and cholinergic (nicotinic) receptor (panel B), respectively. Tyrosine and tryptophan residues depicted by rhombuses and trapezoids build the "hydrophobic pocket", surrounding the ligand bound to specific attachment sites. Their aromatic rings carry an induced negative charge. Weak bonds (π–π interactions, hydrogen bonds) are depicted by dotted lines, salt bridges by dotted ovals around the corresponding charges. (According to Hibert et al. [23].).

The microenvironment of the receptor may also exercise a considerable effect on the transport of the ligand to the binding site. This occurs when the binding site is plunged below the outer membrane interphase. X-ray studies identified such a conformation in several G-protein-coupled receptors [23]; their geometry is schematically depicted in Fig. 7. Fig. 8 shows the putative structures of the binding regions of two of them: the muscarinic (M_2) and the β_2-adrenergic receptors. Individual binding sites are stretched between α-helices 3, 4, 5, and 6; hydrogen bonds, aromatic, and ion-ion interactions participate on the binding which is furthermore stabilized by a hydrophobic pocket built up from the surrounding phenylalanine and tryptophan residues. Similar structures were identified also for other receptors like dopamine and serotonin [23]. Recent studies indicate that negative charges can be induced in these hydrophobic domains [24, 25]. Consquently, the stabilization effects are not solely due to hydrophobic interactions.

16.6 Factors Determining Biological Responses: a Summary

Considering these circumstances, one can reach several tentative conclusions and summarize them in the following way:

1. The receptor compartment may emerge as a rather fictive space since the distribution of the interacting ligand molecules is continuous in the perpendicular direction to the membrane, and very likely quite irregular along the membrane surface (the membrane surface is not homogenous with regard, at least, to the charge distribution and distribution of polar groups in the membrane-intruded proteins). So far, it cannot be proved experimentally which state of the ligand around its receptors is directly determining the rate of interaction and thus, the intensity of a response.

2. The diffusion is not the only process which controls the rate of the ligand-receptor interaction. This mode is probably restricted to the first interaction step; the rest is merely a matter of an induced fit, or of a zipper-like mechanism of closing the entire ligand-receptor bond. Recent studies indicate that even rather rigid ligand molecules (like catecholamines) are not bound by a single linkage to the receptor; one can infer that they must undergo a certain induced fit, similar to molecules with a flexible backbone (like peptides).

3. Dose-response curves reflect features of all phases of response to a ligand and therefore, cannot be employed for any conclusion concerning the ligand-receptor interaction. Also binding studies carried out in the conventional way with membrane fractions, whole cells, etc., may become problematic to this end since they are, more or less, influenced by the same complicating circumstances.

16.7 Partial Agonism and the Role of Lipophilicity

16.7.1 Dose-Response Relationship and the Phenomenon of "Partial Agonism"

The response E of a biological system to a stimulatingsubstance usually follows an asymptotic relationship

$$E = E_{max} \frac{\psi(C)}{\psi(C) + C_\psi} \tag{13}$$

where $\psi(C)$ is an increasing function of the "dose" C defined as the amount of the substance per volume or mass unit of the system, C_ψ is a constant, E_{max} is the ordinate of the (positive) asymptote. Obviously, E reaches a limit value E_{may} at very high doses (ψ $(C) \to \infty$). Properties of the function ψ are generally not known: they reflect kinetic and thermodynamic parameters of individual phases of response and are strongly dependent on the "active" substance concentration at the receptor. The simplest, most popular and in many cases pragmatic solution has been offered by Ariëns and coworkers [11], suitable in particular for *in-vitro* experiments with isolated organs, cells, etc.

According to their suggestion, E_{max} is a product of the total receptor "concentration" r_0 in the stimulated system (precisely: the receptor capacity of the system) and the *intrinsic activity* α of the substance [11], i.e., the ability to utilize the responding capacity of the system:

$$E_{max} = \alpha\, r_0. \tag{14}$$

When, in the most elementary case, $\psi(C) = C$, the relationships (13) becomes

$$E = E_{max}\frac{C}{D_2 + C} \tag{15}$$

where D_2 is the dose C eliciting the half-maximal response, $E_{max}/2$ (frequently assigned in the literature as EC_{50}, the "effective concentration" causing a 50 % of the maximal response). However simplified, this notion defines the maximal response as a property of both the system and the stimulating substance and hence, reflects the old experience that substances may differ in their ability to stimulate under the given conditions. The intrinsic activity α can be conventionally set within the interval $0 < \alpha < 1$, and the active substances ($\alpha > 0$) are assigned as *full* and *partial agonists* ($\alpha = 1$ and $0 < \alpha < 1$, respectively). *Competitive antagonists* are then substances which interact with the receptor, are bound to the same binding domain as full and partial agonists, but possess zero intrinsic activity, $\alpha = 0$. Schematic graphs of the corresponding dose-response curves are shown in Fig. 9. The biophysical background of the differences in α is still unclear; several explanations can be offered but scarcely proved in detail. Most models of partial agonism are based on a degreased efficiency of the stimulus-response coupling [5, 12, 26]. This, however, cannot explain the circumstances that lead to differing E_{max} values of individual substances, and the question arises as to whether an alternative explanation on the level of receptors could be presented.
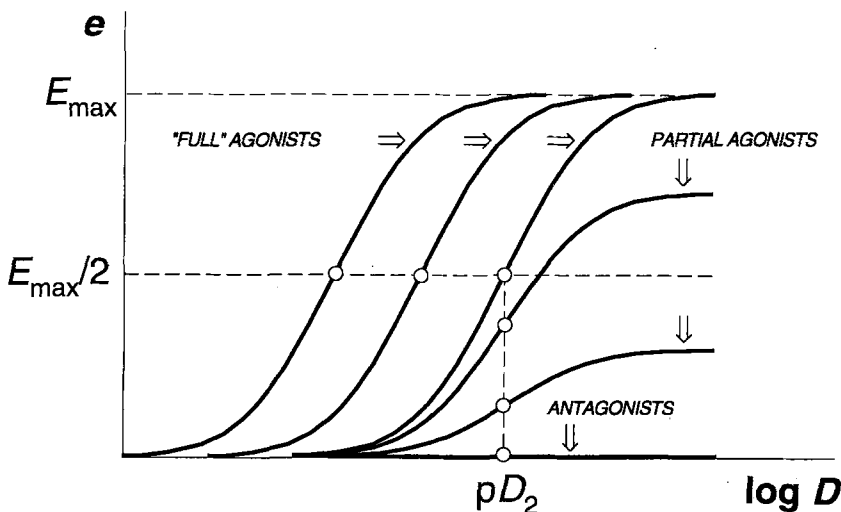
Figure 9.   Dose-response curves for full and partial agonists, and for antagonists. For symbols, see Eq. (15).

A ligand-receptor interaction usually requires a certain lipophilicity of the ligand. Cholinergic receptors can serve as an interesting example.

16.7.2 Partial Agonism in Cholinergic Systems

The stimulation of muscarinic receptors in different smooth muscles *in vitro* by alkylated derivatives of trimethylammonium (TMA) has been investigated by several authors [11, 12, 27, 28]. Results of these studies increasing the length of the alkyl chain (n) causes tighter binding to the receptor for methyl to pentyl derivatives, at full intrinsic activity (maximal E_{max}). Cases of partial agonism have been observed for hexyl to nonyl derivatives, antagonism first occurs for octyl derivatives. Let us now define the "response potential" ξ of a substance in the investigated system (analogously to "binding potential" for the binding a ligand to its binding site) as

$$\xi = \lim_{C \to 0} (dE/dC) = E_{max}/D_2. \tag{16}$$

The potential ξ represents the steepest change of response to an infinitesimally increasing dose C; it can easily be shown that this value is reached for very small doses, close to $C = 0$. When the E-scale is normalized by setting the highest E_{max} in the substance group equal to unit, individual E_{max} values can stand for the intrinsic activity α. Then, ξ values of full agonists are equal to the inverse of D_2, $1/D_2$, those of partial agonists are smaller then $1/D_2$. For antagonists, $\xi = 0$. Thus, the response potential adequately expresses the "agonistic power" of individual substances. Fig. 10 shows the distribution of ξ values with the length of the substituting alkyl group for TMA derivatives. The maximum of ξ occurs for $n = 5$ (pentyl derivative); a decrease has been identified for alkyl derivatives with $n > 6$. The antagonistic potency of these highly hydrophobic
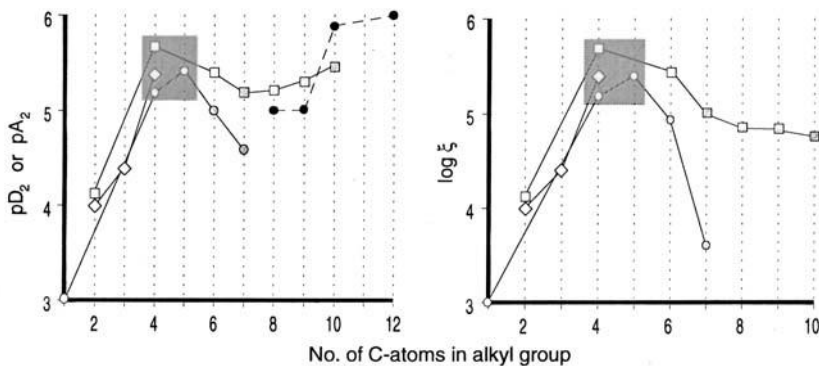
Figure 10. Values of pD_2 or pA_2 (cf. Eq. (15)), and response potential ξ (cf. Eq. (16)) for alkyl trimethylammonium derivatives, as a function of the alkyl-chain length (abscissa). The graph includes contractile responses of rat jejunum [27] (circles), frog rectus abdominis [11] (diamonds), and guinea-pig ileum [12] (squares). Open symbols are for full agonists, closed symbols for antagonists, shaded symbols for partial agonists. The hatched area denotes the region of optimal alkyl chain length. (Note that pA_2 and pD_2 values do not describe the same phenomenon! The graph is discontinuous; the connecting line shows the change between two neighboring points.)

derivatives, however, may even increase with the length of the substituent: pA_2 values seem to indicate a slight increase up to $n = 12$. The symbol A_2 stands for the antagonist concentration which depresses the response to a given agonist concentration to a value which would be achieved by the half of this latter concentration; pA_2 denotes its negative logarithm. The A_2 values are numerically close to the equilibrium dissociation constants of the receptor-antagonist complexes, expressed in terms of the total concentration in the organ medium or in the system.

16.7.3 Molecular Perturbation Hypothesis

These experimental findings motivated Belleau [29, 30] to formulate the "macromolecular perturbation theory" for muscarinic receptors. In short, it has been assumed that the "optimal" hydrophobicity of the ligand, corresponding to the length of five carbon atoms, causes a "specific conformational perturbation" of the receptor molecule followed by establishing a stable binding conformation in the cell membrane. A larger hydrophobicity increment favors a "nonspecific conformational perturbation" with a trend to antagonistic behavior, i.e., to the loss of intrinsic activity. Although being 30 years old, this notion summarizes salient features of the muscarinic systems.

The present knowledge indicates that the five subtypes of this receptor (M_1 to M_5) are associated with at least two types of G proteins (one of them activating the phospholipase C, the other one inhibiting the adenylate caclase), and all contain seven transmembrane spans in their molecules [31]. Only a minor conformational perturbation can therefore be expected as a consequence of ligand-receptor interaction in this, rather rigid, molecular structure. A certain hydrophobicity of the ligand, on the other hand, may be a condition for fitting the ligand into the hydrophobic pocket built up from tryptophan and aromatic amino acid residues. The binding site map is presented in Fig. 8. An excessive hydrophobicity of the side chain may, however, have an influence upon nonspecific (wrong-way) binding in the sense of the scheme on Fig. 6. This might correspond to the "nonspecific conformational perturbation" of Belleau and may be accounted for the change of the membrane environment of the receptor itself.

16.7.4 "Wrong-way" Binding Model of Partial Agonism

Let us consider a kinetic scheme of such a wrong-way binding involving ligand (L), receptor (R), and two intermediate complexes LR and LR', from which only the first one leads to the activated receptor form R*:

$$L + R \overset{K_x}{\leftrightarrow} LR \rightarrow\rightarrow\rightarrow\rightarrow R* \tag{17}$$

$$L + R \overset{K_y}{\leftrightarrow} LR'$$
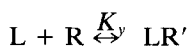
(K_x, K_y are the corresponding dissociation constants). When using simplified assumptions mentioned above, one arrives to a relationship similar to Eq. (15), with E and D given by

$$E_{max} = E^0_{max} \frac{1}{\varkappa + 1}$$ (18)

$$D_2 = K_x \frac{1}{\varkappa + 1}$$ *(19)*

where E^0_{max} is a reference value of maximal response (usually the highest one in the investigated group), $\varkappa = K_x/K_y$. Obviously,

$$E_{max}/E^0_{max} \equiv \alpha = \frac{1}{\varkappa + 1}$$ (20)

Assuming that the minimal required hydrophobicity of the ligand influences solely its fixation within the binding site but not its ability to activate the receptor, only the ligand-receptor dissociation constant of derivatives displaying suboptimal hydrophobicity (in our case, methyl to butyl derivatives) should be changed. Intrinsic activity of these short alkyl derivatives should stay constant at the maximum.

16.7.5 Effect of Lipophilicity on Intrinsic Activity

It seems likely that lipophilicity can exercise effects on both K_x, K_y, and consequently also on $\varkappa$. Then,

$$\alpha = \frac{1}{\varkappa \times \varphi(p) + 1}$$ (21)

where φ denotes a change of $\varkappa$ with changing lipophilicity p of the ligand expressed, for instance, in terms of the Hansch hydrophobic constant π [32],

$$p = 10^\pi.$$ (22)

Molecular mechanisms underlying wrong-way interactions are not sufficiently well known. Therefore, no thermodynamic model is available for an accurate determination of the function $\varphi(p)$. The effect of hydrophobicity on the ratio of the two K constants, however, depends upon the access of the substance to the receptor environment. This, in general, is an asymptotic process, showing a maximum for very high hydrophobicity, and a residual value for very low one. Most likely, $\varphi(p)$ can be matched by the Langmuir isotherm type of relationship

$$\varphi(p) = p_0 + p_{max}\frac{(K_p p)^\nu}{(K_p p)^\nu + 1}$$ (23)

where p_0 is the function value at $p = 0$, p_{max} the constant determining the maximal access $p_0 + p_{max}$, K_p an affinity constant for the access of the substance to the receptor environment, and ν is a corresponding power coefficient of the Hill type (it can both positive and negative values). Since the affinity of the membrane, reflecting mainly ligand partitioning between membrane environment and membrane interior, is low, lipophilic membrane perturbations (with a subsequent effect upon dissociation constants K_x, K_y) may rather be matched by a power function in p,

$$\varphi(p) = p_0(1 + \frac{1}{a_0}p^\nu)$$ (24)

where a_0 is defined as

$$a_0 = \frac{p_0}{p_{max}K^{\gamma}_p} \tag{25}$$

The effect of the lipophilicity p on α is then described by

$$\alpha = \frac{\alpha_{max}}{\varkappa + (a_0 + p^{\gamma}) + 1} \tag{26}$$

where $\varkappa + = \varkappa p_0/a_0$, α_{max} determines the maximal value of α within the investigated group of substances. Scaling of α within the conventional limits $0 < \alpha < 1$ yields

$$\alpha = \frac{\varkappa + a_0 + 1}{\varkappa + (a_0 + p^{\gamma}) + 1} \tag{27}$$

Results for TMA derivatives are summarized in Fig. 11. The graphs in this figure show partial agonism in musculotropic responses in three systems: guinea-pig ileum, rat jejunum and frog rectus abdominis muscle. The legend to Fig. 11 indicates values of individual binding parameters. Nonlinear regression routines applied to Eq. (27) yield somewhat different fits in individual instances but the shape of the α, π relationship is uniform. It is also possible to carry out a common fit with undifferentiated data from all three systems; in all instances, the partial agonists appear in an equal hydrophobicity domain.

16.7.6 Other Examples of Full-to-Partial-Agonism Transition due to Lipophilicity Increase

Increasing lipophilicity of the ligand molecule seems to be a general cause of transition from full to partial agonism and to antagonism (cf. Fig. 9). Several examples are to be found in the older literature [11, 33, 34]. Fig. 12 shows available data and curve fits to Eq. (27).
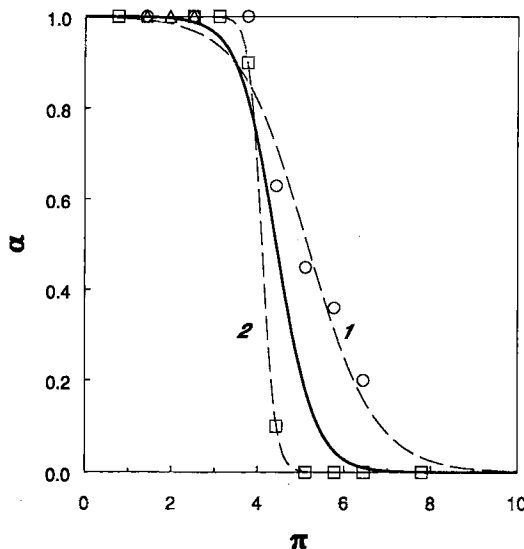
Figure 11. Relationship between hydrophobic constant π of the alkyl groups [32] and intrinsic activity α: TMA alkyl derivatives. Contraction of 1. guinea-pig ileum [12] (circles), 2. rat jejunum [27] (squares) and frog rectus abdominis [11] (triangles). Curves represent fits to Eq. (27): dashed lines to individual sets 1 ($\varkappa'$ = 1.25 $\times 10^{-3}$, $v = 0.56$) and 2 ($\varkappa'$ = 2.01 $\times 10^{-12}$, $v = 2.85$), solid line to pooled data 1 and 2 ($\varkappa'$ = 3.93 $\times 10^{-5}$, $v = 0.99$); in all instances, $a_0 \approx 0$.
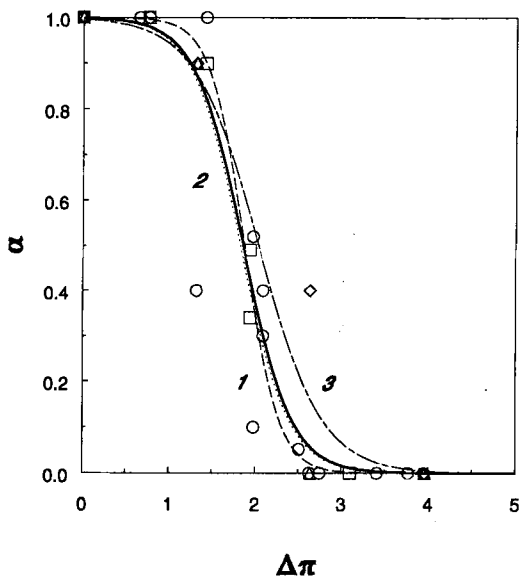
Figure 12. Relationship between hydrophobic constant π of the alkyl groups and intrinsic activity α (cf. Fig. 11): various groups of substances. 1. 1,3-Dioxolane derivatives (2-alkyl$_1$-4-methyl-trialkyl$_2$-ammonium-1,3-dioxole), rat jejunum [33] (circles, dashed line); 2. N-alkylated noradrenaline derivatives, rat vas deferens [34] (squares, dotted line); 3. N-alkylated succinylbischoline (triangles) and adipinylbischoline (diamonds) derivatives, frog rectus abdominis [11] (dashed-dotted line). Since different groups were compared, substituent hydrophobicity was related to the calculated π of all substituents *minus* π of a standard (hydrogen atom in the first two cases, two thrimethylammonium groups in the latter instance). Full line: pooled data. Values of $\varkappa'$ were in the range of 8.7×10^{-5} to 3.0×10^{-3}, those v in the range 1.18 to 2.17; $a_0 \approx 0$.

Values of power coefficients v are presented in legends to Figs. 11 and 12. They are positive, indicating either a negative effect of lipophilicity on the binding to the agonistic site (K_x increases with π), or a positive effect on the wrong-way binding (K_y decreases with π). These two alternatives cannot be differentiated without a further experimental evidence, and moreover, it is likely that both effects are exercised concomitantly. However, strong negative effects of lipophilicity upon "correct" binding are probably not dominating, and thus, the latter mechanism is more plausible.

As a methodological note, it should be mentioned that values of a_0 are generally very small and coefficients of asymptotic correlations with $\varkappa$ are close to -1. Thus, a computation of a_0 can be left out.

16.8 Bell-Shaped Dose-Response Curves

Wrong-way binding can be the cause of another, not satisfactorily explained phenomenon in the dose-response analysis, the down-turn phase of the curve at higher substance concentrations, resulting in a bell-shaped curve [35, 36]. A number of specific

assumptions may explain this behavior [36–38]; their verification, however, is insufficient. When taking a general view, any molecule (particularly a small one) in the receptor environment may at a higher concentration elicit less specific effects on the receptor conformation and change its reactivity. Van der Waals dispersion forces, membrane rearrangements in the vicinity of receptors, or even induced changes in electric charges are likely to play a role in these, not closely recognized, processes. Let us use for all of them the general term *receptor perturbations*.

The perturbation modifies primarily the strength of ligand-receptor bonds. Its effect on the parameter D_2 in Eq. (15) is therefore some function of the concentration C_s, Θ (C_s). Its most likely type resembles Eq. (10), describing the membrane access of ligand, Γ, in which the experimentally not accessible concentration C_s is substituted by C:

$$\Theta(C) = \frac{(K_\Theta C)^\mu}{(K_\Theta C)^\mu + 1};$$ (28)

the values of the function $\Theta(C)$ lie in the interval $[0, 1]$. The association constant K_Θ for ligand-macromolecule interactions causing a perturbation is rather low (mainly weak dispersion forces are engaged), $K_\Theta C \ll 1$, and the denominator in Eq. (28) can simply be set to one. An extended response function (corresponding to Eq. (15)) becomes then

$$E = E_{max}\frac{C^\omega}{D{+_2}^\omega + C^\omega},$$ (29)

where

$$D{+_2} = D_2^\omega (1 + \Theta(C))$$ (30)

and consequently,

$$E = E_{max}\frac{C^\omega}{D_2^\omega (1 + (K_\Theta C)^\mu) + C_\omega}.$$ (31)

The constants D_2, K_Θ, E_{max} and ω (an empirical power coefficient, see above) can be estimated by nonlinear regression methods, as suggested in an earlier communication [36]. At least a part of the dispersion forces leading to a perturbation is associated with membrane effects and therefore with weak hydrophobic ligand-membrane interactions. Thus, the constant K_Θ is likely to be a function of hydrophobicity parameters, e.g., of π. Fig. 13 brings an example from older literature: the contractile response of the from rectus abdominis muscle to alkyltrimethylammonium derivatives (ethyl to dodecyl) [39, 40]. Its upper part shows that, in accord to a general notion, the constant D_2 reflecting the ligand affinity to its receptor displays a clear-cut lipophilicity optimum. The equilibrium constant K_Θ, on the other hand, indicates an increasing effect of lipophilicity on the occurrence of the downhill phase and, assumingly, an increasing perturbation effect.
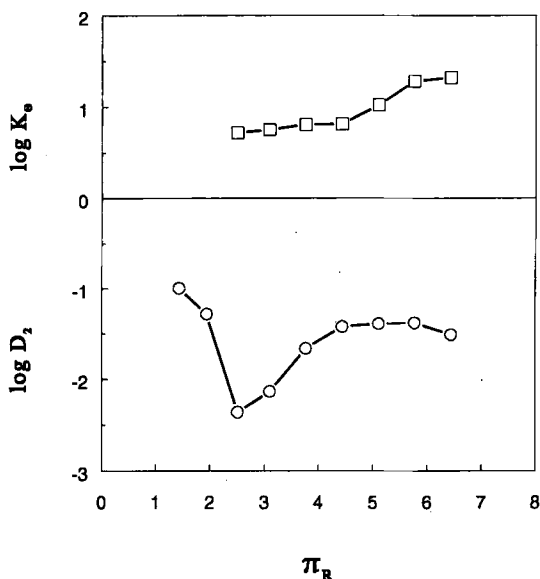
Figure 13. Parameters D_2 and K_Θ (Eqs. (28–31)) of the bell-shaped dose-response curves of alkyl trimethylammonium derivatives [39, 40], as a function of lipophilicity. π_R is the Hansch hydrophobicity constant of the alkyl chain [32]. D_2 and K_Θ are in mmol L^{-1}.

16.9 Thermodynamic Aspects of Variable Intrinsic Activity

16.9.1 Hydrophobic Interactions as an Entropy-Driven Process

The role of hydrophobic forces in ligand-receptor interactions and their relationships to agonistic/antagonistic behavior of active substances indeed awoke considerable interest in the past few decades [41–43].

According to the current notion, a conformational transition of a macromolecule due to hydrophobic forces consists virtually of two subsequent processes: 1) contacts of nonpolar groups with water are diminished, resulting in a positive change of enthalpy; 2) water structure in the environment of the macromolecule becomes "less perfectly" ordered, since water molecules are displaced from the diminished accessible area of the nonpolar groups: water structure is "more perfect" when the surrounding surface of such groups, and therefore, the entropy of the system increases. It is likely that such events follow any ligand-receptor interaction, providing that hydrophobic forces play an important role in the ligand-macromolecule complexes.

Hydrophobic interactions are *entropy-driven* processes [44, 45], i.e., the entropic term $T \Delta S^\circ$ determines the negative sign of the standard Gibbs free-energy change, ΔG°,

$$\Delta G^\circ = \Delta H^\circ - T \Delta S^\circ \tag{32}$$

(ΔH°, ΔS° are changes of standard enthalpy and entropy, respectively, T is the absolute temperature, $R = 8.314 J\ K^{-1}mol^{-1}$ is the gas constant). Processes in living systems are usually exergonic ($\Delta G^\circ < 0$). They are in the rule associated with a rather large positive entropy change that overweights the positive enthalpy change, so that the condition

$$T \Delta S° > \Delta H° > 0 \tag{33}$$

is valid at least for temperatures around 300 K. Values of $\Delta S°$ between 7 and 45 J K^{-1} mol^{-1}, and those for $\Delta H°$ between 1.2 and 7.5 Jmol^{-1} have been reported for systems of two interacting protein molecules [43]. Interactions involving small ligand molecules display enthalpy and entropy changes in a broad range of values (in particular with regard to $\Delta H°$ which is frequently negative) but the Gibbs free-energy change always attains negative values.

Relationship (33) defines conditions necessary for the existence of hydrophobic interactions but is, on itself, no proof of them. Other binding processes may also display a similar pattern. Occasionally, when $\Delta H°$ determines the negative sign of the Gibbs energy change (at $\Delta S° < 0$), the biological process is entropy-driven. Such thermodynamic conditions fully exclude a major role of hydrophobic interactions in the respective process. A diagram of driving forces is presented in Fig. 14.

Thermodynamic data for several ligand-receptor systems have already been published. As mentioned, however, they do not indicate unambiguously the prevailing type of interaction forces. Some of them are presented in more details in the following.

16.9.2 $\Delta S°$, $\Delta H°$ Relationships in Some Receptor Systems

16.9.2.1 Muscarinic Receptors

Intrinsic activities of alkyl trimethylammonium (TMA) derivatives for contraction of guinea pig ileum (muscarinic acetylcholine receptors) were correlated with thermodynamic parameters obtained for binding of individual members of the series to acetylcholinesterase [30]. Values of $\Delta S°$ for agonists (methyl- to hexyl-TMA) occurred in the range of -43 to -13, for antagonists (decyl- to dodecyl-TMA) 12 to 42 J K^{-1}mol$^-$; partial agonists (heptyl- to nonyl-TMA) possessed values around zero. The author speaks about "ordering" and "disordering" perturbation with the putative re-
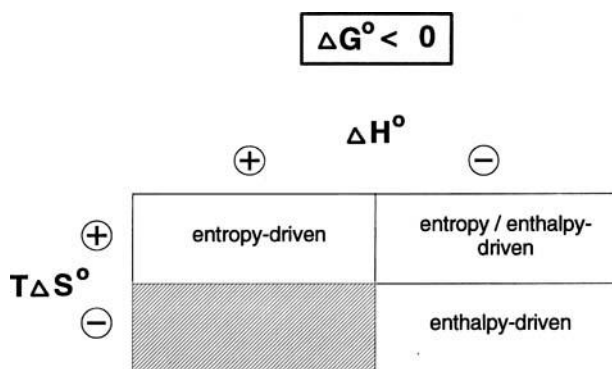
Figure 14. Driving forces in the entropy-, enthalpy-, and entropy/enthalpy-driven processes. Open fields indicate conditions under which spontaneous processes ($\Delta G° < 0$) may take place. The hatched field shows the domain of solely endergonic (nonspontaneous) interactions ($\Delta G° > 0$).

$\Delta S°$ [J K^{-1} mol^{-1}]
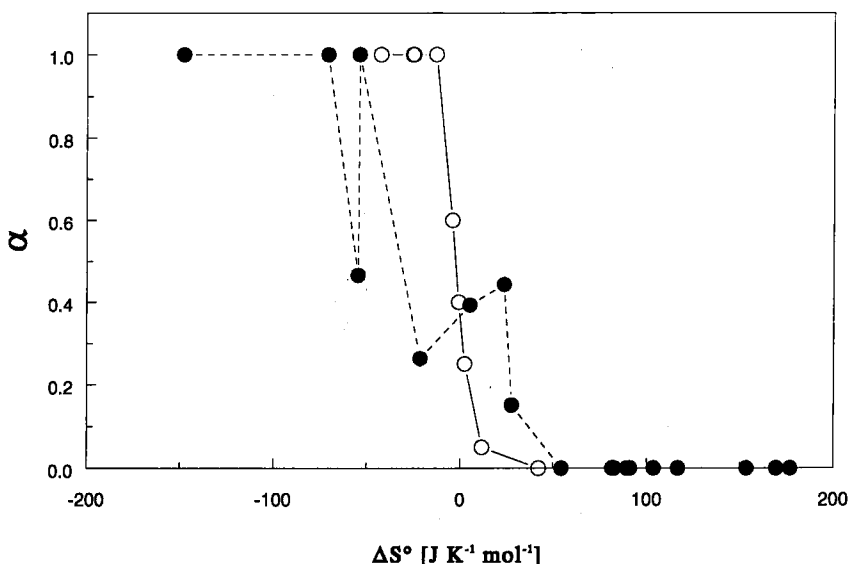
Figure 15. Relationships between the intrinsic activity α and the entropy change of binding for alkyl thimethylammonium derivatives[30] (open circles, full line) and for various substances acting on β_2-adrenergic receptor [46] (closed circles, broken line). First system: α valuesare from the contraction response of the guinea-pig ileum, $\Delta S°$ was estimated from the binding of individual ligands to acetylcholine esterase (correlation with $\Delta S°$ for binding to muscarinic receptors was assumed but not proved). Second system: $\Delta S°$ for binding of ligands to turkey erythrocyte membranes, a related to the cAMP production stimulated (or inhibited) by individual substances.

ceptor (Fig. 15). Hydrophobic interactions cannot in this instance play a decisive role in the binding of agonists which is clearly enthalpy-driven ($\Delta S°$ is negative). It might be argued, however, that thermodynamic parameters of another binding site, not necessarily akin to the one of the responding system on molecular level, were used for comparison.

Fig. 16 (panel A) shows the relationship between $\Delta S°$ and $\Delta H°$ (called *"HS plot"* in the following text) for this system. This is linear and displays a clear-cut clustering of substances according to their intrinsic activities (agonists, partial agonists and antagonists) visualized by different symbols. Such a pattern occurs also in some, but not all, ligand-receptor systems (see below).

16.9.2.2 β-Adrenergic Receptors

A similar relationship, this time for ligand binding and a subsequent activation of adenylate cyclase linked to the β-adrenergic receptor, has been found in turkey erythrocyte membranes [46]. A number of substances active as agonists, partial agonists, or antagonists display the same profile in the intrinsic activity α vs $\Delta S°$ (Fig. 14) as in the previous case.

ΔS° [J K⁻¹ mol⁻¹]

ΔH° [kJ mol⁻¹]
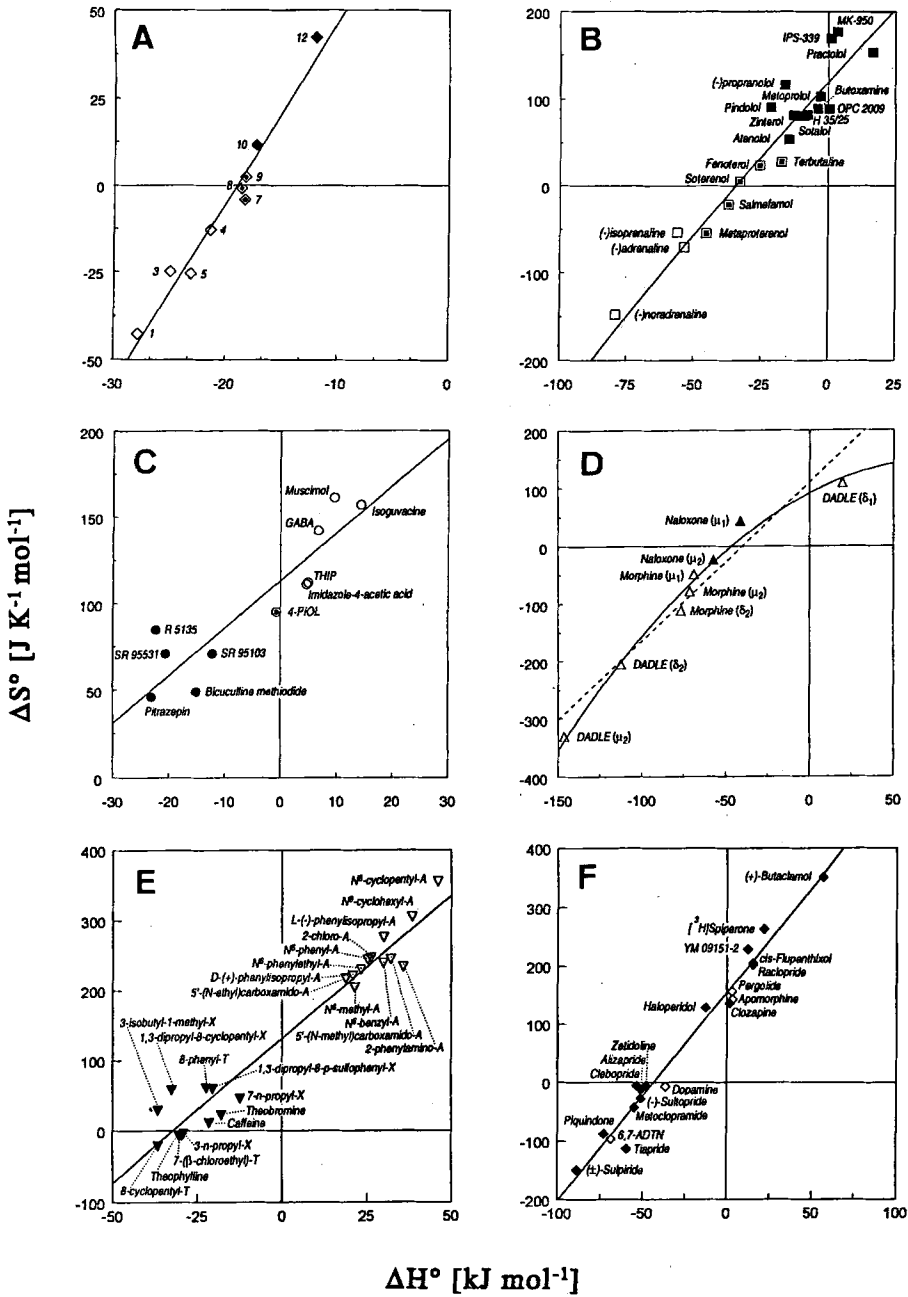
16.9.2.3 GABA$_A$ Receptors

More recently, a similar study has been carried out with the GABA$_A$ receptor in synaptosomal membranes (rat brain) and a number of ligands possessing different chemical compositions [47]. Thermodynamic data correlate also in this instance with the agonistic/antagonistic action on ^{36}Cl uptake in rat-brain synaptosomes. All $\Delta S°$ values were positive, $\Delta H°$ values positive for agonists and negative for antagonists. Thermodynamic parameters of this kind might be expected for binding forces dominated by hydrophobic effects. Contrary to the former two instances, this is valid for agonists. Binding of antagonists is driven by both forces. Unfortunately, lipophilicity data for individual ligands are at present not available.

16.9.2.4 Opioid Receptors

Similar thermodynamic data were found for binding of morphine, (D-Ala^2, D-Leu^5)enkephalin (DADLE), and the opioid receptor antagonist naloxone in rat-brain membranes [48]. Binding to individual opioid receptor classes II (μ_1, μ_2, δ_1, δ_2) was considered. Here, one of the ligands (DADLE) was a peptide; it can be assumed that binding thermodynamics of flexible ligand molecules may differ considerably from those with a rigid molecular backbone. In fact, this could not be proved with these limited data. Both highly negative to positive $\Delta S°$ values were found, without a clear-cut relationship to agonism and antagonism of thee substances in an *in-vivo* assay system. A highly negative value ($\Delta S° = -332$ J K^{-1}mol^{-1}) has been reported for binding of the μ-agonist DADLE to μ_2 receptors, a positive value ($\Delta S° = 44$ J K^{-1}mol^{-1}) for the μ-antagonist naloxone. These differences do not follow a definitive general trend.

16.9.2.5 Adenosine A$_1$ Receptors

A clustering of substances [49] into groups of entropy-driven binding (agonists) and enthalpy- or entropy/enthalpy-driven binding (antagonists) to rat brain membranes can clearly be recognized in the HS plot ion Fig. 16, panel E.

◀ **Figure 16.** HS plots for ligand binding. Open symbols: full agonists; closed symbols: antagonists; combined symbols: partial agonists. A, binding of alkyl trimethylammonium derivatives to acetylcholine esterase [30]; B, binding of various drugs to β_2-receptors in turkey erythrocytes [46]; C, binding of various substances acting upon GABA$_A$ receptor to rat-brain membranes [47]; D, binding of various ligands (peptides and nonpeptides) to opioid μ- and δ-receptors in rat-brain membranes [48] (the broken line shows the linear fit, the solid line the fit by a second-order polynomial which is significantly better than the linear fit); E, binding of agonists and antagonists to the adenosine A$_1$ receptor in rat brain membranes [49]; F, binding of agonists and antagonists to the dopamine D$_2$ receptor in rat striatal membranes [50]. The solid lines in panels A, B, C, E, and F represent the least-squares linear fits.

16.9.2.6 Dopamine D_2 Receptors

Finally, in the case of ligands interacting with D_2 receptors in the rat striatal membranes [50], no particular clustering with regard to their intrinsic activities could be detected in the HS plot. Driving forces are distributed from purely entropy- to purely enthalpy-driven.

16.9.3 Entropy-Enthalpy Compensation

As mentioned, thermodynamic data in all systems presented above obey the rule of linear relationship between $\Delta H°$ and $\Delta S°$:

$$\Delta S° = h_0 + h_1 \, \Delta H° \tag{34}$$

h_0, h_1 are empirical constants. This empirical relationship [51], valid almost universally for families of similar interactions, indicates that an increasing enthalpy is compensated by an increasing entropy of the system; the two thermodynamic properties are strongly interdependent. The intercept, h_0, determines the entropy change of the boundary between endothermic and exothermic interactions in the respective reaction series. The slope h_1 has a dimension of inversed thermodynamic temperature. Its physical meaning is not obvious but, as first time mentioned by Kauzmann [41], it shows numerically similar values only for processes in physically similar systems (for instance, it is very different for the water in ordinary ice and for the one attached to a hydrophobic surface, forming the so-called *"icebergs"*).
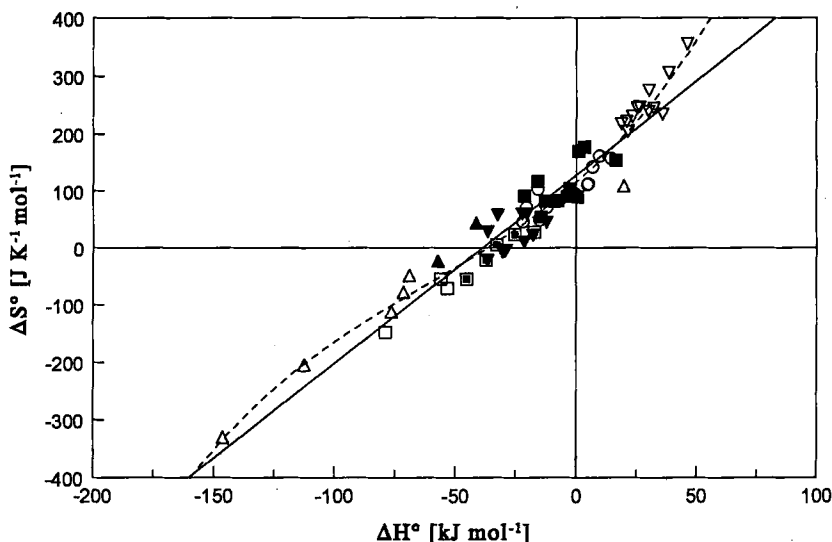
Figure 17. HS plot for ligand binding: pooled data from Fig. 16 (symbols are identical). Solid line: linear fit. Broken line: fit to second order polynomial (not significantly better than linear fit).

Individual plots are presented in Fig. 16. All fits are significantly linear, i.e., they cannot be significantly improved by a fit to a second-orderpolynomial, with the exception of one case (opioid receptors). All sets display similar and insignificantly different values of both h_0 and h_1 coefficients. Pooled data from all systems considered here can also be fitted by Eq. (34) (Fig. 17). There seem to be some domains of agonism, partial agonism, and antagonism in these graphs but they are not uniform and lack sharpness. However, a conclusion that the position of a ligand in the HS plot is closely linked to its intrinsic activity in the given responding system, is by and large correct. Thermodynamic analysis using these approaches may offer additional tools for identifying binding forces acting in qualitatively different ligand-receptor complexes, identify differences in the nature of agonists and antagonists binding, and facilitate an insight into the phenomenon of partial agonism.

Even more important may become the deductive rule of a uniform entropy-enthalpy compensation effects in manifold ligand-receptor systems. Tests of its general validity and its limits require additional extensive studies.

16.10 Outlook

The final comment on this topic can be very short. Besides the generally recognized role of lipophilicity of the bioactive substance in distribution, transport and inactivation processes, the lipophilicity of biologically active compounds influences interactions of these compounds with the cell membrane and thus, the membrane and receptor conformation. Phenomena like partial agonism and bell-shaped dose-response curves may in some instances be explained by these interactions. Such notions are not new. However, systematic studies to these aims have been accomplished in an early phase of molecular pharmacology; additional data are still lacking. They may facilitate a more relevant evaluation of future experiments, and perhaps also a closer look on the cell interactions with smaller molecules.

References

[1] Ariëns, E. J., Introductory remarks. In: *Physico-Chemical Aspects of Drug Action, Proceedings of the Third International Pharmacological Meeting, Sao Paulo, 1966*, Vol. **7**. Ariëns, E. J. (Ed.). Pergamon Press: Oxford 1–4 (1968)

[2] Rudinger, J., Pliška, V., and Kejúcáü, I., *Rec. Prog. Hormone Res.* **28**, 131–166 (1972)

[3] Pliška, V., *Arzneim.-Forsch.* **16**, 886–893 (1966)

[4] Pliška, V., *II Farmaco, Ed. Sci.* **23**, 623–641 (1968)

[5] Pliška, V., *Experientia* **47**, 216–221 (1991)

[6] Pliška, V. The Time course of the biological response to drugs: a mathematical treatment. In: *Physico-Chemical Aspects of Drug Action, Proceedings of the Third International Pharmacological Meeting, Sao Paulo, 1966, Vol.* **7**. Ariëns, E. J. (Ed.). Pergamon Press: Oxford 115–126 (1968)

[7] Hill, A. V., *Biochem. J.* **7**, 471–480 (1913)

[8] Parker, R. B., and Waud, D. R., *J. Pharmacol. Exp. Ther.* **177**, 1–12 (1971)

[9] Pliška, V., *Trends Pharmacol. Sci.* **8**, 50–52 (1987)

[10] Clark, A. J., *Handbuch der experimentellen Pharmakologie, Ergänzungswerk IV: General Pharmacology.* Verlag von Julius Springer: Berlin 1937

[11] Ariëns, E. J., van Rossum, J. M., and Simonis, A. M., *Azneim.-Forsch.* **6**, 282–293 (1956)

[12] Stephenson, R. P., *Br. J. Pharmacol.* **11**, 379–393 (1956)

[13] Schwyzer, R., *CHEMTRACTS – Biochemistry and Molecular Biology* **3**, 347–379 (1992)

[14] Sargent, D. F., and Schwyzer, R., *Proc. Natl. Acad. Sci. USA* **83**, 5774–5778 (1986)

[15] Bean, J. W., Sargent, D. F., and Schwyzer, R., *J. Receptor Res.* **8**, 375–389 (1988)

[16] Sargent, D. F., Bean, J. W., Kosterlitz, H. W., and Schwyzer, R., *Biochemistry* **27**, 4974–4977 (1988)

[17] Kosterlitz, H. W., Paterson, S. J., and Robson, L. E., *J. Receptor Res.* **8**, 363–373 (1988)

[18] Pliška, V., and Marbach, P., *Eur. J. Pharmacol.* **49**, 213–222 (1978)

[19] Schwyzer, R., *J. Receptor Res.* **11**, 45–57 (1991)

[20] Burgen, A. S. V., *J. Pharm. Pharmacol.* **18**, 137–149 (1966)

[21] Burgen, A. S. V., Roberts, G. C. K., and Feeney, J., *Nature* **253**, 753–755 (1975)

[22] Lindeberg, G., Vilhardt, H., Larsson, L. E., Melin, P., and Pliška, V., *J. Receptor Res.* **1**, 389–402 (1980)

[23] Hibert, M., Trumpp-Kallmeyer, S., Bruinvels, A., and Hoflack, J., *Mol. Pharmacol.* **40**, 8–15 (1991)

[24] Honig, B., and Nicholls, A., *Science* **268**, 1144–1149 (1995)

[25] Chen J. K., and Schreiber, S. L., *Angew. Chem. Int. Ed. Engl.* **34**, 953–969 (1995)

[26] Furchgott, R. F., *Adv. Drug. Res.* **3**, 21–55 (1966)

[27] van Rossum, J. M., and Ariëns, E. J., *Arch. Int. Pharmacodyn.* **118**, 418–443 (1959)

[28] Paton, W. D. M., *Proc. R. Soc. B* **154**, 21–69 (1961)

[29] Belleau, B., *J. Med. Chem.* **7**, 776–784 (1964)

[30] Belleau, B., *Adv. Drug Res.* **2**, 89–126 (1965)

[31] Watson, S., and Arkinstall, S., *The G-Protein Linked Receptor Facts Book.* Academic Press: London 1994, 7–18

[32] Hansch, C., and Leo, A. J., *Substituent Constants for Correlation Analysis in Chemistry and Biology.* Wiley: New York, 1979

[33] Fourneau, E., Bovet, D., Bove, F., and Montezin, G., *Bull. Soc. Chim. Biol.* **26**, 516–527 (1944)

[34] Ariëns, E. J., and Simonis, A. M., *Arch. Int. Pharmacodyn.* **127**, 479–496 (1960)

[35] Ariëns, E. J., and Simonis, A. M., Drug-receptor interaction: interaction of one or more drugs with different receptor systems. In: *Molecular Pharmacology. The Mode of Action of Biologically Active Compounds,* Vol. **I**. Ariëns, E. J. (Ed.). Academic Press: Inc: New York, 1964

[36] Pliška, V., *Trends Pharmacol. Sci.* **15**, 178–181 (1994)

[37] Szabadi, E., *J. Theor. Biol.* **69**, 101–112 (1977)

[38] Rovati, G. E., and Nicosia, S., *Trends Pharmacol. Sci.* **15**, 140–144 (1994)

[39] Ariëns, E. J., Simonis, A. M., and de Groot, W. M., *Arch. Int. Pharmacodyn.* **100**, 298–322 (1955)

[40] Ariëns, E. J., van Rossum, J. M., and Simonis, A. M., *Arzneim.-Forsch.* **6**, 611–621 (1956)

[41] Kauzmann, W., *Adv. Protein Chem.* **14**, 1–63 (1959)

[42] Tanford, C., *The Hydrophobic Effect: Formation of Micelles and Biological Membranes,* Wiley: New York 1973

[43] Zahradník, R., and Hobza, P., *Intermolecular Complexes. The Role of van der Waals Systems in Physical Chemistry and in the Biodisciplines.* Academia: Praha, 1988

[44] Lauffer, M. A., *Entropy-Driven Processes in Biology. Polymerization of Tobacco Mosaic virus Protein and Similar Reactions.* Springer: Berlin, 1975

[45] Lauffer, M. A., Entropy-driven polymerization of proteins: Tobacco mosaic virus and other proteins of biological importance. In: *Physical Aspects of Protein Interactions, Proceedings of the Symposium on Protein Interactions (American Chemical Society Meeting), Miami Beach, FA, September 12–13, 1978.* Catsimpoolas, N. (Ed.) Elsevier North Holland: Amsterdam 115–170 (1978)

[46] Weiland, G. A., Minneman, K. P., and Molinoff, P. B., *Nature* **281,** 114–117 (1979)

[47] Maksay, G., *Mol. Pharmacol.* **46,** 386–390 (1994)

[48] Varfolomeev, S. D., Zaitsev, S. V., Kurochkin, I. N., Porodenko, N. V. The kinetics and mechanism of ligand-receptor interactions in the system of opiate receptors. In *Synaptic Transmitters and Receptors.* Tucek, S. (Ed.). Wiley: Chichester 195–201 (1987)

[49] Borea, P. A., Varani, K., Guerra, L., Gilli, P., and Gilli, G., *Mol. Neuropharmacol.* **2,** 273–281 (1992)

[50] Kilpatrick, G. J., El Tayar, N., van de Waterbeemd, H., Jenner, P., Testa, B., and Marsden, C. D., *Mol. Pharmacol.* **30,** 226–234 (1986)

[51] Zahradník, R., Hobza, P., and Slanina, Z., Calculations of Henry constants and partition coefficients using quantum chemical approach. In: *Quantitative Structure-Activity Relationships, Proceedings of the Conference on Chemical Structure-Biological Activity Relationships: Quantitative Approaches, Prague, June 27–29, 1973.* Tichý, M. (Ed.). Birkhauser: Basel 217–230 (1976)

17 Membrane Transport and Cellular Distribution

Štefan Baláž

Abbreviations

A+,A− High and low amphiphilicity
F   Fisher criterion
L+,L− High and low lipophilicity
QSAR   Quantitative structure-activity relationships
sd   Standard deviation
sgn   sgn $= -1$ for acids and sgn $= 1$ for bases

Symbols

A, B, C, D, E, η_i Adjustable parameters
$A(\mathrm{pp}, t)$ Disposition function with the time of exposure (t) and physico-chemical properties (pp) of both the drugs and biological systems as variables
c Actual drug concentration
c_A Concentration of free non-ionized molecules in aqueous compartments
c_X Equipotent concentration eliciting the fraction X of the maximum effect
c_0 Drug concentration in the entry compartment
D Diffusion coefficient in the diffusion layer with the effective thickness h
I_{trans} $I_{\mathrm{trans}} = 0$ for *cis*-derivatives and $I_{\mathrm{trans}} = 1$ for *trans*-derivatives
h Effective thickness of a diffusion layer
$k_{\mathrm{aq}}, k_{\mathrm{org}}$ Rate constant of diffusion in the diffusion layer given by the subscript
l_1, l_2 Rate constants of transport for the direction from water to the organic phase and backwards, respectively
K Drug-receptor association constant
K_i Association constant for protein binding in the *i*-th compartment
n Number of experimental points
$\mathrm{pH_e}, \mathrm{pH_i}$ pH values in the external medium and in the i-th compartment
P Reference partition cofficient (usually in the system 1-octanol/water)
P_M Membrane/water partition coefficient
r Correlation coefficient
Y Surrogate variable
β Empirical exponent from the Collander equation (Eq. (3))
ϑ Sensitivity of the intracellular phases to the change in the external pH

17.1 Introduction

The design of better anthropogenic chemicals requires understanding of their bioavailability in terms of structure and physico-chemical properties. Among attempts to solve the task two main directions can be distinguished, based on either classical or subcellular pharmacokinetics. The chemicals will be called "drugs" in the following, albeit the presented treatment is valid also for other types of biologically active compounds.

1. Classical pharmacokinetics. This describes drug disposition in the terms of the space-averaged drug concentrations in macroscopic nonhomogeneous compartments. Though these concentrations are of great value for the practical purposes of chemotherapy, actual drug concentrations in the immediate surroundings of the receptors are required for the analysis of drug effects at the molecular level. Such data are provided by subcellular pharmacokinetics.

2. Subcellular pharmacokinetics. This aims at a description of the kinetics of drug distribution in individual physically distinct cellular compartments like the extra- and intracellular aqueous phases and membranes. This broadens the scope to the fate of drugs in microorganisms, suspensions of cells or subcellular particles, in addition to higher organisms. If the processes controlling the drug disposition exhibit first-order kinetics [1, 2], the time-course of the drug concentration [D] in the receptor surroundings can be expressed as:

$$[D] = A(pp,t)\, c_0 \tag{1}$$

where c_0 is the drug concentration in the entry compartment and $A(pp,t)$ is the *disposition function* with the time of observation (t) and physico-chemical properties (pp) of both the drugs and biological system as variables [3]. The actual form of the disposition function depends on our need for either kinetic or fixed-time expression, on the complexity of the biological system, on the drug properties, and on the nature of dosing (single, repeated, continuous).

The formulation of model-based QSAR (quantitative structure-activity relationships) for the fate and effects of drugs in biological systems requires: 1) construction of an adequate model; 2) the description of the model rate or equilibrium constants in terms of either physico-chemical properties or structure. Usually, the rate constants are suitable descriptors for the processes achieving the equilibrium or steady state at the time-scale comparable with, or larger than, the duration of the experiment; for faster processes the equilibrium constants are used.

Membrane transport and accumulation, protein binding, ionization, and noncatalyzed reactions can be characterized using physico-chemical properties measured in artificial chemical systems that are usually simpler than the original biological systems. They are celled *structure-nonspecific processes*, in contrast to drug binding to special macromolecules like receptors and metabolizing enzymes, for which artificial systems bearing certain similarity to the biological counterparts do not exist. The interactions of drugs with the two classes of macromolecules are mainly responsible for selectivity of the drug action. Consequently, they must be described in the terms of the exact three-dimensional structures and energetics of the drug molecules. Therefore, they are referred to as *structure-specific processes*.

Occasionally, within a limited drug series, the binding of drugs to receptors and metabolizing enzymes is approximately constant or depends on a single property of the drugs. Solely in these cases a continuous dependence of distribution or biological activity on the drug properties can be observed in experimental data. Such sets provide the basic verification tests of the models of membrane transport and cellular distribution which involve exclusively structure-nonspecific processes.

When the biological response is an immediate consequence of fast and reversible drug-receptor interaction and is proportional to the fraction of the receptors occupied, the biological activity after a single dose reflects the drug concentration in the receptor surroundings [3]. In the following equations, if distribution and biological activity are given interchangeably on the left-hand side, the above conditions are assumed to hold.

17.2 Model

The fate of drugs in a biological system, as a complex event, can be analyzed most conveniently with the help of a model. Physico-chemical properties, chemical structure of the drug, and the time of exposure are input variables in the models of subcellular pharmacokinetics.

17.2.1 Model Construction

Most drugs cross biological membranes by passive diffusion through the lipid core of the membrane that has completely different solvation properties from the aqueous phases. Therefore, subcellular pharmacokinetics considers the biological system as a catenary set of homogeneous compartments representing membranes and extra- and intracellular aqueous phases. Diffusion of the drugs within the bulk of the compartments is assumed to be instantaneous due to the small dimensions of the compartments. The actual number of compartments in the model and their assignment to the individual parts of the biological system do not depend on the system alone, but are influenced also by lipophilicity and amphiphilicity of the distributed drugs. Amphiphilicity can be defined as the degree of separation of hydrophilic and lipophilic parts of the molecule in at least one of its allowed conformations into two subspaces defined by a planar dissection of space. In the membranes lipophilic drugs (L+) are accumulated in the lipid core and amphiphilic drugs (A+) are bound to the interfaces. Thus each membrane can represent none (for hydrophilic and nonamphiphilic drugs – L−, A−), one (the lipid core – for L+, A−), two (the membrane/water interfaces – for L−, A+), or three (the lipid core plus both the membrane/water interfaces – for L+, A+) compartments depending upon the two drug properties. For nonamphiphilic drugs with varying lipophilicity each membrane represents one compartment and the distribution in a biological system can be described by the scheme given in Fig. 1. All the processes taking place in the same compartment, proceeding on similar time scales, and dependent on the same physico-chemical properties (protein binding, spontaneous reactions) have been grouped together and expressed by one variable.
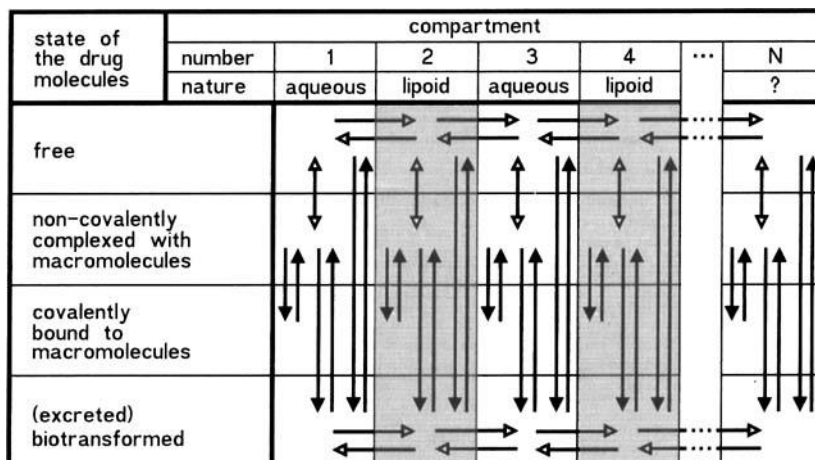
Figure 1. A schematic outline of the drug distribution in a morphologically compartmentalized system consisting of alternating aqueous phases and membranes. Two-ended arrows represent fast processes, one-ended arrows time-dependent processes. Full heads of the arrows indicate processes involving covalent bond formation; empty heads noncovalent processes [4].

17.2.2 Relation Between Individual Distribution Processes and Drug Properties

In order to convert the explicit solution of the set of differential equations describing the scheme in Fig. 1 to a QSAR equation or to perform a simulation *via* numerical integration with a QSAR output, the rate and equilibrium parameters for individual processes outlined in Fig. 1 must be expressed as a function of physico-chemical properties of both drugs and biological system.

17.2.2.1 Transport Through Phase Interface

It is generally accepted that passive membrane permeation through the lipid core is of purely physical nature following Fick's law. In the physical sense, a certain resemblance exists between membrane/water and organic solvent/water interfaces: they both possess the diffusion layers located at both sides of the polar/apolar interface, which are more structured than the bulk due to solvation. The drug transport between two immiscible phases can be studied conveniently in a thermostatted vessel with the organic and aqueous phases sitting one above the other, both being stirred. The rate constants of transport l_1 and l_2 for the direction from water to the organic phase and in the reverse direction, respectively, depend on lipophilicity (the 1-octanol/water partition coefficient P) [5]:

$$l_1 = \frac{AP}{BP + 1} \text{ and } l_2 = \frac{A}{BP + 1} \tag{2}$$

A and *B* are empirical parameters dependent on the organic phase, geometry of the apparatus and the stirring rate, but not on the molecular structure of the transported drugs. A steady-state description of the process [6–8] shows that the empirical parameters in Eq. (2) can be defined as $A = k_{org}$ and $B = k_{org}/k_{aq}$. Here, k is the rate constant of diffusion in the diffusion layer indicated by the subscript that can be expressed using the diffusion coefficient D and the effective thickness h of the corresponding diffusion layer as $k = D/h$ [9]. Interestingly, Eq. (2) holds also for drugs that do not belong to an homologous series, or that may ionize or form ion pairs [10]. The significance for QSAR of the influence of molecular size on the rate permeation through liquid [11] and artificial membranes [12, 13] remains to be assessed.

17.2.2.2 Membrane Accumulation

The lipid core seems to have practically identical solvation properties in all membranes. The membrane/water partition coefficient P_M of nonamphiphilic drugs is, under certain conditions, related to the reference (usually 1-octanol/water) partition coefficient P as [14]:

$$\log P_M = \beta \log P + A \tag{3}$$

with A and β being empirical parameters. Eq. (3) is not invalidated by the observation that binding of some amphiphilic drugs to membrane does not correlate with the partition coefficient [15]. The partition coefficient is a measure of lipophilicity and describes well the accumulation in the lipid core of the membrane. For characterization of the binding to the membrane/water interfaces a measure of amphiphilicity is required.

17.2.2.3 Binding to Cell Constituents

Among the cell constituents, proteins – with their ability to take part in practically all types of weak and strong interactions – are the most probable candidates for association with drug molecules. Protein binding is usually reversible, the equilibrium being reached within milliseconds unless covalent bonds are formed. Proteins to which the drug is bound can be classified into three types: receptors, metabolizing enzymes, and the rest, which are sometimes called "silent receptors". Binding to the former two types represents structure-specific interactions exhibiting high affinity and limited capacity. However, the majority of the drug-binding proteins belong to the silent receptors, which are frequently globular proteins with lipophilic interiors. The averaged association constants K_i depend mostly on lipophilicity according to Eq. (3) with $\log K_i$ on the left-hand side [16]. For binding of various series of drugs to the same protein the slope β often remains identical for all series and A varies according to other than hydrophobic interactions involved in the binding of the parent molecule [17].

17.2.2.4 Enzymatic and Spontaneous Reactions

The covalent reactions (phase 1 and phase 2, for a review see [18]) constitute, together with excretion, the pathway for elimination of drug molecules from biological systems. In addition to the rate parameters measured in separate reaction mixtures imitating the

biological conditions, various empirical substituent constants and quantum chemical indices characterizing the electron density at the reaction center, as well as computed activation energies, can be used to substitute the rate parameters of biological reactions. In some cases the partition coefficient may be of significance, if the noncovalent binding to the enzyme depends on lipophilicity and all the other factors are constant [19].

17.3 Numerical Simulations

The results of numerical integration of the differential equations corresponding to the scheme in Fig. 1 at a fixed time after the drug administration have usually been presented graphically as concentration-lipophilicity profiles. The resulting curves were often described by an empirical equation that could be used for fitting experimental data.

Hansch and Fujita [20] assumed intuitively that the probability of the occurrence of the drug molecules inside the biological system after a predetermined time interval follows a Gaussian distribution with respect to lipophilicity. Nevertheless, the final solution – the parabolic dependence of the biological activity or intracellular drug concentration on lipophilicity (Eq. (4)) – proved to be quite robust and has been used by many subsequent authors:

$$\log c \text{ or } \log \frac{1}{c_x} = A(\log P)^2 + B \log P + C. \tag{4}$$

Here, c is the actual drug concentration, c_x is the equipotent concentration eliciting the effect representing the pre-defined fraction X $(0 \le X \le 1)$ of the maximum effect, and A, B, C are adjustable parameters optimized by linear regression analysis.

The first mathematical treatment of the drug distribution in multimembrane systems with a QSAR output was given by Penniston and coworkers [21]. As the experimentally verified dependence of the rate constants of transport on the partition coefficient (Eq. (2)) was not known that time, they used the assumption $l_1 l_2 = 1$. The observed convex lipophilicity-concentration dependence consisting of two linear parts connected by a curve was interpreted as confirmation of the validity of the parabolic model.

McFarland [22] used a probabilistic approach to describe the drug movement in the Penniston model and obtained symmetrical convex lipophilicity-concentration profiles with linear ascending and descending parts and the maximum at $\log P = 0$.

Kubinyi [23, 24] took into consideration the substantial difference in volumes of aqueous and lipid phases in biological systems. He also obtained symmetrical convex lipophilicity-concentration profiles, but without the restriction encountered by McFarland. Assuming that the receptor binding is also lipophilicity-dependent and making all the parameters A, B, C, D freely adjustable, he derived the so-called bilinear equation (Eq. (5) with $i = \beta = 1$). It has been shown to fit closely many equilibrium lipophilicity-concentration profiles [25] (cf. section 17.3.1.1) as well as those generated by numerical simulations of drug transport during the nonequilibrium period of distribution (cf. section 17.3.1.2) with the dependence of the transport rate constants on lipophilicity as given by Eq. (2) [23–25]. The bilinear equation can be therefore consid-

ered as the model-based description of the drug distribution. Its versatility makes it a valuable tool for the description of the relationships between biological activity and lipophilicity [26].

Dearden and Townend [27, 28] developed the model of Penniston further and drew attention to the importance of time in transport simulations and in QSAR generally. The relation between the time to maximal response and lipophilicity is among the most significant results.

Berner and Cooper [29] have been able to describe the drug partition in multimembrane systems using Fick's law for diffusion within membranes and aqueous phases and assuming instantaneous equilibration at their interfaces. This assumption is, in some sense, equivalent to Eq. (2); therefore, it is not surprising that they observed bilinear lipophilicity-concentration profiles.

In contrast to previous workers simulating only the first periods of distribution, we have examined the complete time-course of the drug transport in closed and open systems with identical lipid phases [30–32]. The lipophilicity-concentration profiles are different in individual time periods and can be described by an empirical equation [30] that could be considered as an extended version of the bilinear equation [23–26]:

$$\log c \text{ or } \log \left(\frac{1}{c_x}\right) = A \log P^\beta + \sum_{i=1}^{3} B_i \log \left(C_i P^\beta + 1\right) + D. \tag{5}$$

The connection between adjustable parameters A, B, C, D and the shape of the corresponding curves (for the exponent from Eq. (3), $\beta = 1$) is clear from Fig. 2. An example of the simulation results is presented in Fig. 3 for the closed system (a), for the open system with hydrolysis as the only elimination route (b), and with metabolism (c). The shapes of the lipophilicity-concentration profiles give an indication of the governing processes of the drug distribution and allow for prediction of its temporal development as is discussed in sections 17.3.1 and 17.3.2.
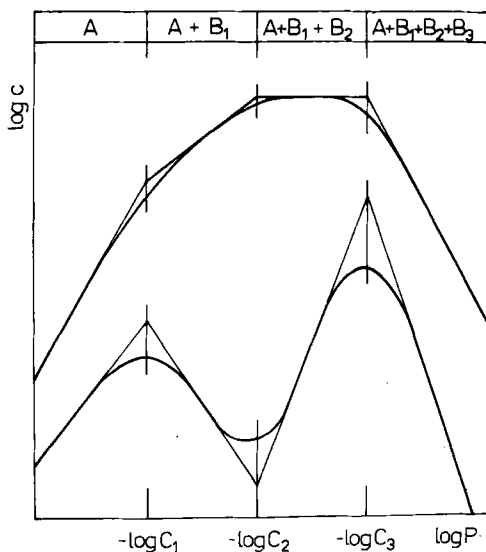
Figure 2. The dependence of the intracellular drug concentration c on the partition coefficient P as described by Eq. (5). The slopes of the linear parts, as dependent on the parameters A and B_i ($i = 1, 2, 3$) from Eq. (5), are given in the upper part. The positions of the curvatures are determined by the values of the parameters C_i ($i = 1, 2, 3$) from Eq. (5) [32]. (Reproduced with permission of Elsevier Science Publishers).
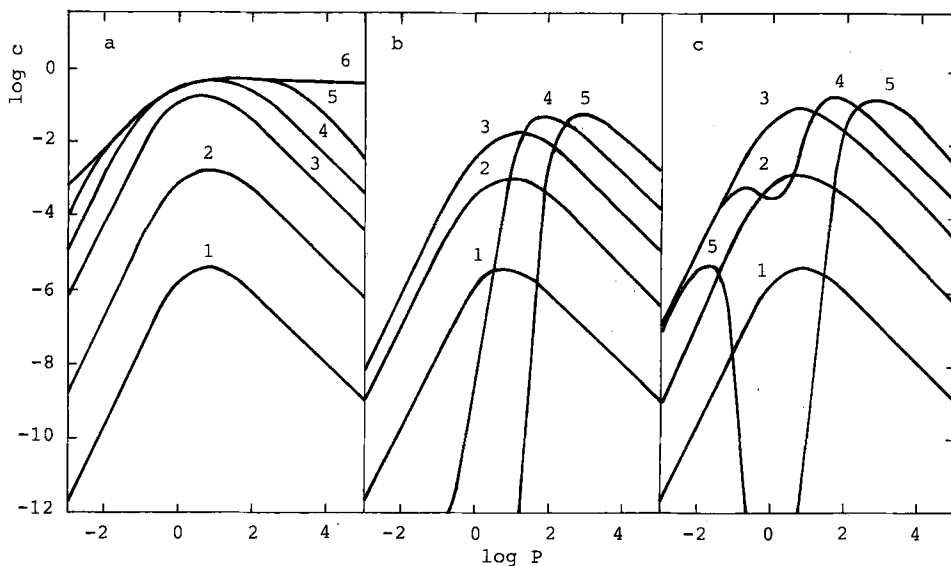
Figure 3. The lipophilicity-concentration profiles in the last compartment of a four-compartment system (a membrane) without elimination (a), with elimination from either both aqueous phases (b), or from the intracellular aqueous phase (c), after the following distribution periods (in time units): 0.1 (curve 1), 1 (2), 10 (3), 100 (4), 1000 (5) and ∞ (6) [32]. (Reproduced with permission of Elsevier Science Publishers.)

17.3.1 Closed Systems

Providing that the reversible processes (transport and protein binding) are much faster than the irreversible elimination (Fig. 1), a closed model (the results in Fig. 3a) is an appropriate representation of the *in vivo* situation for the time interval when the elimination is negligible. The total period of distribution can be subdivided into three parts: nonequilibrium period (A), mixed period (B) and equilibrium period (C). The borders between individual periods are represented by the moments when either the fastest drug (between the periods A and B) or all drugs (between B and C) attain the equilibrium.

17.3.1.1 Nonequilibrium Period

The bilinear lipophilicity-concentration profiles (Fig. 3a, curves 1–3) are simply shifted along the log *c*-axis with the increase in the distribution time. They are symmetrical for the aqueous compartments and asymmetrical for the membranes. The slopes of the linear parts (Table 1) are integers and are characteristic of the corresponding compartments [6, 25, 31]. This fact might significantly promote elucidation of mechanisms of the drug action, as the shape of the relationship between biological activity and lipophilicity could, under nonequilibrium conditions, indicate the sequential number and nature of the receptor compartment.

Table 1. The integer values of the slopes in linear parts of the lipophilicity-concentration profiles (numbered from the left) and the values of the parameters A, B_1, B_2, B_3 in corresponding Eq. (5) for the j-th compartment in nonequilibrium (A), mixed (B) and equilibrium period (C) of distribution [32]

Period	Phase	Slopes				Parameters			
		1	2	3	4	A	B_1	B_2	B_3
A	Aqueous	$(j-1)/2$	$(1-j)/2$	–	–	$(j-1)/2$	$1-j$	–	–
	Membrane	$j/2$	$1-j/2$	–	–	$j/2$	$1-j$	–	–
B	Aqueous	$(j-1)/2$	0	-1	$(1-j)/2$	$(j-1)/2$	$(1-j)/2$	-1	$3-j/2$
	Membrane	$j/2$	1	0	$1-j/2$	$j/2$	$1-j/2$	-1	$2-j/2$
C	Aqueous	0	-1	–	–	0	-1	–	–
	Membrane	1	0	–	–	1	-1	–	–

17.3.1.2 Equilibrium Period

In the closed system the transport of drugs will continue up to the achievement of the partitioning equilibrium (Fig. 3a, curve 6). If all the membranes are of identical composition, the concentration-lipophilicity dependences are again bilinear, with the slopes in the linear parts being equal (from left) to 1 and 0 for the membranes and to 0 and -1 for the aqueous phases. The slopes, however, have different values if the membranes differ in solvation properties [25]. Equilibrium models [33, 34] belong to the first nonempirical models of subcellular pharmacokinetics. They have been extended for ionization of the drug molecules by Martin [35, 36].

17.3.1.3 Mixed Period

In this period the fastest drugs with optimal partition coefficients have already reached equilibrium while the other drugs, which are either more lipophilic or more hydrophilic, have not. The lipophilicity-concentration profiles (Fig. 3a, curves 4 and 5) are a combination of the nonequilibrium (curves 1 and 3) and equilibrium dependences (curve 6) and consist of four linear parts with characteristic integer slopes (Table 1) connected by the curved portions. Eq. (5) with $i = 2$ or 3 is suitable for the description of the curves. They are not frequently observed experimentally because the range of lipophilicity of the tested drugs in usually not sufficiently wide. An example of mixed distribution has been found in the growth inhibition of several fungal and bacterial strains by alkyl amines with 4–18 carbons [37]. The data for most fungal strains (in contrast to the bacteria where the drug distribution is faster) exhibit small but systematic deviations from the bilinear equation [4], as illustrated in Fig. 4.

17.3.2 Open Systems

Metabolism has a pronounced effect on drug disposition. The phenomenon was first described by Dearden and Townend [27], who simulated a double-peaked dependence
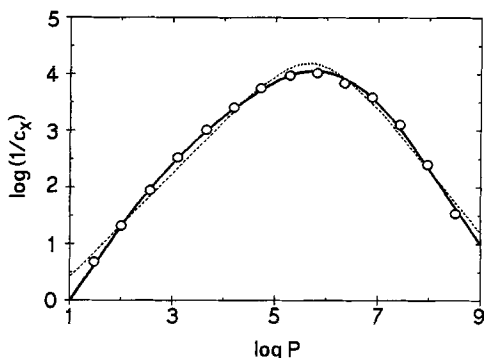
Figure 4. Growth inhibition of *Ctenomyces mentagrophytes* by alkylamines (c_x is the minimum inhibitory concentration [37]) as dependent on lipophilicity [4]. The solid line corresponds to Eq. (5) for mixed period of distribution ($i = 2$) and dotted line to the bilinear equation (Eq. 5 with $i = \beta = 1$).

of the drug concentration on lipophilicity. In the initial period of distribution, corresponding to the nonequilibrium period in the closed system (cf. Fig. 3a, curves 1–3), the curves are bilinear (Figs. 3b and 3c, curves 1–3) with the characteristic integer values of the slopes (Table 1). The influence of metabolism becomes observable at the time when the mixed period of distribution starts in the corresponding closed system (cf. Fig. 3a, curves 4 and 5). The shapes of lipophilicity-concentration profiles depend on the compartment where the metabolism is encountered. For reactions proceeding in all the aqueous compartments, including the extracellular medium (e.g., hydrolysis), the lipophilicity-concentration profiles are deformed in the region of low lipophilicity (Fig. 3b, curves 4 and 5). If only intracellular metabolism is considered, the maximum of the curve is distorted (Fig. 3c, curves 4 and 5). The lipophilicity-concentration profiles have two maxima separated by a minimum and can be described by Eq. (5) with $i = 3$. The drugs with log P values from the distorted regions (log $P < 2$ and $-2 <$ log $P < 2$ for the conditions in Figs. 3b and 3c, respectively) will exhibit much faster elimination in the further course of the distribution than other compounds.

17.4 Explicit Descriptions

The differential equations describing the distribution scheme in Fig. 1 can be solved explicitly only for reduced scenarios like unidirectional transport combined with elimination and protein binding [38], or pure transport in the water/membrane/water system [39]. Due to the time hierarchy of the involved processes (transport is much faster than metabolism), a simplified description for the elimination period of distribution can be obained if transport is considered as instantaneous. The time-course of distribution or biological activity can be described as [40]:

$$\frac{c_A}{c_0} \text{ or } \frac{X}{Kc_x(1-x)} = \frac{1}{AP^\beta + B} \exp\left(-\frac{CP^\beta + D}{AP^\beta + B}t\right). \qquad (6)$$

Here, c_A/c_0 is the ratio of the actual and initial concentration of free and nonionized drug molecules, K is the drug-receptor association constant, and t is the time of expo-
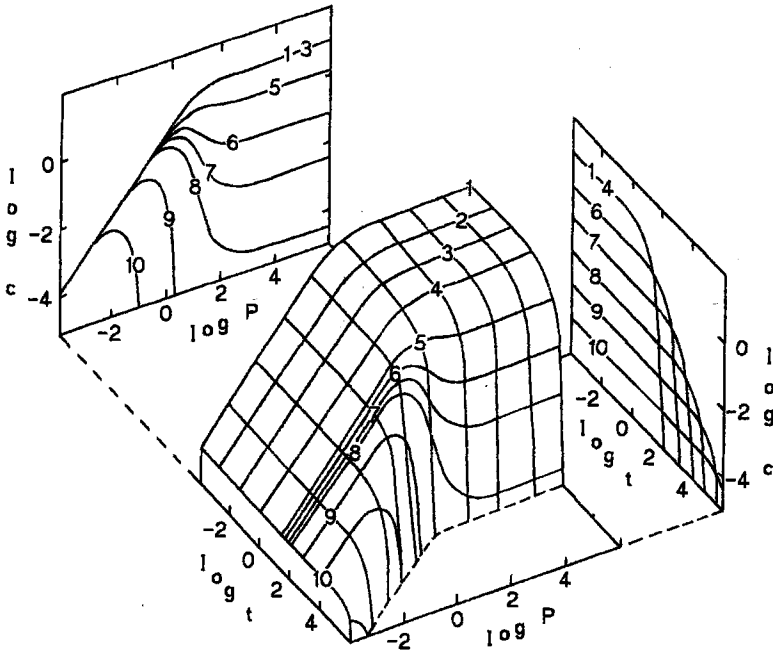
Figure 5. Relationship between the concentration c of the protein-bound drug, the partition coefficient P, and the time of distribution t [35]. The values were calculated from Eq. (6) multiplied by P and $\beta = c_0 = A = 1$, $B = 0.1$, $C = 0.01$, and $D = 0$. (Reproduced with permission of the American Pharmaceutical Association).

sure. The terms A, B, C, D describe the extent of individual processes the drugs undergo in the biological system: A – membrane accumulation and protein binding, B – distribution in aqueous phases which may differ in their pH values, C – lipophilicity-dependent enzymatic metabolism, and D – other first order elimination processes. The parameter β is the exponent from the Collander equation [14] (Eq. (3)).

17.4.1 Nonionizable Compounds

For nonionizable compounds the terms A, B, C, D are constant and can be optimized by nonlinear regression analysis of experimental data according to Eq. (6). The behavior of Eq. (6) – the concentration of the drug bound to the receptors – in the space (log P, t) is depicted in Fig. 5. It is assumed that the compounds are metabolized solely by lipophilicity-dependent enzymatic reactions. The concentration-lipophilicity profiles (the left-hand projection) have first the equilibrium shape (curves 1–4) and later become distorted in the region of high lipophilicity. This is caused by lipophilicity-dependent metabolism, which does not affect significantly hydrophilic compounds. An example of the trilinear dependence of biological activity on lipophilicity similar to that depicted in Fig. 5 is given in Fig. 6.
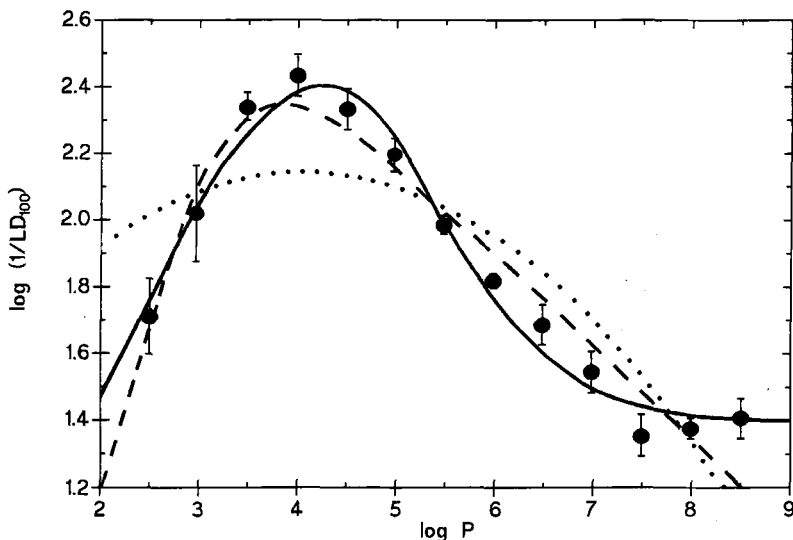
Figure 6. Toxicity of alkanes in mice (LD_{100} in mol kg^{-1}) versus lipophilicity fitted with Eq. (6) with $D = 0$ [40]. The solid line corresponds to Eq. (6), the dashed line to the bilinear equation (Eq. (5) with i = β = 1) and the dotted line to Eq. (4). (Reproduced with permission of the American Pharmaceutical Association).

17.4.2 Ionizable Compounds

In this case individual terms A, B, C, D (described as Y) in Eq. (6) can be expanded as [41]:

$$Y = Y_0 + Y_1/K_a^{sgn} \tag{7}$$

where sgn $= -11$ for acids, and sgn $= 1$ for bases (like the signs of the resulting ions). The terms Y_0 and Y_1 comprise the quantities associated with non-ionized and ionized molecules, respectively. If acidity of individual aqueous compartments is invariant during the experiment, Y_0 and Y_1 are adjustable parameters optimized by nonlinear regression analysis of experimental data according to Eq. (6) as combined with Eq. (7). The approach has been applied to description of the growth inhibition of *Sarcina lutea* by a series of lincomycin derivatives after a fixed exposure time [36]. The most suitable form of Eq. (6) as combined with Eq. (7) for fitting the data proved to be

$$\log(1/c_x) = -\log(A_0P + B_0 + B_1/K_a) - \frac{D_0 + D_1/K_a}{A_0P + B_0 + B_1/K_a} + EI_{trans} \tag{8}$$

The missing parameters were set to $\beta = 1$ and $A_1 = C_0 = C_1 = 0$. The parameter E accounts for the difference in activity of *cis* and *trans* isomers. The fit of Eq. (8) to the data [41] is satisfactory ($n = 31$, $r = 0.984$, sd $= 0.125$, $F = 103.2$).

17.4.3 Varying Acidity of the External Medium

QSAR analysis of the data measured under conditions of varying pH of the external medium can contribute significantly to the elucidation of action mechanisms of ionizable compounds. The buffering capacity of the intracellular aqueous compartments depends heavily on the metabolic and physiologic state of the cells. The influence of the acidity of the external medium on the acidity of the internal phases can be approximated as [42]:

$$pH_i = \vartheta pH_e + \eta_i \tag{9}$$

where the sensitivity ϑ is assumed to be constant for all the intracellular phases. Then all the terms A, B, C, D in Eq. (6) can be described by Eq. (7) with the terms for ionizable compounds, Y_I, being expanded as [42]:

$$Y_1 = Y_{1E} 10^{pH_e} + Y_{1I} 10^{\vartheta pH_e} \tag{10}$$

Here the subscripts E and I refer to the adjustable parameters associated with the external medium and with intracellular phases, respectively. The correlation of experimental data according to Eq. (6) as combined with Eqs. (7) and (10) indicates the influence of ionization on the terms A, B, C, D (cf. the text below Eq. (6)). The approach has been applied to the growth-inhibitory effects of α-bromo-alkanoic acids against *Vibrio cholerae* and other microorganisms in the media with varying pH values [42].

17.5 Outlook

The comparison of the results of numerical simulations with experimental data requires the search for empirical functions like Eqs. (4) and (5). With the growing number of independent variables (lipophilicity, acidity, reactivity, the exposure period) understanding and presentation of the simulation results and formulation of suitable empirical functions become more and more difficult. Therefore, it can be expected that explicit descriptions will be preferred in the future.

Further attempts in this area could extend the existing models for: 1) other time-courses of the drug input (repeated, continuous); 2) multiple ionization; 3) multiplication of the cells for the description of the growth experiments; and 4) structure-specific interactions of the chemicals with metabolizing enzymes and the receptors.

Considering the reviewed results it can be concluded that the methods of subcellular pharmacokinetics may contribute significantly to our understanding of quantitative dependences of biological activity on physico-chemical properties of the drugs, especially on lipophilicity.

References

[1] Pliška, V., *Arzneim.-Forsch.* **16,** 886–893 (1966)

[2] Pliška, V., *Il Farmaco Ed. Sci.* **23,** 623–641 (1968)

[3] Baláž, Š., Šturdík, E., and Ticháy, M., *Quant. Struct.-Act. Relat.* **4,** 77–81 (1985)

[4] Baláž, Š., *Quant. Struct.-Act. Relat.* **13,** 381–392 (1994)

[5] Kubinyi, H., *J. Pharm. Sci.* **67,** 262–263 (1978)

[6] Van de Waterbeemd, J. T. M., Jansen, A. C. A., and Gerritsma, K. W., *Pharm. Weekblad* **113,** 1097–1105 (1978)

[7] Van de Waterbeemd, H., van Boeckel, S., Jansen, A., and Gerritsma, K., *Eur. J. Med. Chem.* **15,** 279–282 (1980)

[8] 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 (1983)

[9] Van de Waterbeemd, J. T. M., and Jansen, A. C. A., *Pharm. Weekblad Sci. Ed.* **3,** 71–78 (1981)

[10] Van de Waterbeemd, H., van Bakel, H., and Jansen, A., *J. Pharm. Sci.* **70,** 1081–1082 (1981)

[11] Leahy, D. E., de Meere, A. L. J., Wait, A. R., Taylor, P. J., Tomenson, J. A., and Tomlinson, E., A general model relating water-oil partitioning rates to physicochemical structure. In: *QSAR in Drug Design and Toxicology,* Hadži, D., and Jerman-Blažiúc, B. eds., Elsevier, Amsteradam (1987) p. 144–146

[12] Walter, A., and Gutknecht, J., *J. Membrane Biol.* **90,** 207–217 (1986)

[13] Xiang, T. X., and Anderson, B. D., *J. Membrane Biol.* **140,** 111–122 (1994)

[14] Collander, R., *Acta Chem. Scand.* **5,** 774–780 (1951)

[15] Mason, R. P., Rhodes, D. G., and Herbette, L. G., *J. Med. Chem.* **34,** 869–877 (1991)

[16] Helmer, F., Kiehs, K., and Hansch, C., *Biochemistry* **7,** 2858–2863 (1968)

[17] Austel, V., and Kutter, E., Absorption, distribution, and metabolism of drugs. In: *Quantitative Structure-Activity Relationships of Drugs,* Topliss, J. G., ed., Academic Press, New York (1983) p. 437–496

[18] Testa, B., and Jenner, P., *Drug Metabolism: Chemical and Biochemical Aspects.* Marcel Dekker: New York, 1976

[19] Hansch, C., Lien, E. J., and Helmer, F., *Arch. Biochem. Biophys.* **12,** 319–330 (1968)

[20] Hansch, C., and Fujita, T., *J. Am. Chem. Soc.* **86,** 1616–1626 (1964)

[21] Penniston, J. T., Beckett, L., Bentley, O. L., and Hansch, C., *Mol. Pharmacol.* **5,** 333–341 (1969)

[22] McFarland, J. W., *J. Med. Chem.* **13,** 1192–1196 (1970)

[23] Kubinyi, H., *Arzneim.-Forsch.* **26,** 1991–1997 (1976)

[24] Kubinyi, H., *J. Med. Chem.* **20,** 626–629 (1977)

[25] Kubinyi, H., *Prog. Drug Res.* **23,** 97–198 (1979)

[26] Kubinyi, H., *QSAR: Hansch Analysis and Related Approaches.* VCH: Weinheim, 1993

[27] Dearden, J. C., and Townend, M. S., Digital computer simulation of the drug transport process. In: *Quantitative Structure-Activity Analysis.* Franke, R. and Oehme, P. (Eds.). Akademie-Verlag: Berlin; 387–393 (1978)

[28] Dearden, J. C., Molecular structure and drug transport. In: *Comprehensive Medicinal Chemistry,* Vol. **4.** Hansch, C., Sammes, P. G., and Taylor, J. B. (Eds.). Pergamon Press: Oxford; 375–411 (1990)

[29] Berner, B., and Cooper, E. R., *J. Pharm. Sci.* **73,** 102–106 (1984)

[30] Baláž, Š., Šturdík, E., Hrmová, M., Breza, M., and Liptaj, T., *Eur. J. Med. Chem.* **19,** 167–171 (1984)

[31] Baláž, Š., and Šturdík, E., Passive transport and lipophilicity. Closed model of drug distribution. In: *QSAR in Design of Bioactive Compounds.* Kuchaúr, M. (Ed.). Prous: Barcelona; 289–300 (1984)

[32] Baláž, Š., and Šturdík, E., Lipophilicity and drug disposition. In: *QSAR in Toxicology and Xenobiochemistry.* Tichý, M. (Ed.). Elsevier: Amsterdam; 257–267 (1985)

[33] Higuchi, T., and Davis, S. S., *J. Pharm. Sci.* **59,** 1376–1383 (1970)

[34] Hyde, R. M., *J. Med. Chem.* **18,** 231–233 (1975)

[35] Martin, Y. C., and Hackbarth, J. J., *J. Med. Chem.* **19,** 1033–1039 (1976)

[36] Martin, Y. C., *Quantitative Drug Design.* Marcel Dekker: New York, 1978

[37] Koelzer, P., and Büchi, J., *Arzneim.-Forsch.* **21,** 1721–1727 (1971)

[38] Baláž, Š., Šturdík, E., and Augustín, J., *Bull. Math. Biol.* **50,** 367–378 (1988)

[39] Baláž, Š., and Šturdík, E., *Gen. Physiol. Biophys.* **4,** 105–108 (1985)

[40] Baláž, Š., Wiese, M., and Seydel, J. K., *J. Pharm. Sci.* **81,** 849–857 (1992)

[41] Piršelová, K., and Baláž, Š., *Chemometrics Intell. Lab. Syst.* **24,** 193–196 (1994)

[42] Baláž, Š., Cronin, M. T. D., and Dearden, J. C., *Pharm. Sci. Commun.* **4,** 51–58 (1993)

18 Applications of a Solvation Equation to Drug Transport Properties

Michael H. Abraham and Harpreet S. Chadha

Abbreviations

BB	Blood-brain
GLC	gas-liquid chromatography
HPLC	High-performance liquid chromatography
LFER	Linear free energy relationship
MLR	Multiple linear regression
PC	Principal component
QSAR	Quantitative structure-activity relationship
RBR	Relative biological response
sd	Standard deviation of a regression
SP	Hydrophilic surface

Symbols

E	An electronic parameter
F	Fisher F-statistic
H	Superscript, a descriptor for use in the general LFER
k'	HPLC capacity factor
k'_w	HPLC capacity factor extrapolated to 100 % water
n	Number of points (compounds) in a correlation
pK_{HB}	Taft's hydrogen-bond basicity parameter
P	Water/solvent partition coefficient
P_{cyc}	Water/cyclohexane partition coefficient
P_{oct}	Water/octanol partition coefficient
P_{16}	Water/hexadecane partition coefficient
r	Correlation coefficient
R_2	Excess molar refraction
S	A steric parameter
V_m	Molar volume
V_x	McGowan's characteristic volume
α_2^H	Solute 1:1 hydrogen-bond acidity
β_2^H	Solute 1:1 hydrogen-bond basicity
$\Delta \log P$	Seiler's delta log P parameter
π_2^H	Solute dipolarity/polarizability
Λ_{cyc}	Interactive polar parameter, from water/cyclohexane partition coefficients
σ	Hammett's substituent constant
$\sum \alpha_2^H$	Solute overall hydrogen-bond acidity

$\sum\beta_2^{\mathrm{H}}$ Solute overall hydrogen-bond basicity
$\sum\beta_2^{\mathrm{O}}$ Solute special overall hydrogen-bond basicity

18.1 Introduction

To date, quantitative structure-activity relationships (QSARs) have mostly been constructed using [1] the Hansch equation,

$$\log \mathrm{RBR} = -a(\log P_{\mathrm{oct}})^2 + b \log P_{\mathrm{oct}} + cE + dS + e \tag{1}$$

where RBR is a relative biological response of a series of substrates, or solutes, in a given system. The descriptors are $\log P_{\mathrm{oct}}$ where P_{oct} is the water/1-octanol partition coefficient, E is an electronic parameter such as Hammett's σ-value, and S is a steric term. Indeed, even today, Eq. (1) is still used to interpret and to predict biological activities [2], especially for a series of structurally related substrates. There is, however, an insuperable difficulty in the application of Eq. (1) when the substrate series includes quite different structures, because it is then no longer possible to define a simple electronic parameter, or a simple steric parameter. Yet there are numerous very important areas in which substrates of all kinds of parent structures are involved. These include aqueous anesthesia, the toxicity of aqueous compounds to various organisms, and the partition of substrates between phases. Young et al. [3], for example, studied the blood-brain distribution of a large number of structurally diverse drugs, where the original Hansch equation cannot be used, because there is no relevant electronic or steric parameter available across the series of compounds. Of course, if Eq. (1) is modified by the deletion of the electronic and steric terms, it can be applied to a structurally diverse series of compounds, but much information is then lost. In any case, an equation in just $\log P_{\mathrm{oct}}$, or a two-term equation in $(\log P_{\mathrm{oct}})^2$ and $\log P_{\mathrm{oct}}$ can hardly be regarded as a Hansch equation, in the spirit of Eq. (1). This problem over electronic and steric parameters in a structurally varied series of compounds is quite general, and can only be solved by the construction of other types of QSAR, that contain radically different descriptors to those used in the Hansch equation. The approach of the present work has been to define and to evaluate other descriptors that can be employed in a new generation of QSARs and linear free energy relationships (LFERs).

In order to set out possible solute descriptors for processes that involve solvation, some model of the solvation process itself is needed. A simple cavity model has often been used to describe solvation [4, 5], as noted by Zahradník and Hobza in Chapter 3. The solvation of a gaseous molecule is broken down into stges as follows:

1. A cavity of suitable size is made in the solvent, with the solvent molecules in the same state as in the bulk solvent (Fig. 1, left panel)
2. The solvent molecules are reorganized into their equilibrium position round the solute (Fig. 1, center panel)
3. The solute is inserted into the cavity and various solute-solvent interactions are set up (Fig. 1, right panel)
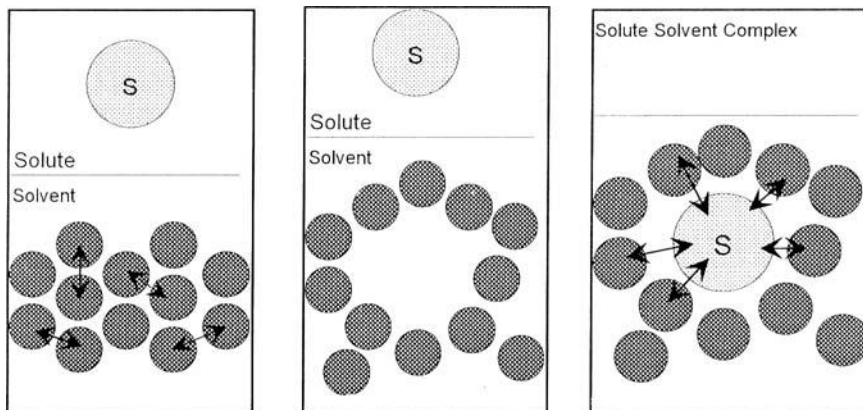
Figure 1. The cavity model of solvation.

The creation of a cavity requires work, and is an endoergic process. If the solvent is constant, and cavities of various sizes are created, the Gibbs energy of cavity formation can be taken as proportional to the solute volume or to the solute surface area.

The Gibbs energy of reorganization can be taken as zero, and so need be considered no further. It should be pointed out, however, that the enthalpy and entropy of reorganization may be considerable – the Gibbs energy is effectively zero because of compensation, as in the melting of ice at 273 K.

Finally, there will be a number of solute-solvent interaction terms, all of which will be energetically favorable, that is, exoergic. What are now required are solute parameters or descriptors that can be identified with these interaction terms, again for the case in which the solvent is constant and the solute is varied.

The most general solute-solvent interaction involves dispersion (or London) forces. A suitable solute descriptor might be the molar refraction, but this is too well related to solute volume to be an independent quantity. Indeed, molar refraction is so well correlated with volume, that if the solute volume is used as a descriptor for the cavity effect, it will also be a descriptor for London forces. There is no simple way to separate these, and so volume is a combined descriptor for the endoergic cavity effect, and the exoergic solute-solvent London dispersion forces.

Another type of solute-solvent interaction of very considerable importance indeed, will involve hydrogen-bonding of the type solute acid-solvent base, and solute base-solvent acid. Until quite recently, no solute parameters that could be used to provide quantitative descriptors of hydrogen-bonding were available. Taft et al. [6] had put forward the pK_{HB} scale of solute hydrogen-bond basicity as long ago as 1969, but the lack of a defined origin had prevented any general use in QSARs or in LFERs.

A useful starting point was the construction of scales of solute hydrogen-bond acidity, α_2^H, and hydrogen-bond basicity, β_2^H, using $1:1$ complexation constants in tetrachloromethane. Values of α_2^H and β_2^H were obtained for numerous monofunctional solutes [7, 8], and the scales were subsequently extended by the group of Laurence and Berthelot [9–11]. It was recognized, however, that such scales were not adequate for

the situation in which a solute molecule was surrounded by an excess of solvent molecules [12] (see Fig. 1). What were then required were scales that described the "overall" or "effective" hydrogen-bond acidity or basicity of a solute molecule. In order to construct such scales, it was first necessary to develop a general LFER that included not only hydrogen-bond descriptors, but also the other descriptors necessary for an analysis of physico-chemical and biochemical phenomena [12].

A list of descriptors had thus to be drawn up, bearing in mind that for use in a multiple linear regression equation, the various descriptors should be as orthogonal as possible. In order to separate out (at least to some extent) the general dispersion interaction from the volume term, an excess molar refraction R_2, was defined [13] as the compound molar refraction less that of an alkane of the same characteristic volume. R_2 thus defined has the advantage that it is almost independent of volume. Values of R_2 can be obtained experimentally for liquids from the refractive index at 293 K. Like the molar refraction itself, R_2 is an almost additive quantity which can easily be estimated for solids, and for structures in general, from fragment or substructure values.

The final descriptor must take care of interactions of the dipole-dipole and dipole-induced dipole type. Attempts were first made [13] to use the solute dipole moment, as such or squared, as a descriptor, but this was found to be unsatisfactory, and so a new descriptor, π_2^H, the solute dipolarity/polarizability, was developed [14]. This was first obtained through gas-liquid chromatographic (GLC) measurements [14–16], but is now more generally obtained from water/solvent partition coefficients [17], as will be shown later.

The various solute descriptors were combined linearly to give a general equation,

$$\log SP = c + rR_2 + s\pi_2^H + a\sum \alpha_2^H + b \sum \beta_2^H + vV_x \tag{2}$$

In this equation, SP is some property of a series of solutes in a given solvent system, and the explanatory variables, or descriptors, are solute properties as follows. R_2 is the excess molar refraction, π_2^H is the solute dipolarity/polarizability, $\sum \alpha_2^H$ is the solute overall or effective hydrogen-bond acidity, $\sum \beta_2^H$ is the solute overall or effective hydrogen-bond basicity, and V_x is the McGowan characteristic volume that can be calculated for any solute simply from molecular structure, using a table of atomic and bond constants [18], and the algorithm of Abraham [12] for the number of bonds in any molecule. Note that we use subscript "2" to denote a solute, and superscript "H" to indicate that this descriptor is for use in the general hydrogen-bond equation (Eq. (2)).

The general approach leading to Eq. (2) is exactly the same as that used by the group of Abboud, Abraham, Doherty, Kamlet and Taft [4, 19, 20], although the definition of the descriptors is quite different. Indeed, Eq. (2) contains not a single descriptor that is used in the "solvatochromic" equation of Kamlet et al. [20]. It is important to note that several of the solute descriptors in the "solvatochromic" equation were derived from solvent properties [20] whereas the corresponding terms in Eq. (2), π_2^H, $\sum \alpha_2^H$ and $\sum \beta_2^H$ are all rigorously solute properties. Taylor et al. [21, 22] have also used an LFER constructed from solute properties to investigate partitioning processes, although their solute descriptors differ in several ways from those in Eq. (2).

A preliminary version of Eq. (2) was set up using α_2^H and β_2^H as the hydrogen-bond descriptors, and was applied to various water/solvent partitions [17]. The α_2^H and β_2^H descriptors were then modified where necessary, in order to obtain the effective $\sum \alpha_2^H$

and $\sum \beta_2^H$ values [17]. A new set of equations was then constructed, and the process repeated until a self-consistent set of equations and $\sum \alpha_2^H$ and $\sum \beta_2^H$ values was obtained. Since the solutes in water/solvent partitions are surrounded by solvent molecules, the $\sum \alpha_2^H$ and $\sum \beta_2^H$ scales of hydrogen-bonding will, indeed, be those required. It was observed that values of $\sum \alpha_2^H$ were constant along any homologous series, except perhaps for the first one or two members, and so once a few values have been established, it is easy to deduce values for the rest of the homologous series. This was also the case for homologous series of bases.

The first four descriptors in Eq. (2) can be regarded as measures of the propensity of a solute to undergo various solute-solvent interactions, all of which are energetically favorable, i.e, exoergic. The V_x descriptor models the (endoergic) cavity effect that arises through the disruption of solvent-solvent interactions when a solute is placed in a solvent, together with the (exoergic) solute-solvent general (London) dispersion interactions. Note that solute size is closely related to molar refraction [23]. Because the descriptors in Eq. (2) refer to rather specific interactions, the coefficients in Eq. (2) will contain information on the particular solvent phase in question. The r-coefficient, is a measure of the phase polarizability, the s-coefficient measures the phase dipolarity/ polarizability, the a-coefficient is a measure of the phase hydrogen-bond basicity (because an acidic solute will interact with a basic phase), and the b-coefficient is a measure of the phase acidity. The v-coefficient is a measure of the phase hydrophobicity. Of course if Eq. (2) is applied to distribution between two phases, the coefficients will then refer to differences between the phases concerned.

A problem over variable solute hydrogen-bond basicity was first noticed by Taylor et al. [21]. They showed that for certain solutes, $\sum \beta_2$ as calculated from equations on the same lines as Eq. (2) was not constant, but varied with the partitioning system. This was later confirmed by Abraham [17] who managed to incorporate these solutes into the usual system by defining an additional basicity parameter, $\sum \beta_2^O$. This is used instead of $\sum \beta_2^H$ for solutes such as sulfoxides (but not sulfones or sulfonamides), anilines, pyridine and alkyl pyridines, and some heterocyclic compounds in water/solvent partitioning systems where the organic phase is quite aqeous. The latter include octanol, ethyl acetate, butyl acetate, diethyl ether, and dibutyl ether. For nonaqueous phases such as chloroform, alkanes, benzene, and the gas phase, the original $\sum \beta_2^H$ descriptor can be used for all solutes.

18.2 The Determination of Descriptors

The R_2 descriptor can be obtained very simply by additivity of fragments, and the V_x descriptor can be calculated from atomic constants as shown by McGowan [18] and by Abraham [12]. There are therefore three other descriptors that have to be determined by experiment, π_2^H, $\sum \alpha_2^H$, and $\sum \beta_2^H$. For volatile solutes, the π_2^H descriptor can be obtained by GLC on a polar, nonacidic, stationary phase [14–16]. Some values out of the 700 determined in this way are in Table 1, where n is the number of separate determinations and sd is the standard deviation in π_2^H. It can be seen that quite large molecules, such as pyrene, can be investigated in this way, and that values of π_2^H can be obtained to around 0.03 units. Some values of π_2^H for more complicated solutes, obtained

Table 1. Some values of π_2^H for solutes

Solute	π_2^H	n^a	sd^b
Cyclohexane	0.10	13	0.03
Biphenyl	0.99	5	0.05
Pyrene	1.76	3	0.03
Fluorobenzene	0.57	32	0.02
Chlorobenzene	0.65	36	0.01
Hexachlorobenzene	0.99	4	0.01
Bromobenzene	0.73	30	0.02
Benzaldehyde	1.00	22	0.01
Benzonitrile	1.11	21	0.02
Aniline	0.96	29	0.02
4-Methoxyaniline	1.19	5	0.04
4-Nitroaniline	1.91	4	0.02
Phenol	0.89	46	0.02
4-Methoxyphenol	1.17	9	0.01
4-Cyanophenol	1.63	3	0.01
4-Nitrophenol	1.72	5	0.03
Pyridine	0.84	16	0.02
Indole	1.12	7	0.05
Carbazole	1.42	4	0.03
Benzo[*b*]thiophene	0.88	2	0.04
Cocaine	1.92[c]		
Caffeine	1.60		
Morphine	2.35		
Codeine	2.40		
Heroin	3.05		
Cimetidine	1.73		
Icotidine	3.30		
Lupitidine	3.39		
Clonidine	1.83		
Mepyramine	1.92		
Imipramine	1.75		
Ranitidine	1.63		
Tiotidine	1.98		
Zolantidine	2.64		
Temelastine	3.24		
Progesterone	3.29		
Deoxycorticosterone	3.39		
Testosterone	2.59		
Corticosterone	3.43		
Cortisone	3.50		
Hydrocortisone	3.49		
Estrone	3.10		
Estradiol	3.30		

[a] Number of determinations.
[b] Standard deviation.
[c] From water/solvent partition coefficients.

from water/solvent partitions, are also given in Table 1 for comparison. For volatile solutes, the GLC method can in principle be used to obtain $\sum \alpha_2^H$ values through measurements on highly basic stationary phases, and $\sum \beta_2^H$ values via measurements on highly acidic phases. Some $\sum \alpha_2^H$ and $\sum \beta_2^H$ values have been determined in this way, but the use of partition coefficients for various water/solvent systems is usually better.

Our general procedure for the determination of descriptors, is now through the use of water/solvent partitions. The method is based on the construction of LFERs using Eq. (2), where log SP is log P, a partition coefficient in a given system. For example, the water/cyclohexane partition coefficient, as log P_{cyc}, can be correlated with the descriptors in Eq. (2) to yield [24],

$$\log P_{cyc} = (0.13 \pm 0.03) + (0.82 \pm 0.04)R_2 - (1.73 \pm 0.04)\pi_2^H -$$
$$(3.78 \pm 0.04)\sum \alpha_2^H - (4.91 \pm 0.06)\sum \beta_2^H + (4.65 \pm 0.05)V_x \quad (3)$$

$$n = 170, \ r^2 = 0.994, \ sd = 0.131, \ F = 5123$$

Here, and elsewhere, n is the number of data points, r is the correlation coefficient, sd is the standard deviation of the regression, and F is the F-statistic. The sd values of the coefficients are given in the parentheses. The correlation matrix for the descriptors in Eq. (3), in terms of r^2, shows that the maximum cross-correlation is between R_2 and π_2^H,

	R_2	π_2^H	$\sum \alpha_2^H$	$\sum \beta_2^H$
π_2^H	0.545			
$\sum \alpha_2^H$	0.158	0.116		
$\sum \beta_2^H$	0.001	0.145	0.005	
V_x	0.192	0.265	0.016	0.241

However, deletion of the R_2 descriptor leads to a poorer correlation ($r^2 = 0.979$, sd $= 0.241, F = 1874$) and deletion of the π_2^H descriptor while retaining the R_2 descriptor leads to a very much poorer correlation ($r^2 = 0.933$, sd $= 0.424$, F $= 578$). Whether the five descriptors in Eq. (3) contain any redundancy can be examined by a PC analysis of the 170 sets of descriptors. The cumulative proportions are, with the number of PCs in parentheses: 0.495(1), 0.748(2), 0.876(3), 0.966(4) and 1.000(5) so that at least four and possibly all five PCs are required unless significant information is lost.

Similar equations to Eq. (3) can be constructed for the correlation of numerous water/solvent partitions [17, 24, 25]. The descriptors R_2 and V_x can always be calculated, and so π_2^H, $\sum \alpha_2^H$ and $\sum \beta_2^H$ remain to be determined. In principle, log P values in three different water/solvent systems for a given solute could be used to calculate the three unknown descriptors through three simultaneous equations. But in practice, this method will only work if the three log P equations have quite different coefficients in the three descriptor terms. The preferred method [17] is to use log P values for as many systems as possible, and then to calculate the set of descriptors that best describes the log P values. For a solute that is not acidic, the calculation is quite rapid, because only

Table 2. Regression constants for partition from water[a]

$$\log P = c + rR_2 + s\pi_2^H + a\sum\alpha_2^H + b\sum\beta_2^H + vV_x$$

Phase	c	r	s	a	b	v
Isobutanol	0.227	0.514	−0.693	0.020	−2.258	2.776
	0.107	0.133	0.104	0.107	0.140	0.080
Pentanol	0.175	0.575	−0.787	0.020	−2.837	3.249
	0.167	0.129	0.129	0.098	0.181	0.152
Hexanol	0.143	0.718	−0.980	0.145	−3.214	3.403
	0.155	0.169	0.103	0.098	0.147	0.169
Octanol	0.088	0.562	−1.054	0.034	−3.460	3.814
	0.015	0.014	0.021	0.021	0.026	0.015
Decanol	0.008	0.485	−0.974	0.015	−3.798	3.945
	0.049	0.106	0.103	0.081	0.095	0.119
Oleyl alcohol	−0.359	−0.270	−0.528	−0.035	−4.042	4.204
	0.089	0.091	0.087	0.085	0.140	0.065
Diethyl ether	0.462	0.571	−1.035	−0.024	−5.508	4.346
	0.071	0.115	0.098	0.090	0.122	0.093
Ethyl acetate	0.253	1.157	−1.397	−0.054	−3.755	3.736
	0.082	1.101	0.125	0.102	0.117	0.100
n–Butyl acetate	−0.468	0.712	−0.397	0.010	−3.743	3.865
	0.179	0.107	0.151	0.106	0.212	0.136
Olive oil	−0.035	0.574	−0.798	−1.422	−4.984	4.210
	0.033	0.058	0.061	0.061	0.063	0.044
PGDP	0.287	0.338	−0.638	−0.908	−5.038	4.093
	0.129	0.103	0.159	0.133	0.161	0.153
CH_2Cl_2	0.326	0.097	−0.037	−3.312	−4.128	4.252
	0.088	0.138	0.128	0.127	0.124	0.107
$CHCl_3$	0.125	0.118	−0.372	−3.390	−3.467	4.521
	0.037	0.046	0.042	0.043	0.065	0.048
CCl_4	0.223	0.564	−1.151	−3.510	−4.536	4.501
	0.029	0.038	0.043	0.041	0.060	0.042
CH_2ClCH_2Cl	0.161	0.124	−0.001	−3.047	−4.290	4.300
	0.050	0.084	0.084	0.069	0.112	0.063
Alkane	0.287	0.649	−1.657	−3.516	−4.818	4.282
	0.028	0.033	0.038	0.034	0.045	0.037
Hexadecane	0.087	0.667	−1.617	−3.587	−4.869	4.433
	0.025	0.030	0.035	0.041	0.040	0.027
Cyclohexane	0.127	0.816	−1.731	−3.778	−4.905	4.646
	0.033	0.041	0.044	0.040	0.061	0.046
Isooctane	0.288	0.382	−1.668	−3.639	−5.000	4.561
	0.026	0.056	0.050	0.055	0.054	0.046
Benzene	0.017	0.490	−0.604	−3.013	−4.628	4.587
	0.036	0.047	0.055	0.048	0.074	0.063
Toluene	0.015	0.594	−0.781	−2.918	−4.571	4.533
	0.038	0.051	0.061	0.052	0.080	0.069
Chlorobenzene	0.046	0.259	−0.466	−3.047	−4.819	4.660
	0.030	0.062	0.067	0.071	0.074	0.045

Phase	c	r	s	a	b	v
Nitrobenzene	-0.181	0.576	0.003	-2.356	-4.420	4.263
	0.054	0.096	0.106	0.120	0.122	0.083
Gas phase	0.994	-0.577	-2.549	-3.813	-4.841	0.869
	0.031	0.032	0.037	0.040	0.040	0.031

[a] Coefficient sd values are below the coefficient.

the best combination of π_2^H and $\sum \beta_2^H$ need be found. In any case, the various determined log P values are listed, and the combination of descriptors (π_2^H, $\sum \alpha_2^H$, and $\sum \beta_2^H$) that gives the smallest standard deviation in observed and calculated log P values is selected as the "best" set.

In Table 2 are given the coefficients in Eq. (2) that have been obtained for various water/phase partitions, including the water/gas system. Of course, more systems are gradually being added to the list, but even now there are a sufficient number to allow the determination of descriptors for a very wide variety of compounds.

An example of the determination of π_2^H, $\sum \alpha_2^H$ and $\sum \beta_2^H$ for the solute 2,4-dinitrophenol is given in Table 3. The value of 1.20 for R_2 was easily estimated by the addition of fragments, and V_x was simply calculated. The other three descriptor values of $\pi_2^H = 1.50$, $\sum \alpha_2^H = 0.10$, and $\sum \beta_2^H = 0.55$ were obtained by the best fit of calculated and observed partition coefficients, and enable 14 log P values to be calculated and matched to the observed values, with an sd value of only 0.07 log units. The example illustrates other uses of the method, also. First, outlying log P values can be identified: the given value of -0.52 for the water/oleyl alcohol system is clearly grossly in error, and the n-butyl acetate value of 2.38 is some way out-of-line. Second, log P values can be predicted for many other water/phase systems: some examples are given in Table 3, including an estimate of -5.59 log units for the water/gas partition, that is almost impossible to measure directly.

For many solutes there will be but few log P values available for different water/solvent systems [26], but even so, it is usually still possible to assign descriptors. This is illustrated for amphetamine (Table 4), where only five water/solvent systems are given. Nearly always, a value of log P_{oct} will be available, but the method does not depend on this; any set of water/solvent partitions for which the coefficients in Eq.(2) are known, will suffice.

A more difficult situation arises if only few partition coefficients are available and if the solute exhibits "variable basicity". In this case $\sum \beta_2^o$ must be obtained from log P values that refer to solvents that contain appreciable quantities of water at saturation (i.e., octanol, butyl acetate, and diethyl ether), and $\sum \beta_2^H$ from log P values in solvents that contain little water at saturation (i.e., alkanes, chloroform, and benzene). As always, R_2 and V_x will be known, and π_2^H and $\sum \alpha_2^H$ must take the same values throughout. Sometimes one or more descriptors can be estimated by analogy with similar structures, to leave fewer descriptors to be obtained from log P values. For example, log P values for 4-n-propylaniline are known [26] only for water/octanol (2.40), water/chloroform (2.99), and water/alkane (1.61). But π_2^H can be estimated as 0.90 by com-

Table 3.   Calculation of descriptors of 2,4-dinitrophenol

$\log P = c + rR_2 + s\pi_2^H + a\sum\alpha_2^H + b\sum\beta_2^H + Vv_x$

$R_2 = 1.20$ $\pi_2^H = 1.50$ $\sum\alpha_2^H = 0.10$ $\sum\beta_2^H = 0.55$ $V_x = 1.1235$

Phase	$\log P$(calc)	$\log P$(obs)
Isobutanol	1.70	
Pentanol	1.78	1.67
Hexanol	1.61	1.63
Octanol	1.56	1.67
Decanol	1.47	1.41
Oleyl alcohol	1.02	(−0.52)
Diethyl ether	1.45	1.41
Ethyl acetate	1.66	
n-Butyl acetate	2.08	(2.38)
Olive oil	1.27	1.35
PGDP	1.47	
CH_2Cl_2	2.56	2.60
$CHCl_3$	2.54	2.48
CCl_4	1.38	1.34
$CH_2Cl.CH_2Cl$	2.48	
Alkane	0.37	0.33
Hexadecane	0.44	
Cyclohexane	0.63	0.60
Benzene	2.01	2.06
Toluene	1.84	1.94
Chlorobenzene	1.94	
Nitrobenzene	2.64	2.53
Gas phase	−5.59	

parison with 4-ethylaniline (0.91), and $\sum\alpha_2^H$ can be taken as the same as that for 4-ethylaniline (0.23), to leave only $\sum\beta_2^O$ and $\sum\beta_2^H$ to be estimated from the log P measurements (see Table 5).

It is not possible to list all the solutes for which we have now determined descriptors, but Table 6 shows values of $\sum\alpha_2^H$ for a number of acids. It is very clear that

Table 4.   Calculation of descriptors for amphetamine

$\log P = c + rR_2 + s\pi_2^H + a\sum\alpha_2^H + b\sum\beta_2^H + vV_x$

$R_2 = 0.795$ $\pi_2^H = 0.82$ $\sum\alpha_2^H = 0.13$ $\sum\beta_2^H = 0.77$ $V_x = 1.24$

Phase	$\log P$(calc)	$\log P$(obs)
Octanol	1.76	1.76
CH_2Cl_2	2.03	1.80
$CHCl_3$	2.41	2.41
Heptane	0.58	0.53
Isooctane	0.55	0.60

Table 5. Calculation of descriptors for 4-*n*-propylaniline

$\log P = c + rR_2 + s\pi_2^{H} + a\sum\alpha_2^{H} + b\sum\beta_2 + vV_x$

$R_2 = 1.922 \quad \pi_2^{H} = 0.90 \quad \sum\alpha_2^{H} = 0.23 \quad \sum\beta_2^{O} = 0.58 \quad \sum\beta_2^{H} = 0.48 \quad V_x = 1.239$

Phase	log P(calc)	log P(obs)
Octanol	2.40[a]	2.40
Chloroform	3.06[b]	2.99
Alkane	1.58[b]	1.61

[a] Using $\Sigma\beta_2^{O}$.
[b] Using $\Sigma\beta_2^{H}$.

Table 6. Some values of $\sum\alpha_2^{H}$ for solutes

Solute	$\Sigma\alpha_2^{H}$
4-Nitrophenol	0.82
Hexafluoropropan-2-ol	0.77
Methyl 4-hydroxybenzoate	0.69
Methyl 3-hydroxybenzoate	0.66
Acetic acid	0.61
Phenol	0.60
Benzoic acid	0.59
Ethanol	0.37
2,4-Dinitrophenol	0.10
Methyl 2-hydroxybenzoate	0.04
Cocaine	0.00
Caffeine	0.00
Morphine	0.86
Codeine	0.31
Heroin	0.00
Cimetidine	0.67
Icotidine	0.60
Lupitidine	0.60
Clonidine	0.35
Mepyramine	0.00
Imipramine	0.00
Ranitidine	0.25
Tiotidine	1.18
Zolantidine	0.40
Temelastine	0.60
Progesterone	0.00
Deoxycorticosterone	0.15
Testosterone	0.32
Corticosterone	0.40
Cortisone	0.35
Hydrocortisone	0.70
Estrone	0.56
Estradiol	0.88

Table 7. Some values of $\sum \beta_2^H$ for solutes[a]

Solute	$\Sigma \beta_2^H$	
Piperazine	1.17	
Urea	0.90	
Dimethyl phthalate	0.88	
2-Methoxyethanol	0.84	
Triethylamine	0.79	
Ethan-1,2-diol	0.78	
Dimethylformamide	0.74	
Urethane	0.64	
Butanone	0.51	
Diethyl ether	0.45	
Ethyl acetate	0.45	
1,4-Dinitrobenzene	0.41	
Acetonitrile	0.32	
Nitromethane	0.31	
Phenol	0.30	
Nitrobenzene	0.28	
3,5-Dichlorophenol	0.00	
Aniline	0.41	(0.50)
Pyridine	0.52	(0.47)
Cocaine	1.50	
Caffeine	1.35	
Morphine	1.79	
Codeine	1.79	
Heroin	2.02	
Cimetidine	2.21	(1.93)
Icotidine	2.16	(2.02)
Lupitidine	2.78	(2.64)
Clonidine	1.08	
Mepyramine	1.59	
Imipramine	1.19	
Rantidine	2.33	
Tiotidine	2.23	
Zolantidine	1.38	
Temelastine	2.00	(1.87)
Progesterone	1.14	
Deoxycorticosterone	1.13	
Testosterone	1.19	
Corticosterone	1.63	
Cortisone	1.84	
Hydrocortisone	1.87	
Estrone	0.91	
Estradiol	0.95	

[a]Values of $\Sigma \beta_2^O$ in parentheses for variable basicity solutes.
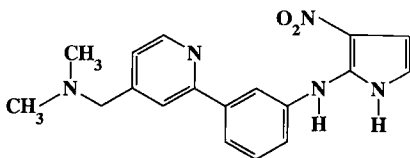
Figure 2. 2-[[3-[4-[(dimethylamino)methyl]pyrid-2-yl]phenyl]amino]-3-nitropyrrole (compound 29).

the OH hydrogen-bond acidity depends on the surrounding molecular structure, and that little can be achieved by use of a single "indicator variable" for the OH group. The situation is even more pronounced for the NH group; Abraham et al. [27] have shown from complexation constants towards *N*-methylpyrrolidinone that the NH acid, 5-phenyl-1,2,3,4-tetrazole is as strong a hydrogen-bond acid as is trifluoroacetic acid, that 3-(3-phenylpropyl)-1,2,4-triazole is as strong as acetic acid, and that alkylamines have almost no hydrogen-bond acidity at all.

In a similar vein, there is little connection between our $\sum\beta_2^H$ hydrogen-bond basicity scale and indicator variables such as the number of lone pairs, or the number of oxygen/nitrogen atoms in a molecule. Some examples are given in Table 7, including values of $\sum\beta_2^H$ for multifunctional bases.
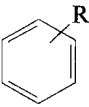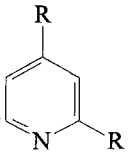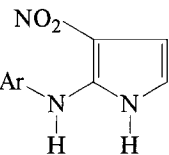
fragment	R_2	π^H_2	$\sum\alpha^H_2$	$\sum\beta^H_2$	V_x
R_3N	0.140	0.15	0.00	0.67	
	0.601	0.52	0.00	0.14	
	0.630	0.76	0.00	0.63	
	1.220	1.08	0.46	0.70	
TOTAL	**2.591**	**2.51**	**0.46**	**2.14**	**2.55**

Figure 3. Addition of fragments for 2-[[3-[4-[(dimethylamino)methyl]pyrid-2-yl]phenyl]amino]-3-nitropyrrole (compound 29).

Another very valuable procedure for the estimation of descriptors for complicated structures is by the summation of descriptors for fragments or substructures. This method has been outlined for a number of structurally diverse drug molecules [28], and for a set of steroids [29], and tables of descriptors for important fragments or groups have been set out [28, 29]. As pointed out [28], in the summation of fragments it is very important to avoid intramolecular interactions between the specified fragments, and a structure must be broken down into fragments in such a way that interactions between the fragments in the molecule are minimized. An advantage of the method is that it is possible to assign descriptors to drug molecules by examination of their structure, and hence to be able to predict properties before synthesis. As an example we use the compound designated before [3, 28] as "compound 29" (see Fig. 2), and which is 2-[[3-[4-[(dimethylamino)methyl]pyrid-2-yl]phenyl]amino]-3-nitro-pyrrole. The fragments and their addition are set out in Fig. 3. It is important that the 2-amino-3-nitropyrrole entity is considered as a single fragment or substructure, because any intramolecular interactions between e.g., the amino and the nitro group will then be accounted for in the substructure.

18.3 Applications of the Solvation Equation (Eq. (2))

There is little point showing the application of Eq. (2) to all of the water/solvent partitions that we have investigated, but we do give one interesting example, that of the water/octanol partition,

$$\log P_{oct} = (0.09 \pm 0.02) + (0.56 \pm 0.01)R_2 - (1.05 \pm 0.02)\pi_2^{H} +$$
$$(0.03 \pm 0.02)\sum \alpha_2^{H} - (3.46 \pm 0.03)\sum \beta_2^{H} + (3.81 \pm 0.01)V_x \quad (4)$$

$$n = 613, \ r^2 = 0.995, \ \text{sd} = 0.116, \ F = 23162$$

The correlation matrix for Eq. (4) is very nearly the same as that for Eq. (3), with the maximum cross-correlation between R_2 and π_2^{H} ($r^2 = 0.531$). Note that $\sum \beta_2^{H}$ is used in Eq. (4) because no "variable" basicity compounds were included in the analysis. As in the water/cyclohexane partition, Eq. (3), solute excess molar refraction (weakly) and solute volume (strongly) favor the organic phase, whereas solute dipolarity and solute hydrogen-bond basicity favor water. Thus, for both partitions, it can be deduced that the organic phase is more polarizable, and more hydrophobic than water, and is less dipolar and less acidic than water. Note that the negative b-coefficient in Eq. (4), -3.46, refers to the difference in hydrogen-bond acidity between the phases, since solvent acidity is the complementary property to solute basicity. An interesting feature of Eq. (4) is the effectively zero value of the a-coefficient, 0.03, which means that water and octanol (or more correctly, wet octanol) have the same hydrogen-bond basicity. This effect is not unique to wet octanol, as we have recently shown that the a-coefficient is nearly zero for all the water/alcohol partitions were have studied [25]. How the various solute factors influence the water/octanol partition is shown in Fig. 4.

We have checked that the previous equations [30] on the inhibition of firefly luciferase activity by aqueous solutes, and on general anesthesia by aqueous solutes, are confirmed when Eq. (2) is used instead, but now deal with a number of special topics.
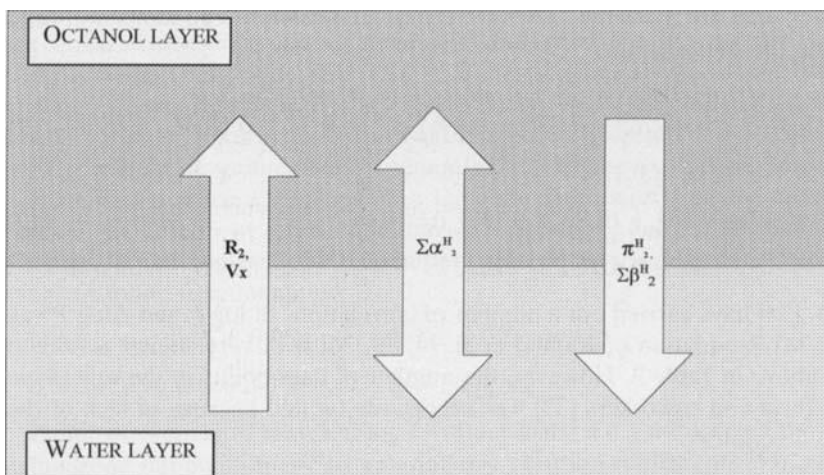
Figure 4. Solute factors that influence log P_{oct}.

18.3.1 Seiler's Δlog P Parameter

The Δlog P parameter, defined by Seiler [31] as the difference between log P_{oct} and log P for alkane or cyclohexane partition, has become an important descriptor in biological QSARs. In the present context, it makes little difference if Δlog P is obtained from water/alkane or water/cyclohexane partitions, but more values are available for Δlog P defined as,

$$\Delta\log P_{16} = \log P_{oct} - \log P_{16} \tag{5}$$

where P_{oct} and P_{16} are water/octanol and water/hexadecane partition coefficients. Application [24] of Eq. (2) yields,

$$\Delta\log P_{16} = -(0.07 \pm 0.04) - (0.09 \pm 0.04)R_2 + (0.53 \pm 0.05)\pi_2^H + \\ (3.65 \pm 0.06) \sum\alpha_2^H + (1.40 \pm 0.06)\sum\beta_2^H - (0.52 \pm 0.04)V_x \tag{6}$$

$n = 288$, $r^2 = 0.967$, sd $= 0.173$, F $= 1646$

As El Tayar et al. [32] pointed out, the main contributing term to Δlog P is that of solute hydrogen-bond acidity. However, the other terms cannot be ignored, and if they are left out to give a regression against $\sum\alpha_2^H$ only, it becomes clear [24] that the Δlog P parameter will provide only a very rough estimate of solute hydrogen-bond acidity,

$$\Delta\log P_{16} = (0.04 \pm 0.03) + (4.50 \pm 0.13)\sum\alpha_2^H \tag{7}$$

$n = 288$, $r^2 = 0.823$, sd $= 0.396$, F $= 1329$

These results can be checked using Δlog P now defined by the original [31] equation,

$$\Delta\log P_{cyc} = \log P_{oct} - \log P_{cyc} \tag{8}$$

Although fewer data are available [24], the resulting correlation equations are quite similar to those for $\Delta\log P_{16}$,

$$\Delta\log P_{\mathrm{cyc}} = - (0.07 \pm 0.05) - (0.35 \pm 0.06)R_2 + (0.76 \pm 0.06)\pi_2^{\mathrm{H}} + (3.82 \pm 0.06)\sum\alpha_2^{\mathrm{H}} + (1.38 \pm 0.09)\sum\beta_2^{\mathrm{H}} - (0.77 \pm 0.08)V_x \quad (9)$$

$n = 128$, $r^2 = 0.979$, sd $= 0.178$, F $= 1138$

$$\Delta\log P_{\mathrm{cyc}} = (0.17 \pm 0.05) + (3.90 \pm 0.13)\sum\alpha_2^{\mathrm{H}} \quad (10)$$

$n = 128$, $r^2 = 0.979$, sd $= 0.420$, F $= 915$

El Tayar et al. [32] have carried out a number of correlations of log P and $\Delta\log P$ values using the MLR equation of Kamlet et al. [4, 19, 20] with very similar results to those shown above in Table 2. However, the number of data points in the regression equations of Testa and coworkers [32] was necessarily limited because of lack of the Kamlet et al. descriptors.

18.3.2 Reversed-phase HPLC

The use of reversed-phase HPLC to estimate partition coefficients is a well established practice, especially with respect to log P_{oct} values [33–38]. A correlation is established between log k' in a given system, or log k'_{w}, the capacity factor extrapolated to 100 % water, and log P_{oct} for a training set, and further measurements of log k' or log k'_{w} are used to estimate log P_{oct} from the correlation equation. It is usually found that the correlation equation holds for solutes of similar hydrogen-bond functionality, but breaks down for solutes which are of different hydrogen-bond ability to the training set [37, 38]. Kamlet and coworkers [39–43] applied their solvatochromic equation to several sets of reversed-phase HPLC capacity factors, as log k', and showed that the factors that influenced HPLC retention were in general not the same as those that influenced log P_{oct} values [39, 41]. More recently, Miller and Poole [44] have applied Eq. (2) to reversed-phase HPLC log k' values, and Larrivee and Poole [45] have used Eq. (2) to interpret breakthrough volumes in solid phase extraction. The relationship between the log k' values obained by Yamagami and Takao [37] and log P_{oct} values has been investigated [46] through the application of Eq. (2), and the role of hydrogen-bonding established explicitly.

Much larger data sets to any of the above were used by Abraham and Roses [47] who applied Eq. (2) to log k' values on a number of C_{18} phases with various mobile phases. As an example, the following equation was found for a Spherisorb ODS-2 phase with 60 % methanol/water as the eluate; log k' values were from the extensive work of Smith and Burr [48–53].

$$\log k' = - (0.322 \pm 0.047) + (0.252 \pm 0.043)R_2 - (0.651 \pm 0.032)\pi_2^{\mathrm{H}} - (0.429 \pm 0.030)\sum\alpha_2^{\mathrm{H}} - (1.529 \pm 0.042)\sum\beta_2^{\mathrm{O}} + (1.773 \pm 0.040)V_x \quad (11)$$

$n = 126$, $r^2 = 0.984$, sd $= 0.072$, F $= 1408$

Note that in Eq. (11) the alternative $\sum\beta_2^{\mathrm{O}}$ descriptor was used. The correlation matrix for the 126 solutes shows the greatest cross-correlation to be between R_2 and π_2^{H} ($r^2 = $

0.329). Although the main factors that influence HPLC retention in this typical system are qualitatively the same as those that influence log P_{oct} partitions, i.e., solute hydrogen-bond basicity decreases log k' and log P_{oct} and solute volume increases log k' and log P_{oct}, quantitatively they are not the same [46, 47]. Hence, for a varied group of solutes, especially with considerable differences in hydrogen-bond acidity, log k' and log P_{oct} will not be well related, in general. There can be exceptions; thus Minick et al. [35] showed that log k'_w values on a C_8 silica phase with aqueous methanol mobile phases modified by the addition of octanol and n-decylamine with a 4-morpholinopropanesulfonic acid buffer (pH 7.4) matched log P_{oct} values over a wide range of solute functionalities,

$$\log P_{oct} = (0.146 \pm 0.058) + (1.090 \pm 0.023)\log k'_w \qquad (12)$$

$$n = 24, \ r^2 = 0.990, \ sd = 0.172, \ F = 2180$$

Other workers have matched log k' values using a C_{18} derivatized polystyrene/divinylbenzene stationary phase to water/alkane partition coefficients [54], and have used immobilized artificial membranes as stationary phases in order to predict drug absorption [55].

18.3.3 Water/Micelle Partition

The partition of solutes between water and micelles is important in separation science [56] and in biological chemistry where the use of micelles to facilitate intermolecular hydrogen-bonding is a recent important advance [57]. Hitherto, the factors that govern water/micelle partitions have not been elucidated, although it is known that water/micelle partition coefficients are not well related to water/octanol partitions [58–60]. Eq. (2) has been applied [60] to partition of solutes between water and sodium dodecylsulfate micelles, with the mol fraction partition coefficient defined as,

$$K_x = \frac{[\text{mol fraction solute in micelle}]}{[\text{mol fraction solute in water}]}$$

For 138 varied solutes, application of Eq. (2) yielded,

$$\log K_x = (1.280 \pm 0.060) + (0.484 \pm 0.063)R_2 - (0.431 \pm 0.079)\pi_2^H$$
$$- (0.183 \pm 0.067)\sum\alpha_2^H - (1.721 \pm 0.088)\sum\beta_2^O + (2.878 \pm 0.079)V_x \qquad (13)$$

$$n = 138, \ r^2 = 0.962, \ sd = 0.192, \ F = 668$$

However, it has been suggested [58] that alkanes behave differently to other solutes in that they are sorbed by the hydrophobic chain part of the micelle. Six alkanes were therefore left out, and the preferred regression equation was,

$$\log K_x = (1.201 \pm 0.058) + (0.542 \pm 0.057)R_2 - (0.400 \pm 0.071)\pi_2^H$$
$$- (0.133 \pm 0.060)\sum\alpha_2^H - (1.580 \pm 0.082)\sum\beta_2^O + (2.793 \pm 0.073)V_x \qquad (14)$$

$$n = 132, \ r^2 = 0.970, \ sd = 0.171, \ F = 817$$

The correlation matrix for the 132 sets of descriptors shows that the greatest cross correlations are between R_2 and π_2^H ($r^2 = 0.533$) followed by R_2 and V_x ($r^2 = 0.365$) and

$\pi_2{}^H$ and V_x ($r^2 = 0.317$). The main factors that influence the log K_x values are solute hydrogen-bond basicity that decreases partition, and solute volume that increases partition. Quantitatively these are rather similar to the factors that influence water/iso-butanol partition (see Table 2) and it was suggested [60] that sodium dodecylsulfate micelles were about as hydrophobic as wet isobutanol.

There was not a very good correlation between the log K_x values and log P_{oct} for the 132 solutes,

$$\log K_x = (2.01 \pm 0.05) + (0.693 \pm 0.023)\log P_{oct} \tag{15}$$

$$n = 132, \, r^2 = 0.873, \, sd = 0.346, \, F = 896$$

However, incorporation of the V_x descriptor yielded a reasonable equation,

$$\log K_x = (1.13 \pm 0.07) + (0.504 \pm 0.019)\log P_{oct} + (1.216 \pm 0.084)V_x \tag{16}$$

$$n = 132, \, r^2 = 0.952, \, sd = 0.215, \, F = 1269$$

The correlation between log P_{oct} and V_x for the 132 sets of descriptors is not large ($r^2 = 0.454$) and Eq. (16) is a reasonable predictive equation for water/sodium dode-cylsulfate micelle partition coefficients, especially since both V_x and log P_{oct} can be calculated from structure [12, 18, 26, 61].

18.3.4 The Blood-Brain Barrier

As mentioned by Zahradník and Hobza (Chapter 3) and by Richards (Chapter 10), statistical mechanical calculations could in principle be applied to many of the systems discussed above. This is not the case for the real biological systems we now turn to, for the reason that the solvent phase is simply too complicated to deal with in this way, at least at present.

There are various measures of the propensity of a drug molecule to cross the blood-brain barrier, and some confusion exists in the literature through failure to define terms. Thus Seelig et al. [62] attempted to relate "... the ability of drugs to diffuse through the blood-brain barrier" to the drug lipophilicity. However, the measure of this ability ranged from studies on tissue and plasma distribution [63] to binding studies [64]. Not surprisingly, no general relationship was found.

The oft-quoted [65] rule that a log P_{oct} value of about 2 is optimal for ready entry into brain [66] is derived, not from rates of permeation or from equilibria, but from studies on biological activity – for example the potency of hypnotics [67].

We shall not deal with the physiology of the blood-brain barrier, as this has been re-viewed in some depth by Bradbury [68] quite recently, but will concentrate on a well-defined measure of passive transport across the blood-brain barrier. The distribution of drugs and other molecules between blood and brain can be defined as,

$$BB = \frac{[\text{conc. of solute in brain}]}{[\text{conc. of solute in blood}]} \tag{17}$$

In a key paper, Young et al. [3] determined BB values for 39 drug molecules by a radio-active assay method. These showed that for a set of 20 of these drug molecules there was a poor correlation between log BB and log P_{oct}, but a markedly better one between log BB and Δlog P,

$$\log BB = -1.22 + 0.266 \log P_{oct} \tag{18}$$

$$n = 20, \; r^2 = 0.190, \; sd = 0.711, \; F = 4.2$$

$$\log BB = 0.889 - 0.485 \, \Delta\log P \tag{19}$$

$$n = 20, \; r^2 = 0.691, \; sd = 0.439, \; F = 40.2$$

This was later [28] confirmed for a data set that included all the compounds studied by Young et al. [3], as well as other compounds to make a much wider data set than hith-erto.

In another approach, van de Waterbeemd and Kansy [69] defined a solute parameter from log P_{cyc} and a calculated molar volume,

$$\Lambda_{cyc} = 1.098 - \log P_{cyc} + 0.039 \, V_m \tag{20}$$

and then obtained a correlation for the 20 drug molecules in the Young-Mitchell [3] data set,

$$\log BB = 1.730 - 0.338 \, \Lambda_{cyc} + 0.007 \, V_m \tag{21}$$

$$n = 20, \; r^2 = 0.872, \; sd = 0.290, \; F = 58$$

It was noted [28], however, that the very positive constant (1.730) in Eq. (21) was unre-alistic, and could lead to difficulties in the estimation of log BB values for compounds that were outside the range of the training set. For example, Eq. (21) predicts log BB for methane to be 1.94, but the actual value is 0.04 log unit [70]. Another correlation reported by van de Waterbeemd and Kansy [69] uses the hydrophilic or polar surface of the drug molecule, SP, as a descriptor,

$$\log BB = 1.643 - 0.021 \, SP - 0.003 \, V_m \tag{22}$$

$$n = 20, \; r^2 = 0.697, \; sd = 0.448, \; F = 19.5$$

but the sd value of 0.45 log units seems rather too high for the equation to be very use-ful. Calder and Ganellin [71] have examined Eq. (19) and Eq. (22). Their results for the test set of compounds are shown in Table 8. The calculated log BB values for the final two compounds are too high by over a log unit, but a possible reason for such dis-crepancies is biological degradation of the test compound. This will invariably result in the observed log BB value being much lower than the "correct" value. In addition, the efflux mechanism mentioned in Chapter 14 will also result in a lower observed log BB value. Even if we exclude the final two compounds, the estimations through Eq. (19) and through the van de Waterbeemd and Kansy method, Eq. (22), are not very good. As Calder and Genallin [71] reported, both equations overestimate the blood-brain distribution.

It appears that part of the difficulty in constructing a general equation for the es-timation of log BB values is the restriction to 20 of the Young-Mitchell training set. It

Table 8. Estimations of log BB through Eq. (19) and Eq. (22)[a]

Compound	log BB(obs)[b]	log BB(calc)[b]	
		Eq. (19)	Eq. (22)
Thioperamide	− 0.16		0.08
2-(CH₂CH₂NHMe)pyridine	− 0.30	0.08	0.63
2-(CH₂CH₂NMe₂)pyridine	− 0.06	0.26	0.85
2-(CH₂CH₂NH₂)thiazole	− 0.42	0.02	0.43
4-Phenyl-2-(CH₂CH₂NH₂)thiazole	− 1.30	0.18	0.27
4-(CH₂CH₂NH₂)-1,2-benzimidazole	− 1.40	− 0.72	0.28

[a] A slightly different version of Eq. (22) was used [71].
[b] log BB is defined as log (C_{brain}/C_{blood}).

is possible to use all the 39 compounds of the Young-Mitchell set for which log BB values had been determined, together with the data collected by Abraham and Weathersby [70] on small molecules. Of the total 65 compounds, eight were very pronounced outliers, possibly for the reasons mentioned above, but for the remaining 57 compounds a reasonable correlation was found [28],

$$\log BB = - (0.04 \pm 0.06) + (0.20 \pm 0.10)R_2 - (0.69 \pm 0.12)\, \pi_2^{H}$$
$$- (0.71 \pm 0.33)\sum \alpha_2^{H} - (0.70 \pm 0.11)\sum \beta_2^{H} + (1.00 \pm 0.10)V_x \tag{23}$$

$$n = 57,\ r^2 = 0.907,\ sd = 0.197,\ F = 99$$

Although 57 solutes covering a wide range of values of the descriptors have been used in Eq. (23), they do not constitute an ideal set at all, there being considerable cross-correlations between the descriptors. In terms of r^2 the main ones are R_2/π_2^{H} (0.929), $\pi_2^{H}/\sum \beta_2^{H}$ (0.885), $R_2/\sum \beta_2^{H}$ (0.844), and $R_2/\sum \alpha_2^{H}$ (0.537). There is therefore some redundancy among the descriptors, and the $r \cdot R_2$ term can be left out to yield,

$$\log BB = - (0.08 \pm 0.06) - (0.54 \pm 0.10)\pi_2^{H} - (0.61 \pm 0.13)\sum \alpha_2^{H}$$
$$- (0.71 \pm 0.11)\sum \beta_2^{H} + (1.03 \pm 0.10)V_x \tag{24}$$

$$n = 57,\ r^2 = 0.899,\ sd = 0.202,\ F = 117$$

However, if any other term is omitted, the equation is significantly worse.

These equations show, for the first time, exactly the solute factors that influence equilibria across the so-called blood-brain barrier. Solute dipolarity, hydrogen-bond acidity, and hydrogen-bond basicity all reduce the log BB value, while solute excess molar refraction (marginally) and solute size increase log BB and favor brain. It is now possible to design drugs that will specifically target brain instead of blood, or conversely will remain in blood. As an example, in Table 9 are given some calculated log BB values for substituted zolantidines (Fig. 5), showing how it is possible to modify structures in a controlled way to achieve either more positive or more negative log BB values, at will [72].
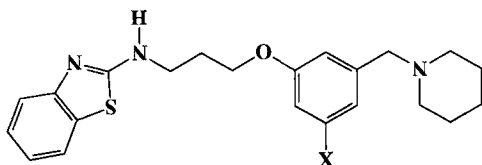
Figure 5. Zloantidine; X = H.

Not only is it possible to predict log BB values for related structures, but log BB for any structure can be predicted, provided the necessary descriptors in Eq. (23) are available. As shown above, these may be obtained from various log P measurements, but naturally such measurements require that the compound has already been synthesized. By suitable choice of fragments or substructures, it is possible to obtain descriptors for large drug molecules by addition of fragment values. Thus it is feasible to calculate log BB values for a range of molecules simply from molecular structure, that is, before synthesis. These molecules are quite complicated, for example zolantidine

Table 9. Calculation of log BB for substituted zolantidine derivatives

X	R_2	π_2^H	$\Sigma\alpha_2^H$	$\Sigma\beta_2^H$	V_x	log BB(calc)
H	2.69	2.64	0.48	1.38	2.995	0.36[a]
Me	2.69	2.67	0.48	1.39	3.135	0.48
Et	2.69	2.67	0.48	1.39	3.276	0.62
Cl	2.81	2.75	0.48	1.31	3.117	0.48
OMe	2.80	2.90	0.48	1.54	3.194	0.29
OH	2.82	3.06	1.07	1.48	3.053	− 0.33
NH$_2$	3.01	3.11	0.73	1.64	3.094	− 0.16

[a] Observed value is 0.14 log units.

(Fig. 5) and the Young-Mitchell compound 29, which is 2-[[3-[4-[(dimethylamino)methyl]pyrid-2-yl]phenyl]amino]3-nitropyrrole (see Fig. 2), so that the procedure is general. In the case of compound 29, no log P measurements were available, and so the descriptors had all to be calculated by the summation of fragment values. The final result [28] was a calculated value for log BB through Eq. (23) of − 0.54, as compared with the observed value [3] of − 0.28 log units. Hence, the method can be applied to more-or-less any structure, whether or not the compound has been synthesized.

18.3.5 Permeation Through Skin

The permeation or penetration of compounds through skin is of increasing importance. In the pharmaceutical industry, interest in skin permeation is not only confined to the traditional application of, e.g., anti-inflammatory agents, but also as an alternative drug delivery system. Skin permeation is also of obvious interest in the cosmetic industry. A different, but important area is that of occupational toxicology, because ab-
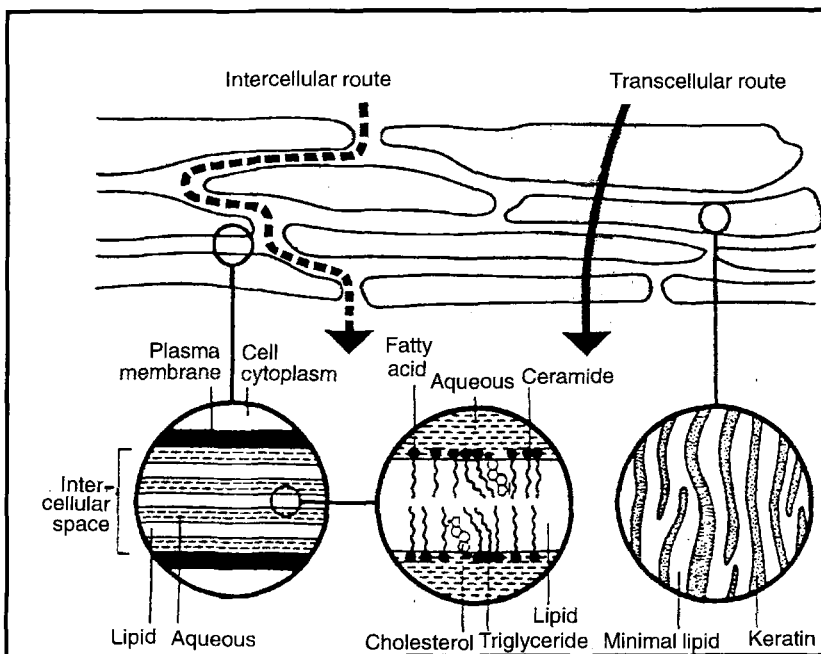
Figure 6. The postulated intercellular and transcellular permeation routes through skin.

sorption of noxious chemicals through the skin is acknowledged to be a significant route uptake. The measurement of skin permeation is laborious, however, and so there is a growing need for a method of predicting percutaneous permeation.

The principal barrier to skin permeation is the outermost skin layer, the stratum corneum, composed of dead cells surrounded by an intercellular matrix of lipid and aqueous layers. An important property of the stratum corneum is its ability to become heavily hydrated on contact with water. Quite typically, the hydrated stratum corneum contains about 70 % w/w water. Two main pathways have been suggested for the permeation of compounds through hydrated skin, the intercellular and the intracellular route [73–76], as shown in Fig. 6.

Two distinct types of measurement have been made on the interaction of compounds with skin. Firstly, water/skin partition coefficients, K_m , have been determined, using excised human skin as a partitioning phase. Secondly, solute permeation rates through skin, k_p, have been determined using the stratum corneum as a membrane between two aqueous compartments. The partition coefficient is defined as,

$$K_m = \frac{[\text{mol solute absorbed per unit mass of dry tissue}]}{[\text{mol solute in solution per unit mass of water}]} \tag{25}$$

El Tayar et al. [76] have examined the partition data of Scheuplein and colleagues [73, 74] for alcohol and steroid solutes, and showed that log K_m was correlated with the water/octanol partition coefficient,

$$\log K_m = 0.104 + 0.514 \log P_{oct} \tag{26}$$

$$n = 19, \; r^2 = 0.940, \; \text{sd} = 0.163, \; F = 265$$

Although Eq. (26) is quite good, it conveys no information as to the factors that influence $\log K_m$. In order to apply Eq. (2), a knowledge of the descriptors for alcohols and steroids is necessary. Those for the alcohol solutes were available [12], but those for steroids had to be obtained, either from $\log P$ values as shown above, or by the method of fragment summation. It was relatively easy to construct a scheme for the calculation of R_2, $\sum \alpha_2^H$, and $\sum \beta_2^H$ from fragments [29], but less easy for the π_2^H descriptor. However, since π_2^H was the only remaining unknown descriptor, it could be obtained from known $\log P$ values. Application of Eq. (2) to the $\log K_m$ values for alcohols and steroids then yielded,

$$\begin{aligned}
\log K_m = &- (0.03 \pm 0.14) - (0.37 \pm 0.11)\pi_2^H + (0.33 \pm 0.15)\sum \alpha_2^H \\
&- (1.67 \pm 0.16)\sum \beta_2^H + (1.87 \pm 0.17)V_x
\end{aligned} \tag{27}$$

$$n = 22, \; r^2 = 0.943, \; \text{sd} = 0.166, \; F = 70$$

As expected for a rather restricted set of solute, there exists cross-correlations between the descriptors, the largest being, in terms of r^2; π_2^H/V_x (0.925), $\pi_2^H/\sum \beta_2^H$ (0.783), and $\sum \beta_2^H/V_x$ (0.797). Hence Eq. (27) cannot be regarded as very firmly based, but will suffice to give information as to the nature of hydrated skin by comparison to various water/solvent partitions (see Table 8). Note that in Eq. (27) the R_2 descriptor has already been left out. If the $\sum \alpha_2^H$ descriptor is also omitted we find,

$$\begin{aligned}
\log K_m = &(0.18 \pm 0.12) - (0.27 \pm 0.10)\pi_2^H \\
&- (1.66 \pm 0.18)\sum \beta_2^H + (1.71 \pm 0.17)V_x
\end{aligned} \tag{28}$$

$$n = 22, \; r^2 = 0.926, \; \text{sd} = 0.184, \; F = 75$$

If further descriptors are left out, the statistics become very much worse.

It is clear that hydrated skin is much more water-like than are the wet alcohols, there being a clear trend of both the b-coefficient and the v-coefficient with the % w/w water in the organic phase (given in the last column of Table 10). The coefficients in Eq. (26) thus fairly reflect the general chemical make-up of the hydrated stratum corneum. The fact that this tissue is 70 % w/w water, and behaves as an aqueous, not very hydrophobic phase, is crucial in the understanding of water-skin permeability coefficients.

Table 10. Regression constants for partition from water
$\log P = c + rR_2 + s\pi_2^H + a\sum \alpha_2^H + b\sum \beta_2^H + vV_x$

Phase	c	r	s	a	b	v	%aq^a
Skin	− 0.03	0.00	− 0.37	0.33	− 1.67	1.87	70
Isobutanol	0.23	0.51	− 0.69	0.02	− 2.26	2.78	17
Pentanol	0.18	0.57	− 0.79	0.02	− 2.84	3.25	9
Hexanol	0.14	0.72	− 0.98	0.14	− 3.21	3.40	7
Octanol	0.09	0.56	− 1.05	0.03	− 3.46	3.81	5
Decanol	0.01	0.49	− 0.97	0.01	− 3.80	3.95	4

[a] Percentage water w/w in the organic layer or in hydrated skin.

El Tayar et al. [76] found only a very poor correlation of the permeability coefficient (k_p in cm s^{-1}) with log P_{oct}, for a set of alcohols and steroids, although a better correlation was found with Seiler's Δlog P parameter, again for alcohols and steroids. Unfortunately, the Δ log P parameter does not correlate with the log k_p values found by Roberts et al. [77] for phenols, so that neither log P_{oct} nor Δlog P can be used as a general descriptive indicator.

However, it was found [29] that Eq. (2) could be applied to the combined data set of alcohols, steroids, and phenols (and a few other compounds studied also by Scheuplein and Roberts), to yield the first general equation for skin permeation,

$$\log k_p = -(5.05 \pm 0.14) - (0.59 \pm 0.12)\pi_2^H - (0.63 \pm 0.16)\sum\alpha_2^H$$
$$- (3.48 \pm 0.16)\sum\beta_2^H + (1.79 \pm 0.19)V_x \tag{29}$$

$$n = 46, \ r^2 = 0.958, \ \text{sd} = 0.249, \ \text{F} = 235$$

Again, the data set is not ideal, with cross-correlations, in terms of r^2, between π_2^H/V_x (0.912), $\pi_2/\sum\beta_2^H$ (0.728). There is little that can be done about such cross-correlations that arise because of a particular data set that has been studied, rather than because of any inherent cross-correlation of the descriptors themselves. Again, in Eq. (29) the R_2 descriptor has already been left out. If the $\sum\alpha_2^H$ descriptor is also omitted we find,

$$\log k_p = -(5.46 \pm 0.11) - (0.81 \pm 0.12)\pi_2^H - (3.24 \pm 0.18)\sum\beta_2^H$$
$$+ (2.10 \pm 0.21)V_x \tag{30}$$
$$n = 46, \ r^2 = 0.943, \ \text{sd} = 0.287, \ \text{F} = 231$$

If π_2^H is left out but $\sum\alpha_2^H$ retained, the equation is not so good as Eq. (30), and if both π_2^H and $\sum\alpha_2^H$ are left out (as well as R_2) the statistics become very poor.

As for water-skin partition, the two main factors that influence the rate of permeation are solute hydrogen-bond basicity that decreases rate, and solute volume that increases it. On our analysis, there is only one general route, and it is not necessary to specify two separate routes at all. This analysis is in accord with the recent suggestion of Guy and Potts [78], and also with the nature of the stratum corneum. We feel that any division into intercellular and transcellular routes must be arbitrary, given that the intercellular phase and the dead cells both contain lipids, and that both are hydrated.

A comparison of Eq. (29) with the log P_{oct} Eq. (4) shows that the relative values of the coefficients in the two equations differ substantially. Thus the v/b ratio in Eq. (29) is only -0.465, but in Eq. (4) is -1.101, so that volume has relatively a much larger effect on skin permeation than on the water/octanol distribution. This is why the latter distribution is a poor model for skin permeation. In the Δlog P_{16} Eq. (6) the v/b ratio is -0.371, much closer to that in Eq. (29), but now v/a in Eq. (6) -0.142 whereas it is -2.823 in Eq. (29), so that Δlog P_{16} will not be a good model for skin permeation, either. Since volume has relatively a much larger effect on skin permeation than on log P_{oct} values, it should be possible to construct a better correlation equation than by using log P_{oct} alone, simply by incorporating volume as another descriptor. For the 46 solutes in Eq. (29), log P_{oct} values are available for 43, but these are not correlated to log k_p ($r^2 = 0.130$) [29].

Incorporation of a V_x descriptor yields a reasonable predictive equation for log k_p,

$$\log k_p = -(5.63 \pm 0.12) + (0.812 \pm 0.053)\log P_{oct} - (0.727 \pm 0.037)V_x \tag{31}$$

$$n = 43, r^2 = 0.918, \text{sd} = 0.340, F = 224$$

For the 43 solute data set in Eq. (31) the value of r^2 between log P_{oct} and V_x is only 0.167; although the sd value of 0.34 in Eq. (31) is more than the value of 0.25 in Eq. (29), it is possible that Eq. (31) will be useful in providing a first estimate of log k_p values.

This analysis illustrates another facet of the general Eq. (2), namely that it is now possible to analyse influences on processes and various models, and to examine exactly why a particular model is, or is not, a good model process.

18.4 Conclusions

The general Eq. (2) can be applied to all kinds of processes that involve transport of solutes – either in terms of equilibrium partition between phases or rates of transfer from one phase to another. We have given a few examples; others are aqueous toxicity towards various organisms [12, 30], and a general equation for the solubility of vapors in water [79].

A key point in the application of Eq. (2) as an LFER or as a QSAR is the determination of the necessary solute descriptors. Methods of determination from experimental water/solvent partition coefficents have been devised, but a very valuable extension is through summation of values for fragments or substructures. It is therefore now possible to examine numerous biological processes that are transport-dominated, in which a set of structurally varied compounds is used, and to predict biological properties. All that is required is a knowledge of the descriptors for any particular solute. Our own particular aim is to determine descriptors by the methods outlined above for drug and other molecules in order to apply the general LFERs and QSARs to further processes of biological and environmental importance.

References

[1] Hansch, C., *Acc. Chem. Res.* **2**, 232–239 (1969)

[2] Hansch, C., *Acc. Chem. Res.* **26**, 147–153 (1993)

[3] 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)

[4] Kamlet, M. J., Doherty, R. M., Abboud, J.-L. M., Abraham, M. H. and Taft, R. W., *CHEMTECH* **16**, 566–576 (1988)

[5] Abraham, M. H., and Liszi, J., *J. Chem. Soc. Faraday Trans. 1* **74**, 1604–1614 (1978)

[6] Taft, R. W., Gurka, D., Joris, L., von R. Schleyer, P., and Rakshys, J. W., *J. Am. Chem. Soc.* **91**, 4801–4808 (1969)

[7] Abraham, M. H., Grellier, P. L., Prior, D. V., Duce, P. P., Morris, J. J., and Taylor, P. J., *J. Chem. Soc. Perkin Trans.* **2**, 699–711 (1989)

[8] Abraham, M. H., Grellier, P. L., Prior, D. V., Morris, J. J., and Taylor, P. J., *J. Chem. Soc. Perkin Trans.* **2**, 521–529 (1990)

[9] Laurence, C., and Queignec, R., *J. Chem. Soc. Perkin Trans.* **2**, 1915–1917 (1992)

[10] Le Questel, J. Y., Laurence, C., Lachkar, A., Helbert, M., and Berthelot, M., *J. Chem. Soc. Perkin Trans.* **2**, 2091–2094 (1992)

[11] Berthelot, M., Helbert, M., Laurence, C., and Le Questel, J.-Y., *J. Phys. Org. Chem.* **6,** 302–306 (1993)

[12] Abraham, M. H., *Chem. Soc. Revs.* **22,** 73–83 (1993); Pure Appl. Chem. **65,** 2503–2512 (1993)

[13] Abraham, M. H., Whiting, G. S., Doherty, R. M., and Shuely, W. J., *J. Chem. Soc. Perkin Trans.* **3,** 1451–1460 (1990)

[14] Abraham, M. H., Whiting, G. S., Doherty, R. M., and Shuely, W. J., *J. Chromatogr.* **587,** 213–228 (1991)

[15] Abraham, M. H., and Whiting, G. S., *J. Chromatogr.* **594,** 229–241 (1992)

[16] Abraham, M. H., *J. Chromatogr.* **644,** 95–139 (1993)

[17] Abraham, M. H., *J. Phys. Org. Chem.* **6,** 660–684 (1993)

[18] Abraham, M. H., and McGowan, J. C., *Chromatographia* **23,** 243–246 (1987)

[19] Abraham, M. H., Doherty, R. M., Kamlet, M. J., and Taft, R. W., *Chem. Brit.* **22,** 551–554 (1986)

[20] Kamlet, M. J., Doherty, R. M., Abraham, M. H., and Taft, R. W., *Quant. Struct.-Act. Relat.* **7,** 71–78 (1988)

[21] Leahy, D. E., Morris, J. J., Taylor, P. J., and Wait, A. R., *J. Chem. Soc. Perkin Trans.* **2,** 705–722 (1992)

[22] Leahy, D. E., Morris, J. J., Taylor, P. J., and Wait, A. R., *J. Phys. Org. Chem.* **7,** 743–750 (1994)

[23] Dearden, J. C., Bradburne, S. J. A., and Abraham, M. H., The nature of molar refractivity. In: QSAR: *Rational Approaches to the Design of Bioactive Compounds.* Silipo, C., and Vittoria, A. (Eds.). Elsevier: Amsterdam; 143–150 (1991)

[24] Abraham, M. H., Chadha, H. S., Whiting, G. S., and Mitchell, R. C., *J. Pharm. Sci.* **83,** 1085–1100 (1994)

[25] Abraham, M. H., Chadha, H. S., Dixon, J. P., and Leo, A. J., *J. Phys. Org. Chem.* **7,** 712–716 (1994)

[26] Leo, A. J., The Pomona Medicinal Chemistry Project

[27] Abraham, M. H., Duce, P. P., Prior, D. V., Barratt, D. G., Morris, J. J., and Taylor, P. J., *J. Chem. Soc. Perkin Trans.* **2,** 1355–1375 (1989)

[28] Abraham, M. H., Chadha, H. S., and Mitchell, R. C., *J. Pharm. Sci.* 83, 1257–1268 (1994)

[29] Abraham, M. H., Chadha, H. S. and Mitchell, R. C., *J. Pharm. Pharmacol.* **47,** 8–16 (1995)

[30] Abraham, M. H., Lieb, W. R., and Franks, N. F., *J. Pharm. Sci.* **80,** 719–724 (1991)

[31] Seiler, P., *Eur. J. Med. Chem.* **9,** 473–479 (1974)

[32] El Tayar, N., Tsai, R.-S., Testa, B., Carrupt, P. A., and Leo, A., *J. Pharm. Sci.* **80,** 590–598 (1991)

[33] Nahum, A., and Horvath, Cs., *J. Chromatogr.* **192,** 315–322 (1980)

[34] Xie, T. M., Hulthe, B., and Folestad, S., *Chemosphere* **13,** 445–459 (1984)

[35] Minick, D. J., Brent, D. A., and Frenz, J., *J. Chromatogr.* **461,** 177–191 (1989)

[36] Yamagami, C., Ogura, T., and Takao, N., *J. Chromatogr.* **514,** 123–136 (1990)

[37] Yamagami, C., and Takao, N., *Chem. Pharm. Bull.* **40,** 925–929 (1992)

[38] Vallat, P., Van, W., El Tayar, N., Carrupt, P.-A., and Testa, B., *J. Liq. Chromatogr.* **15,** 2133–2151 (1992)

[39] Sadek, P. C., Carr, P. W., Doherty, R. M., Kamlet, M. J., Taft, R. W., and Abraham, M. H., *Anal. Chem.* **57,** 2971–2978 (1985)

[40] Carr, P. W., Doherty, R. M., Kamlet, M. J., Taft, R. W., Melander, W., and Horvath, Cs., *Anal. Chem.* **58,** 2674–2680 (1986)

[41] Kamlet, M. J., Abraham, M. H., Carr, P. W., Doherty, R. M., and Taft, R. W., *J. Chem. Soc. Perkin Trans.* **2,** 2087–2092 (1988)

[42] Park, J. H., Carr, P. W., Abraham, M. H., Taft, R. W., Doherty, R. M. and Kamlet, M. J., *Chromatographia* **25,** 373–381 (1988)

[43] Leahy, D. E., Carr, P. W., Pearlman, R. S., Taft, R. W., and Kamlet, M. J., *Chromatographia* **21**, 473–477 (1986)

[44] Miller, K. G., and Poole, C. F., *J. High Res. Chromatogr.* **17**, 125–134 (1994)

[45] Larrivee, M. L., and Poole, C. F., *Anal. Chem.* **66**, 139–146 (1994)

[46] Abraham, M. H., Chadha, H. S., and Leo, A. J., *J. Chromatogr. A* **685**, 203–211 (1994)

[47] Abraham, M. H., and Roses, M., *J. Phys. Org. Chem.* **7**, 672–684 (1994)

[48] Smith, R. M., and Burr, C. M., *J. Chromatogr.* **485**, 325–340 (1989)

[49] Smith, R. M., and Burr, C. M., *J. Chromatogr.* **550**, 335–356 (1991)

[50] Smith, R. M., and Burr, C. M., *J. Chromatogr.* **475**, 57–74 (1989)

[51] Smith, R. M., and Burr, C. M., *J. Chromatogr.* **475**, 75–83 (1989)

[52] Smith, R. M., and Burr, C. M., *J. Chromatogr.* **481**, 71–84 (1989)

[53] Smith, R. M., and Burr, C. M., *J. Chromatogr.* **481**, 85–95 (1989)

[54] Lambert, W. J., Wright, L. A., and Stevens, J. K, ., *Pharmaceut Res* **7**, 577–586 (1990)

[55] Pidgeon, C., Ong, S., Liu, H., Qui, X., Pidgeon, M., Dantzig, A. H., Munroe, J., Hornback, W. J., Kasher, J. S., Glunz, L., and Szczerba, T., *J. Med. Chem.* **38**, 590–594 (1995)

[56] Christian, S. D., and Scamehorn, J. F. Use of micellas-enhanced ultrafiltration to remove dissolved organics from aqueous streams. In *Surfactant-based Separation Processes*. Scamehorn, J. F., and Harwell, J. H. (Eds.). Marcel Dekker: New York 1989

[57] Nowick, J. S., Chen, J. S., and Noronha, G., *J. Am. Chem. Soc.* **115**, 7636–7644 (1993)

[58] Treiner, C., and Mannebach, M.-H., *J. Colloid Interfac. Sci.* **118**, 243–251 (1987)

[59] Garrone, A., Marengo, E., Fornatto, E., and Gasco, A., *Quant. Struct.-Act. Relat.* **11**, 171–175 (1992)

[60] Abraham, M. H., Chadha, H. S., Dixon, J. P., Rafols, C., and Treiner, C., *J. Chem. Soc. Perkin Trans.* **II**, 1843–1851 (1995)

[61] Leo, A. J., *Chem. Rev.* **93**, 1283–1306 (1993)

[62] Seelig, A., Gottschlich, R., and Devant, R. M., *Proc. Natl. Acad. Sci. USA* **91**, 68–72 (1994)

[63] Heykants, J., Michielis, H., Knaeps, A., and Burgmanns, J., *Arzneim.-Forsch.* **24**, 1649–1653 (1974)

[64] Awouters, F. H. L., Niemegeers, C. J. E., and Janssen, P. A. J., *Arzneim.-Forsch.* **33**, 381–388 (1983)

[65] Gupta, S. P., *Chem. Rev.* **89**, 1765–1800 (1989)

[66] Hansch, C., Bjorkroth, J. P., and Leo, A. J., *J. Pharm. Sci.* **76**, 663–687 (1987)

[67] Hansch, C., Steward, A. R., Anderson, S. M., and Bentley, D. L., *J. Med. Chem.* **11**, 1–11 (1968)

[68] Bradbury, M. W. B., *Exp. Physiol.* **78**, 453–472 (1993)

[69] van de Waterbeemd, H., and Kansy, M., *Chimia* **46**, 299–303 (1992)

[70] Abraham, M. H., and Weathersby P. K., *J. Pharm. Sci.* **83**, 1450–1455 (1994)

[71] Calder, J. A. D., and Ganellin, C. R., *Drug Design and Disc.* **11**, 259–268 (1994)

[72] Chadha, H. S., Abraham, M. H., and Mitchell, R. C., *Bioorg. Med. Chem. Letters* **4**, 2511–2516 (1994)

[73] Scheuplein, R. J., *J. Invest. Dermatol.* **45**, 334–346 (1965)

[74] Scheuplein, R. J., Blank, I. H., Brauner, M. D., and MacFarlane, D. J., *J. Invest. Dermatol.* **52**, 63–70 (1969)

[75] Barry, B. W., *Int. J. Cosmet. Sci.* **10**, 281–293 (1988)

[76] El Tayar, N., Tsai, R. S., Testa, B., Carrupt, P. A., Hansch, C., and Leo, A., *J. Pharm. Sci.* **80**, 744–749 (1991)

[77] Roberts, M. S., Anderson, R. A., and Swarbrick, J., *J. Pharm. Pharmacol.* **29**, 677–683 (1977)

[78] Guy, R. H., and Potts, R. O., *Am. J. Ind. Med.* **23**, 711–719 (1993)

[79] Abraham, M. H., Andonian-Haftvan, J., Whiting, G. S., Leo, A., and Taft, R. W., *J. Chem. Soc. Perkin Trans.* **2**, 1777–1791 (1994)

19 Environmental Hazard Assessment Using Lipophilicity Data

Robert L. Lipnick

Abbreviations

BCF   Bioconcentration factor – ratio of concentration in biotic relative to non-biotic phase (usually water)

EPA   U. S. Environmental Protection Agency

DOM   Dissolved organic matter – organic compounds contained in a dissolved or colloidal state in natural waters that alter the properties of the waters with respect to lipophilic organic chemicals

FATS   Fish acute toxicity syndromes

LOEC   cf. MATC

MATC   Maximum acceptable toxicant concentration, lying between the no observed effect concentration (NOEC) and lowest observed effect concentration (LOEC)

MP   Melting point in degrees centigrade

NOEC   cf. MATC

PCBs   Polychlorobiphenyls

QSAR   Quantitative structure-activity relationships

SAR   Structure-activity relationships

SASA   Solvent-accessible surface area

TSCA   Toxic Substances Control Act

Symbols

EC_{50} Concentration (in aquatic toxicology test) producing 50 % effect response

LC_{50} Concentration (in aquatic toxicology test) producing 50 % mortality

pK_a Negative logarithm to the base 10 of acid dissociation constant

P *1*-octanol/water partition coefficient

S Water solubility

T_e Excess toxicity parameter – ratio of baseline predicted toxicity divided by experimentally observed value

T_{obs} Observed toxicity value

T_{pred} Predicted baseline toxicity value

19.1 Introduction

The passage of the Toxic Substances Control Act (TSCA) [1] in the United States in 1976, and the introduction of similar laws elsewhere for the assessment and control of potential risk from the manufacture, use, and disposal of industrial chemicals [2, 3],

demonstrated the recognition that there are limits to the resources available for testing the chemical and toxicological properties of such new substances. Methods involving the use of quantitative structure-activity relationships (QSAR) and structure-activity relationships (SAR) to predict toxicity have played a significant role in filling data gaps and setting testing priorities within both the United States [4–6] and European Union countries [7].

19.2 Historical Perspective

The history of the development of the relationship of toxicity to lipophilicity and partition coefficient is pertinent to current applications in this field because it reflects efforts to understand the mechanistic basis of both toxicity potency and type of effect, and the parameters or molecular descriptors themselves. For this reason, the relationship has been reviewed in some detail, both in terms of its utility and its limitations.

19.2.1 Nonlinear Relationship to Water Solubility

The correlation of lipophilicity with toxicity potency can be traced to the 1863 thesis of Cros [8] at the Faculty of Medicine at the University of Strasbourg, "Action of Amyl Alcohol on Organisms". Cros measured the toxicity to pigeons, rabbits, and dogs of methyl and amyl alcohol, via oral, inhalational, intraperitoneal, dermal, and intravenous administration. The two alcohols produces similar effects of depression, regardless of the species or route of administration. In each case, death followed at higher doses. Cros noted that the more toxic of the two, amyl alcohol, was also the less soluble, and that with oxygenated refined substances having even lower solubility – such as camphor or oil of absinthe – the potency continued to increase. On the other hand, he found that this trend no longer continued for substances of low solubility – such as palmityl alcohol – which produce no toxic response. From these early observations, Cros in fact demonstrated a nonlinear relationship between toxicity and water solubility for such substances. This loss of toxicity at low water solubility was confirmed for cetyl alcohol by Dujardin-Beaumetz and Audigé [9] in 1876.

19.2.2 Relationship of Toxicity to Chain Length and Molecular Weight

Unaware of the work of Cros, Richardson in 1869 [10] in England reported a similar relationship for methyl, ethyl, propyl, butyl, amyl, and capryl alcohols dosed to guinea pigs, pigeons, rabbits, and frogs by oral, subcutaneous, and inhalation administration. Four years later in France, Rabuteau [11] came to a similar conclusion[1] by immersing frogs in aqueous solutions of ethyl, butyl, and amyl alcohols. Richardson's work was

---

[1] De quelques propriétés nouvelles ou peu connues de l'alcool ethylique; déductions thérapeutiques de ces propriétés – des effets toxiques des alcools butylique et amylique – application à l'alcoolisation du vin improprement appelée vinage.

cited by later investigators, with the relationship between chain length and toxicity being referred to as Richardson's law [12]. Cros' far more important finding of a nonlinear relationship of toxicity to solubility was virtually unknown in the literature.

19.2.3 Chemical Constitution Theory of Hypnotic Activity

Bauman and Kast [13] studied the hypnotic activity of sulfones and concluded that potency was a function of their chemical constitution, including the number of ethyl groups present, as opposed to physico-chemical properties, and that potency results from metabolic release of toxicants.

19.2.4 Richet's Law

In 1893, Georges Houdaille [14], a student of Richet at the University of Paris, published in his medical thesis careful studies of the activity of various hypnotic agents to fish, which he found to vary inversely with water solubility. This relationship is referred to as Richet's law [15].

19.2.5 Development of the Lipoid Theory of Narcosis

In 1899, Hans Horst Meyer [16] and Fritz Baum [17] at the University of Marburg demonstrated that narcotic or hypnotic activity was related to neither chemical constitution nor water solubility, but to the relative affinity of substances to water and lipid sites within the organism. At the same time, working alone at the University of Zürich, Ernest Overton [18] independently came to the same conclusion, providing substantial documentation two years later, in a book [19] that has been widely cited as a classic in this field.

Meyer's student Diehl [20], using as an endpoint the minimum concentration needed to anesthetize tadpoles and fish, tested 14 sulfones to assess the Bauman and Kast chemical constitution theory of narcosis potency. Contrary to Bauman and Kast, he discovered that narcotic potency was unrelated to the number of ethyl groups, nor was metabolism involved since the effect was quickly reversed when the organisms were returned to clean water.

Bucholz [21], also working in Meyer's laboratory, showed in 1895 that a wide variety of nonelectrolytes produce a narcotic response, but that a certain degree of solubility in both water and fat is required. He demonstrated that in the case of acetamide, metabolism is responsible for non-narcotic effects which tend toward narcosis with increasing chain length and slower metabolism.

In opposition to Richet's law, Meyer's student Dunzelt [22] discovered two instances in which water solubility did not correlate with narcotic potency. Bromal hydrate was found to be more potent, and methyl urethane less potent than predicted. Dunzelt's discovery of these exceptions lay the groundwork for the discovery of the lipoid theory.

In 1899, Fritz Baum published measurements of the correlation of partition coefficient with narcotic potency, providing the basis for Meyer's paper on the theory itself. Baum measured the olive oil/water partition coefficients of 11 compounds previously

studied for narcotic potency with tadpoles in Meyer's laboratory. To measure these co-efficients, he agitated the compounds between the two phases, and determined the concentration remaining in the aqueous phase by one of the following methods: 1) extraction and evaporation; 2) freezing point depression; 3) density; 4) Kjeldahl nitrogen measurement; or 5) saponification. Bromal hydrate and methyl urethane, which did not fit the water solubility correlation (Richet's law), were no longer outliers with respect to the excellent rank correlation that Baum reported between the narcotic concentration and measured partition coefficients.

Unlike Meyer, who worked in a pharmacological laboratory, Overton began the research that led to his independent discovery working as a botanist, attempting to understand what aspects of a substance's constitution are responsible for its ability to permeate plant cells and the rate at which it does so [23, 24].

Overton selected olive oil as a surrogate for what he considered to be a lipoid site of action because of its ready availability and ease of purification. In addition, there was no problem separating the oil and water phases following shaking. By contrast, this was not the case with what he considered to be more realistic models such as cholesterol.

19.2.6 QSAR and More Quantitative Use of Lipophilicity Data

Pioneering work in the early 1960s led to the development of quantitative structure-activity relationships (QSARs), which are being applied increasingly as a tool in hazard assessment of industrial organic chemicals. Corwin Hansch and coworkers [25] at Pomona College, Claremont, California employed octanol/water as a standard lipophilic partitioning system, demonstrating that this parameter on a logarithmic scale could be estimated by a simple additivity scheme, and that regression analysis would provide a statistical means of treating the development of QSAR models.

19.3 Toxicological Applications

Most applications of lipophilicity both past and present are related to the development of new drugs and attempts to understand their mechanism of action. The importance of lipophilicity in toxicology, from the standpoint of quantitative modeling, continues to be less well developed, particularly in the area of mammalian toxicology.

19.3.1 Contributions of Lazarev

The earliest application of lipophilicity data for hazard assessment appears to be by Lazarev [26] in St. Petersburg, who, beginning in the early 1930s, applied the findings of Overton and Meyer to predictive toxicology. Lazarev's approach to assessing the toxicological properties of industrial organic chemicals was first organized in a 1944 book [27], in which a number of ideas were presented such as:

1. Defining partition coefficients on a logarithmic scale for classification into groups.
2. Plotting partition coefficients as their logarithmic group values against chain length, observing the effect of increasing numbers of carbon on chain length, and demon-
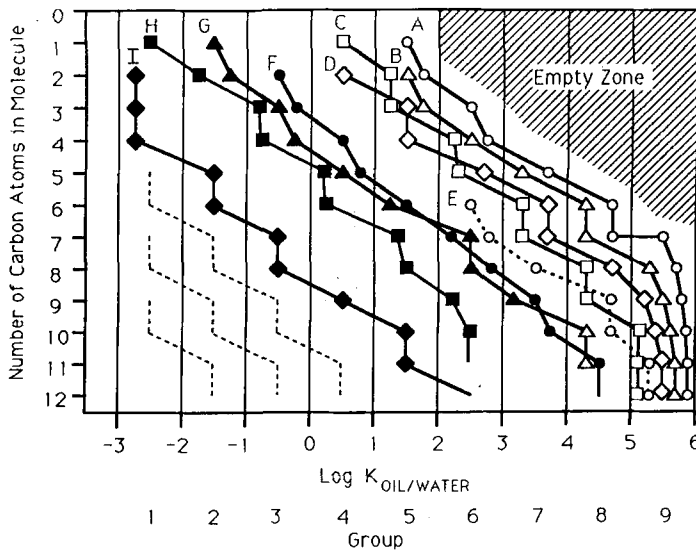
Figure 1. Effect of increasing number of carbon atoms within homologous series on Lazarev's group number of the olive oil/water partition coefficient. (○-○ alkanes; (△) alkenes; (□) alkynes; (◇) phenyl-substituted; (○---○) monochloro-substituted; (●) saturated esters of fatty acids; (▲) primary unbranched monohydric alcohols; (■) saturated fatty acids; (◆) saturated dibasic acids. The "empty zone" refers to that domain in which no structures were known to satisfy these conditions. (Reproduced from [26], Table 3 by permission from Elsevier.)

Table 1. Correlations of Lazarev with increasing partition coefficient or decreasing water solubility. (Adapted from [23].)

Effect	Subject
Increase in degree of irritancy of organic liquids	Skin
Increase in degree of reversible aggregation of liquid particles	Coacervate emulsion (phospholipid and oleate)
Decrease in concentration needed to produce a 6 % reduction in staining	Fixed frog gastrocnemius
Decrease in concentration needed	*In vitro* hemolysis
Decrease in concentration required	50 % Reduction of bird erythrocyte respiration
Decrease in concentration	Arrest of isolated frog heart
Decrease in minimum concentration	Contraction changes in isolated segments from heart ventricle
Decrease in concentration	Paralyzing action on isolated rabbit intestine
Decrease in concentration	Narcosis in tadpoles and small fish
Decrease in concentration in blood of mammals	Change in reflex time, narcosis, respiratory failure, or death
Decrease in blood concentration	Respiratory failure in frog
Decrease in concentration	Irritation of the eye or tongue
Decrease in concentration	Anesthesia via intradermal administration

strating that within each series, there was a systematic increase of olive oil/water partition coefficient with chain length. (Fig. 1)

3. Development of equations relating one partition coefficient to a second partition coefficient, as well as to water solubility on a double logarithmic scale,

$$\log P_{benzene/water} = 1.05 \log P_{oil/water} + 0.32 \tag{1}$$

$$\log S = -0.89 \log P_{oil/water} + 3.15 \tag{2}$$

4. Demonstration that a wide variety of physiological effects are correlated with partition coeffcient (Table 1), and can be expressed mathematically as an equation such as,

$$\log C_{tadpole\ narcosis} = -0.69 \log P_{oil/water} + 1.06 \tag{3}$$

19.3.2 Development of QSAR in Aquatic Toxicology

Following work by Hansch and coworkers that laid the foundation for QSAR, work was reported from this laboratory on numerous such linear correlations, including those for the toxicity to five species of fish (carp, goldfish, goby, roach, and tench) based upon literature data [28] on simple alcohols. For example, for goldfish, where C is the 24-h our minimum lethal concentration (in mol l^{-1}):

$$\log (1/C) = 0.881 \log P - 5.011$$

$$n = 5, r = 0.958, sd = 0.250, F \text{ not given} \tag{4}$$

Almost a decade later, apparently unaware of Hansch's work, Könemann [29] in The Netherlands reported a correlation between log P and the 7- to 14-day LC_{50} for the guppy (*Poecilia reticulata*):

$$\log(1/C) = 0.871 \log P - 4.87 \tag{5}$$

$$n = 50, r = 0.988, sd = 0.237, F \text{ not given}$$

A number of similar QSAR equations [30–36] for the toxicity of nonreactive non-electrolytes as illustrated in Table 2 have appeared in the literature and proven useful for hazard assessment for such chemicals for which little or no test data are available.

Table 2. QSAR equations for the toxicity of nonreactive nonelectrolytes to aquatic organisms.

Organism	Endpoint effect	Reference
Selanastrum capricornutum (algae)	Growth inhibition (96-h EC_{50})	[30]
Leuciscus idus melanotus (golden orfe)	Lethal concentration (48-h LC_{50})	[31, 32]
Pimephales promelas (fathead minnow)	Lethal concentration (96-h LC_{50})	[33]
Pimephales promelas (fathead minnow)	Subchronic effects (32-day MATC)	[34]
Daphnia magna (water flea)	Immobilization (48-h EC_{50})	[35]
Daphnia magna (water flea)	Chronic growth (16-day NOEC)	[36]

Success has also been shown with the use of neural networks to develop QSARs for the toxicity of 50 alkylated and halogenated phenols to *Tetrahymena pyriformes* (concentration for 50 % growth inhibition) using 15 global electronic and steric molecular descriptors [37]. However, since these data had already been modeled using only log P and pK_a as descriptors with a high-quality model, the rationale for this approach for such simple compounds, except for developmental purposes, seems questionable.

19.3.3 Water Solubility and Pharmacokinetic Cutoff: QSAR Limitations

For simple nonelectrolytes acting by narcosis, one can expect to observe toxicity, provided that the required narcotic concentration in water does not exceed its water solubility. Algae and other plant cells require higher concentrations of simple nonelectrolytes to produce a depressant response than do tadpoles and other organisms with nervous systems. Because of this difference in sensitivity, certain substances may produce effects, on animals but not algae and other plant cells in which the needed toxic concentration exceeds available water solubility. In fact, as reported by Overton, the existence of such a difference in the behavior of plant and animal cells toward such substances has been known for more than a century. However, the variation was originally attributed to a difference in mechanism of action, in which those substances able to affect both plant and animal cells were termed anesthetics, but substances which acted only upon animal ganglia cells were considered narcotics. Overton observed that concentrations 6 to 10 times greater were required to produce effects on plant cells.

The concentration needed to produce an observed effect can be estimated using a QSAR model such as one of the above. If the log P value and melting point (MP) of a candidate chemical are available, the water solubility can be estimated [38], where MP is in °C (for liquid solutes, a nominal value of 254 °C is used); S is the water solubility in μmol l^{-1}, and P is the n-octanol/water partition coefficient:

$$\log P = 6.5 - 0.89 \log S - 0.015 \text{ MP} \tag{6}$$

$n = 27, r = 0.96$, sd not given, F not given

Solution of this equation for substances that are liquids at room temperature indicates a cutoff at log P of about 8. In practice, at such a high value, another cutoff resulting from insufficient test duration to achieve steady state between the external and internal site of action is expected to dominate. For higher melting compounds, correspondingly lower cutoffs will be observed.

19.3.4 Additive Effect of Toxicants

While so-called narcotics such as sulfonal produce no effect on algal cells at saturation, Overton found that their presence reduced the required concentration of a second substance as a result of their additive contribution. Overton's findings on additive behavior of toxicants have been confirmed by more recent aquatic toxicology studies [39–41]. In fact, for a mixture of a large number of toxicants acting by a variety of

mechanism, concentration addition is observed consistent with a narcosis mechanism for each contributing chemical. Thus, the individual concentrations are too low to exert their specific responses, and only baseline narcosis is expressed [42]. The degree to which a contaminated water sample is likely to approach baseline narcosis through mixture additivity can be estimated by "counting the number of molecules" collected in a simulated biphase using vapor pressure depression [43].

The toxicity of simple nonreactive nonelectrolytes to aquatic organisms can be modeled using a bilinear equation. The toxicity of such chemicals to the fathead minnow was found to be related by the following, where P is the n-octanol/water partition coefficient:

$$\log 1/LC_{50} = 0.94 \log P - 0.94 \log (0.000068\ P + 1) - 1.25 \tag{7}$$

$$n = 65,\ r^2 = 0.999,\ \text{sd not given},\ F\ \text{not given}$$

19.3.5 Bioconcentration

A fish or other aquatic organism exposed to a lipophilic xenobiotic chemical in an aquatic environment tends to bioconcentrate that substance in proportion to its relative lipophilicity. In a laboratory study designed to simulate environmental behavior, bioconcentration is followed from several days to several months, or longer, depending upon the rate at which equilibrium steady state is achieved. For so-called superlipophilic chemicals, in which $\log P$ is > 6, uptake requires increasingly long periods with increasing partition coefficient [44, 45], and occurs increasingly via dietary exposure as opposed to direct uptake via the gills. In addition to test duration, bioconcentration may be limited by one or more of the following: biotransformation to less lipophilic substances, insufficient solubility in lipids, and molecular size and cross-section. The bioconcentration factor (BCF) organic chemicals in fish (with these of limitations in mind) can be estimated using a QSAR based upon a large number of tested substances [46],

$$\log BCF = 0.79 \log P - 0.40 \tag{8}$$

$$n = 122,\ r^2 = 0.86,\ \text{sd not given},\ F\ \text{not given}$$

Estimating the true bioavailability of a lipophilic organic compound in natural water is complicated by the presence of dissolved organic matter (DOM) consisting of humic acid, carbohydrates and proteins. Thus, mirex (perchloropentacyclo[5.2.1.0$^{2.6}$. 0^{3,9}.0$^{5.8}$]decane, $C_{10}Cl_{12}$) and seven other chlorinated hydrocarbons subjected to liquid-liquid extraction in the presence of humic acid or other DOM were only incompletely extracted until the DOM was destroyed by chromic acid oxidation [47]. Such recoveries were found to decrease with increasing DOM. Furthermore, it has been demonstrated [48] using filtered Niagara River water that spiked chlorinated chemicals are recovered more poorly with increasing lipophilicity. Since the DOM-bound lipophilic chemicals are similarly less available for bioconcentration in fish and other aquatic organisms, it is important to understand what measured concentrations actually represent for predictive purposes.

19.3.6 Thermodynamic Approaches

Ferguson [49] provided a thermodynamic interpretation for the correlations of partition coefficient with biological activity, based upon the concept that narcotic chemicals produce an equivalent effect at some site of action. Like Overton, Ferguson distinguished between narcotic acting substances and ones exhibiting more specific action. He concluded that the former group act by means of "...adsorption of the substance on certain cell structures, or solution in the lipids, or coagulation of cell proteins..," and noted that "all substances apparently exert a physiological action by a physical mechanism... [which], however, may be, and frequently is, masked by a specific chemical effect." We shall see how Ferguson's observation is important in toxicological assessment in the identification of mechanisms of action and the corresponding appropriate model for predictive purposes.

19.3.7 Excess Toxicity as a Measure of a Specific Mechanism of Action

The prediction of toxicity of a simple nonelectrolyte industrial organic compound by a QSAR, based upon test data for similar substances such as monohydric alcohols which act by a narcosis mechanism represents a baseline, or minimum toxicity prediction. Chemicals for which toxicity is not water solubility-limited or limited by insufficient test duration will exhibit toxicity at a concentration at least as low as that predicted by the baseline narcosis model. On the other hand, those substances showing effects at lower concentrations (more toxic) than baseline toxicity can be considered to be acting by more specific mechanisms. For the sake of such comparisons, a parameter, *excess toxicity* (T_e), can be defined [50] as,

$$T_e = T_{obs}/T_{pred} \tag{9}$$

where T_{obs} is the observed toxic concentration, and T_{pred} is the concentration predicted to produce the defined toxic response by a narcosis or baseline mechanism. The calculation of excess toxicity values provides a means of identifying categories of toxicants and assigning a putative molecular mechanism of action that can serve as a hypothesis for further experiments. Compounds showing excess toxicity include electrophiles, proelectrophiles, and cyanogenic toxicants.

Various schemes have been presented to use chemical structure moieties to classify compounds by mechanism of action [51–53]. In addition, work at the EPA Environmental Research Laboratory in Duluth, Minnesota, resulted in the development of Fish Acute Toxicity Syndromes (FATS), based upon the use of principal components analysis and clustering of a set of physiological responses by the dosed fish [54–57]. This work provided the rationale for distinguishing between narcosis and "polar narcosis", and for defining the responses associated with respiratory irritancy, uncoupling of oxidative phosphorylation, acetylcholinesterase inhibition, and central nervous system seizure agents.

Table 3. Classification of electrophile toxicity mechanisms and comparison of excess toxicities. (Adapted from [42].)

Compound	Mechanism type	T_e (excess toxicity)[a]	log P	Species[b]
1,4-Dibromo-butene	Allylic activation	> 304	1.97	F
		61		RO
3-Chloro-propyne	Propargylic activation	> 340	0.59	F
α,α-Dichloro-xylene	Benzylic activation	86	3.87	F
Chloracetonitrile	α-Halo activation	1960	0.22	F
		32		RO
Succinic anhydride	Acid anhydride activation	> 8530	− 0.87	F
Ethylene oxide	Strained 3-membered heterocyclic ring (epoxide)	490 54	− 0.79	F RO
N-Vinylethyleneimine	Strained 3-membered heterocyclic ring (imine)	69	0.26	RO
Acrolein	Michael-type addition (C=C−C=O moiety)	> 81 000 134	0.10	F RO
1,4-Napthoquinone	Michael-type addition (quinoid structure)	> 3800	0.9	F
Acrylonitrile	Michael-type addition (C=C−C≡N type)	68	0.05	RO
Divinyl sulfone	Michael-type addition	1066	− 0.56	RO
Pentafluorobenzaldehyde	Schiff base formation	51	2.45	F

[a] See Eq. (7).
[b] F, toxicity to fish as effect concentration; RO, rat oral LD_{50}

19.3.7.1 Electrophile Toxicants

Electrophiles act by direct covalent bond formation with sulfhydryl and other nucleophilic moieties present on enzymes and other biochemical sites, resulting in loss of biological activity. Examples of electrophile toxicants [58] are provided in Table 3. Electrophilic behavior requires the ability to undergo either displacement of an activated heteroatom moiety or direct addition reaction. The latter can take place via 1,4-conjugate Michael-type addition or Schiff base formation. The assessment of the potential toxicity of untested electrophile toxicants has been facilitated by the availability of QSAR models using measured [59, 60] or calculated reactivity descriptors to model degree of electrophilicity in addition to lipophilicity, as has been done in the case of phosphorothionates [61], organophosphates [62], and epoxides [63–64]. A wide variety of chemical structure types show electrophilic behavior [65].

19.3.7.2 Proelectrophile Toxicants

While the activity of electrophiles can be observed readily with *in vitro* experiments, proelectrophiles require metabolic activation to a corresponding electrophile, for example, via the enzymes alcohol dehydrogenase, monooxygenase, and glutathione

Table 4. Classification of proelectrophile toxicity mechanisms and comparison of excess toxicities. (Adapted from reference [42][a].)

Compound	Mechanism type	T_e	log P	Species
Allyl alcohol	Alcohol dehydrogenase activation (to α, β-unsaturated aldehyde)	16 000 166	−0.25	F RO
3-Butyl-2-ol	Alcohol dehydrogenase activation (to α,β-unsaturated ketone)	383	−0.06	F
3-Butyl-1-ol	Alcohol dehydrogenase activation (with rearrangement to allene conjugated with aldehyde carbonyl)	321	−0.50	F
Pentaerythritol triallyl ether	Monooxygenase activation (to α,β-unsaturated aldehyde)	18 000	−1.60	F
1,3-Dibromopropane	Glutathione transferase activation (to activated 4-membered ring sulfonium derivative)	87	1.99	F

[a] See footnotes to Table 3.

transferase, as illustrated in Table 4. Propargylic alcohols provide an instructive example to illustrate this behavior. Study of the toxicity to fish of a series of propargylic alcohols demonstrated [66] that those in which the hydroxyl group is tertiary produce solely narcosis effects, and fit a simple lipophilicity QSAR model. In contrast, secondary and tertiary propargylic alcohols exhibit excess toxicity, with more specific toxic effects. The role of alcohol dehydrogenase in this activation step has been confirmed by showing that narcosis action alone is produced in the presence of an inhibitor of this enzyme [67]. QSAR models based upon quantum mechanical electronic descriptors appear to account for this metabolic transformation and resulting reactivity of the corresponding electrophile products, in relation to observed toxicity [68], and provide a means of predicting the behavior of untested members of this class.

19.3.7.3 Cyanogenic Toxicants

Cyanogenic toxicants can act by either a direct hydrolytic release of cyanide, as in the case of cyanohydrins such as lactonitrile, or via metabolic activation such as monooxygenase activation of malonitrile to an unstable cyanohydrin intermediate, as illustrated in Table 5.

Table 5. Classification of cyanogenic toxicity mechanisms and comparison of excess toxicities. (Adapted from reference [42][a].)

Compound	Mechanism type	T_e	log P	Species
Lactonitrile	Hydrolysis (unstable addition product of cyanide and acetaldehyde)	361 23 800	−0.85	RO F
Malononitrile	Monooxygenase oxidation to cyanohydrin and hydrolysis to cyanide and formic acid	88 700	−1.20	F

[a] See footnotes to Table 3.

19.4 Biodegradation

The potential environmental effects of a chemical are dependent upon its relative toxicity and bioconcentration, as well as on the probability that it will survive microbial action in a waste treatment plant. Although the site and rate of biodegradation are dependent upon bond-breaking processes and therefore electronic influences, lipophilicity controls the rate of uptake into the bacteria. Thus, within simple series of compounds including alcohols and ketones, log P was found to play an important role with the development of QSAR models [69, 70]. In the case of halogenated aromatic and aliphatic compounds, a good correlation with rate of biodegradation could be obtained with the use of electronic and steric parameters [71]. Schüürmann and Müller [72] have examined the utility of back-propagation neural networks for prediction of biodegradation kinetics for 26 organic compounds using fragment count descriptors, and demonstrated the utility of the leave-n-out procedure.

Klopman and coworkers [73] have developed a computer program META that employs an expert system for recognizing likely molecular fragment sites of attack and predicting most likely metabolites.

19.5 Outlook

The ability to predict the acute toxicity of simple industrial organic chemicals acting by a narcosis mechanism is now considered to be very reliable using lipophilicity (log P) as the sole molecular descriptor. Log P is the preferred measure of lipophilicity, since most of the derived predictive models are based upon it. In addition, it offers the largest database of measured values of such descriptors, as well as the ability to be estimated directly from chemical structure by a variety of different methods. New methods add further to the utility of this parameter. Thus, the slow-stirring method [74] has provided a means of direct measurement of extremely lipophilic substances beyond the capability of shake-flask studies.

More global approaches are now available to estimate log P values. Thus, for a series of polychlorobiphenyls (PCBs), an excellent correlation was found between solvent-accessible surface area (SASA) and log P [75],

$$\log P = 0.0270 \, \text{SASA} - 7.12 \tag{10}$$

$$n = 18, \, r^2 = 0.986, \, s = 0.12, \, F = 1116$$

In the future, increasing emphasis will be placed on understanding mechanistic details behind these and other correlations to provide guidance to the predictive assessment of less routine chemicals. As an example, pharmacokinetic modeling is used increasingly in studies of bioconcentration of lipophilic organic compounds in fish [76].

References

[1] Toxic Substances Control Act, U.S. Public Law 94–469, October 1, 1976

[2] ECETOC: *Structure-Activity Relationships in Toxicology and Ecotoxicology: An Assessment*, European Chemical Industry Ecology and Toxicology Centre, Brussels, February 24, 1986, Monograph 8, 86 pp.

[3] Hart, J. W., *Sci. Total Environ.* **109/110**, 629–633 (1991)

[4] Lipnick, R. L., *Environ. Toxicol. Chem.* **4**, 255–257 (1985)

[5] Auer, C. M., Zeeman, M., Nabholz, J. V., and Clements, R. G., *QSAR SAR Environ. Res.* **2**, 29–38 (1994)

[6] Zeeman, M. and Gilford, J., Ecological hazard and risk assessment under EPA's Toxic Substances Control Act (TSCA): an introduction. In: *Environmental Toxicology and Risk Assessment, ASTM STP 1179*, Landis, W. G., Hughes, J. S., and Lewis, M. A., eds., American Society for Testing and Materials, Philadelphia (1993), p. 7–21

[7] Verhaar, H. J. M., Van Leeuwen, C. J., Bol, L., and Hermens, J. L. M., *SAR QSAR Environ. Res.* **2**, 39–58 (1994)

[8] Cros, A. F. A., *Action de l'alcool amylique sur l'organisme*. Thesis, Faculté de Médecine, Strasbourg, Strasbourg, 1863

[9] Dujardin-Beaumetz and Audigé. *Comptes Rendus Hebdomadaires des Séances de L'Académie des Sciences* **83**, 192–194, 1876

[10] Richardson, B. W., Physiological research on alcohols. *Medical Times and Gazette (London)* **2**, 703–706, December 18, 1869

[11] Rabutea., *L'Union Medicale*, [3rd ser.], **10**, 165–173 (1870)

[12] Munch, J. C., and Schwartze, E. W., *J. Lab. Clin. Med.* **10**, 985–996 (1925)

[13] Baumann, E. and Kast, A., *Z. Physiol. Chem.* **14**, 52–74 (1890)

[14] Houdaille, G., *Étude Expérimentale et Critique sur les Nouveaux Hypnotiques*. Thesis, Le Doctorat en Médecine, Faculté en Médecine de Paris, Paris, 1893

[15] Richet, C., *Comptes Rendus Societé Biologie Paris* **54**, 775–776 (1893)

[16] Meyer, H., *Naunyn-Schmiedebergs Arch. Exp. Pathol. Pharmakol.* **42**, 109–118 (1899)

[17] Baum, F., *Naunyn-Schmiedebergs Arch. Exp. Pathol. Pharmakol.* **42**, 119–137 (1899)

[18] Overton, E., *Vierteljahrsschr. Naturforsch. Ges. Zürich* **44**, 88–135 (1899)

[19] Overton, E., *Studien über die Narkose*, Gustav Fischer, Jena, 1901; English translation: *Studies of Narcosis*, R. L. Lipnick, ed., Chapman & Hall, London, 1991

[20] Diehl, D., *Vergleichende Experimentaluntersuchungen über die Stärke der narkotischen Wirkung einiger Sulfone, Säureamide und Glycerinderivate*, Inaugural Dissertation, University of Marburg, 1894

[21] Bucholz, F. A., *Beiträge zur Theorie der Alkoholwirkung*, Inaugural Dissertation, University of Marburg, 1895

[22] Dunzelt, W., *Vergleichende Experimentaluntersuchungen über die Stärke der Wirkung einiger Narcotica*, Inaugural Dissertation, University of Marburg, 1896

[23] Overton, E., *Vierteljahrsschr. Naturforsch. Ges. Zürich* **40**, 159–201 (1895)

[24] Overton, E., *Vierteljahrsschr. Naturforsch. Ges. Zürich* **41**, 383–406 (1896)

[25] Martin, Y. C., *Quantitative Drug Design,* Marcel Dekker, New York, 1978

[26] Lipnick, R. L., and Filov, V. A., *Trends Pharmacol. Sci.* **13**, 56–60 (1992)

[27] Lazarev, N. V., *Neelektrolity* [Nonelectrolytes], Voennomorskaya Medicinskaya Akademiya, Leningrad, 1944

[28] Hansch, C., and Dunn, W. J. III., *J. Pharm. Sci.* **61**, 1–19 (1972)

[29] Könemann, H., *Toxicology* **19**, 223–228 (1981)

[30] Calamari, D., Galassi, S., Setti, F., and Vighi, M., *Chemosphere* **12**, 253–262 (1983)

[31] Lipnick, R. L. and Dunn, W. J., A MLAB Study of aquatic structure-toxicity relationships.

In: *Quantitative Approaches to Drug Design,* Dearden, J. ed., Elsevier, Amsterdam (1983) p. 265–266

[32] Lipnick, R. L., Validation and extension of fish toxicity QSARs and interspecies comparisons for certain classes of organic chemicals. In: *QSAR in Toxicology and Xenobiochemistry,* Tichý, M. ed., Elsevier, Amsterdam (1985), p. 39–52

33] Veith, G. D., Call, D. J., and Brooke, L. T., *Can. J. Fish. Aquat. Sci.* **40,** 743–748 (1983)

[34] Call, D. J., Brooke, L. T., Knuth, M. L., Poirer, S. H., and Hoglund, M. D., *Environ. Toxicol. Chem.* **4,** 335–341 (1985)

[35] Hermens, J., Canton, H., Janssen, P., and De Jong, R., *Aquat. Toxicol.* **5,** 143–154 (1984)

[36] Hermens, J., Broekhuyzen, E., Canton, H., and Wegman, R., *Aquat. Toxicol.* **6,** 209–217 (1985)

[37] Xu, L., Ball, J. W., Dixon, S. L., and Jurs, P. C., *Environ. Toxicol. Chem.* **13,** 841–851 (1994)

[38] Banerjee, S., Yalkowsky, S. H., and Valvani, S. S., *Environ. Sci. Technol.* **14,** 1227–1229 (1980)

[39] Könemann, H., *Ecotox. Environ. Saf.* **4,** 415–421 (1980)

[40] Könemann, H., *Toxicology* **19,** 209–221 (1981)

[41] Broderius, S. J., and Kahl, M., *Aquat. Toxicol.* **6,** 307–322 (1985)

[42] Deneer, J. W., Sinnige, T. L., Seinen, W., and Hermens, J. L. M., *Aquat. Toxicol.* **12,** 33–38 (1988)

[43] Verhaar, H. J. M., Busser, J. M., and Hermens, J. L. M., *Environ. Sci. Technol.* **29,** 726–734 (1995)

[44] Bruggeman, W. A., Opperhuizen, A., Wijbenga, A., and Hutzinger, O., *Toxicol. Environ. Chem.* **7,** 173–189 (1984)

[45] Hawker, D. W., and Connell, D. W., *Chemosphere* **14,** 1205–1219 (1985)

[46] Veith, G. D., and Kosian, P., Estimating bioconcentration potential from octanol/water partition coefficients. In: *Physical Behavior of PCBs in the Great Lakes,* Mackay, D. (Ed.). Ann Arbor Science: Ann Arbor; 269–282 (1982)

[47] Maguire, R. J., Batchelor, S. P., and Sullivan, C. A., *Environ. Toxicol. Chem.* **14,** 389–393 (1995)

[48] Driscoll, M. S., Hassett, J. P., Fish, C. L., and Litten, S., *Environ. Sci. Technol.* **25,** 1432–1439 (1991)

[49] Ferguson, J., *Proc. Roy Soc. (Lond.) Ser. B* **127,** 387–404, (1939)

[50] Lipnick, R. L., Watson, K. R., and Strausz, A. K., *Xenobiotica* **17,** 1011–1025 (1987)

[51] Verhaar, H. J. M., Van Leeuwen, C. J., and Hermens, J. L. M., *Chemosphere* **25,** 471–491 (1992)

[52] Bradbury, S. P., and Lipnick, R. L., *Environ. Health Persp.* **87,** 181–182 (1990)

[53] Lipnick, R. L., *Sci. Total Environ.* **109/110,** 131–153 (1991)

[54] McKim, J. M., Schmieder, P. K., Carlson, R. W., Hunt, E. P., and Niemi, G. J., *Environ. Toxicol. Chem.* **6,** 295–312 (1987)

[55] McKim, J. M., Schmieder, P. K., Niemi, G. J., Carlson, R. W. and Henry, T. R., *Environ. Toxicol. Chem.* **6,** 313–328 (1987)

[56] Bradbury, S. P., Henry, T. R., Niemi, G. J., Carlson, R. W., and Snarski, V. M., *Environ. Toxicol. Chem.* **8,** 247–261 (1989)

[57] Bradbury, S. P., Carlson, R. W., Niemi, G. J., and Henry, T. R., *Environ. Toxicol. Chem.* **10,** 115–131 (1991)

[58] Lipnick, R. L., *Sci. Total Environ.* **109/110,** 131–153 (1991)

[59] Hermens, J., De Bruijn, J., Pauly, J., and Seinen, W., QSAR studies for fish toxicity data of organophosphorus compounds and other classes of reactive organic compounds. In: *QSAR in Environmental Toxicology-II.* Kaiser, K. L. E. (Ed.). Reidel: Dordrecht, 135–152 (1987)

[60] De Bruijn, J., and Hermens, J., *Aquat. Toxicol.* **24,** 257–274 (1993)

[61] Schüürmann, G., *Environ. Toxicol. Chem.* **9,** 417–428 (1990)

[62] Verhaar, H. J. M., Eriksson, L., Sjöstrom, M., Schüürmann, G., Seinen, W., and Hermens, J. L. M., *Quant. Struct.-Act. Relat.* **13,** 133–143 (1994)

[63] Purdy, R., *Sci. Total Environ.* **109/110** 553–556 (1991)

[64] Eriksson, L., Verhaar, H. J. M., and Hermens, J. L. M., *Environ. Toxicol. Chem.* **13,** 683–691 (1994)

[65] Hermens, J., *Environ. Health Persp.* **87,** 219–225 (1990)

[66] Veith, G. D., Lipnick, R. L., and Russom, C. L., *Xenobiotica* **19,** 555–565 (1989)

[67] Bradbury, S. P., and Christensen, G. M., *Environ. Toxicol. Chem.* **10,** 1155–1160 (1991)

[68] Mekenyen, O. G., Veith, G. D., Bradbury, S. P., and Russom, C. L., *Quant. Struct.-Act. Relat.* **12,** 132–136 (1993)

[69] Vaishnav, D. D., Boethling, R. S., and Babeu, L., *Chemosphere* **16,** 695–703 (1987)

[70] Paris, D. F., Wolfe, N. L., and Steen, W. C., *Appl. Environ. Microbiol.* **47,** 7–11 (1984)

[71] Peijnenburg, W. J. G. M., Hart, M. J. T., den Hollander, H. A., van de Meent, D., Verboom, H. H., and Wolfe, N. L., *Sci. Tot. Environ.* **109/110,** 283–300 (1991)

[72] Schüürmann, G., and Müller, E., *Environ. Toxicol. Chem.* **13,** 743–747 (1994)

[73] Klopman, G., Zhang, Z., Balthasar, D. M., and Rosenkranz, H. S., *Environ. Toxicol. Chem.* **14,** 395–403 (1995)

[74] De Bruijn, J., Busser, F., Seinen, W., and Hermens, J., *Environ. Toxicol. Chem.* **8,** 499–512 (1989)

[75] De Bruijn, J., and Hermens, J., *Quant. Struct.-Act. Relat.* **9,** 11–21 (1990)

[76] Lien, G. J., Nichols, J. W., McKim, J. M., and Gallinat, C. A., *Environ. Toxicol. Chem.* **13,** 1195–1205 (1994)

20 Lipophilicity in Peptide Chemistry and Peptide Drug Design

Jean-Luc Fauchère

Abbreviations

Abbreviations for amino acids and peptides are in accordance with recommendations of the IUPAC-IUB Joint Commission on Biochemical Nomenclature (see [97]).

ACE   Angiotensin I-converting enzyme
CLOGP   Calculated log P (fragment method)
CNS   Central nervous system
HPLC   High-performance liquid chromatography
QSAR   Quantitative structure-activity relationships

Symbols

π Amino acid side chain lipophilicity; for definition, see legend of Table 1 and [13].
Λ Amino acid side chain polarity; for definition, see [19].
P Partition coefficient (one single species)
D P at a given pH (several ionic forms may be present)

20.1 Introduction

The hydrophobic effect [1], i.e., the tendency of a solute to expose the smallest possible hydrophobic surface to the solvent (ordered water molecules), is an important factor in drug design, since drug action very often implies the transport of the active species from an aqueous fluid to a less hydrophobic environment such as an embedded membrane receptor. Natural peptides are relatively polar molecules due to the amide linkages and the presence of a number of ionizable side chains. Peptide hormones and neurotransmitters act as endogenous drugs which are recognized by specific (mostly G-protein-coupled) receptors and which produce strong and short effects, before being inactivated into nontoxic metabolites. Pseudopeptides and peptide mimetics act on the same receptors, either as agonists or antagonists, although their access to the receptor strongly depends on the exogenous site of administration and often implies crossing of physiological barriers. Obviously, the lipophilic properties of peptide and peptide-derived drugs are of considerable interest for the understanding of not only the drug-receptor interaction, but also of their transport and absorption properties and the prediction of their partitioning behavior at the cell membrane receptors.

In this chapter, the estimation and parametrization of the hydrophobicity of amino acids and peptides will be briefly overviewed. The effects of conformation on the hydrophobicity of a peptide as well as the effects of three-dimensionally organized hydrophobicity on its biological activity will be examined. A special note on hydrophobic collapse [2] will be made as a clear example of the interplay of hydrophobic and conformational effects in water and organic phases. Finally, it will be shown that optimization of the partition coefficient and monitoring of peptide absorption are useful tools in the design of peptide and peptidomimetic drugs.

20.2 Lipophilicity of Amino Acids and Parametrization of Side Chain Hydrophobicity

Experimental methods for the measurement of partition coefficients (P) at a given pH (D) of amino acids have been reviewed [3]. Due to the high hydrophilicity of free amino acids, direct determination of P, e.g., in the octanol/water system is difficult [4, 5], even when using countercurrent partition chromatography [6]. Retention parameters in thin-layer chromatography [7] and in HPLC [8], although not easily obtained with free amino acids, can be converted, after calibration, to log P values in octanol/water. Several theoretical scales have also been derived from accessible molecular surface areas and the corresponding hydration potentials [9] and from the amino acid side chain exposure in water-soluble globular proteins [10].

The large number of hydrophobic scales proposed for amino acids led Eisenberg and McLachlan [11] to estimate a consensus scale already in 1986. Abraham and Leo [12] reviewed later the state of the art (1987), when extending their general fragmental method to include amino acids. Finally, van de Waterbeemd et al. (1994) [3] comprehensively compared the available lipophilicity scales, including a few unusual amino acids.

The hydrophobicity of free amino acids is of interest in a number of biological processes such as, e.g., amino acid transport and uptake in neurotransmission or amino acid incorporation in plant physiology. In addition, amino acid *side chain* parameters have to be considered in any QSAR study or any calculation of peptide hydrophobicity on the basis of additivity models or fragment constants. A general problem in this case is the presence of charges in the backbone. In order to get free of the interaction of these charges, a number of substituted derivatives such as *N*-acetyl-*C*-amides [13], *N*-acetyl-*C*-methyl amides [14] have been used as model compounds for the assessment of side chain hydrophobicity, either for direct partitioning in octanol/water or for determination of HPLC retention times. The resulting data produced the first "reliable" scales and helped to set a standard to this field. This approach, first used for amino acids, was extended later to N- and C-protected peptides such as Ac-Ala-Xaa-Ala-NHtBu [15], and Ac-Gly-Xaa-Xaa-(Leu)$_3$-(Lys)$_2$-NH$_2$ [16] with the aim of estimating side chain hydrophobicity in increasingly larger model peptides after elimination of any possible interaction with N- and C-terminal end groups. The general observation was made that the hydrophilicity of polar amino acid side chains was markedly reduced by flanking peptide bonds [17]. In addition, the local environment of a side chain cer-

tainly affects hydrophobicity, a fact which points to the questionable value of hydrophobicity profiles [18] which only consider the adjacent residues in the sequence but not the realthree-dimensional vicinity of the given residue in the folded protein.

There are still considerable differences between the available lipophilicity scales for amino acids, even between those resulting only from experimentation [3]. However, a good correlation is observed between the calculated values of Abraham and Leo [12] and the experimental values obtained from direct partitioning of *N*-acetyl-*C*-amides in 1-octanol/water at pH 7.1 [13]. The residual discrepancy observed for Asp and Glu (ionized side chains) has been eliminated by the data obtained by Kim and Szoka [15] with protected tripeptides: new values for Asp $\pi = -2.57$ and Glu $\pi = -2.29$ (instead of -0.77 and -0.64, respectively (Table 1)). The side chain parameter for Pro cannot be defined in the usual way, due to its cyclic structure [13]. An updated list of π values for noncoded amino acids is presented in Chapter 21 of this book. The polarity scale (Table 1) established by Vallat et al. [19] (see Section 20.3.2) usefully complements the available hydrophobicity scales.

Using the fragmental method of Rekker [20], hydrophobicity parameters for 22 posttranslationally modified amino acids have been calculated [21], which can be used

Table 1. Lipophilicity π (pH 7.1) and polarity Λ side chain parameters of the 20 coded amino acids

Side chain of	π exp.[a]	π calc.[c]	Λ[d]
Ala	0.31	0.32	0.2
Arg	−1.01	−1.78	4.0
Asn	−0.60	−0.97	2.0
Asp	−2.57[b]	−3.18	2.0
Cys	1.54	0.43	0.6
Gln	−0.22	−0.97	2.2
Glu	−2.29[b]	−3.84	2.3
Gly	0	0	0
His	0.13	0.01	2.1
Ile	1.80	1.81	0.3
Leu	1.70	1.81	0.4
Lys	−0.99	−1.80	3.6
Met	−1.23	0.81	1.0
Phe	1.79	1.87	1.1
Pro	(0.72)	0.95	0.7
Ser	−0.04	−0.62	0.8
Thr	0.26	−0.30	1.0
Trp	2.25	1.88	1.5
Tyr	0.96	1.20	2.1
Val	1.22	1.27	0.4

[a] $\pi = \log D$ (AcNHCH(R)CONH$_2$) − log D (Ac-Gly-NH$_2$); Fauchère and Pliška [13].
[b] Corrected according to Kim and Szoka [15].
[c] CLOG P, Abraham and Leo [12].
[d] Λ, side chain polarity, Vallat et al. [19].

as a first estimate of the contribution to log D of these frequently appearing residues in physiological peptides. However, *experimental* values for these residues are lacking as are also amino acid side chain parameters for important posttranslationally modified peptides, such as tyrosine-O-sulfate in gastrin, cholecystokinin, and hirudin.

20.3 Lipophilicity of Peptides, Pseudopeptides, and Mimetics

20.3.1 Experimental P Values for Peptides

These values are still scarce and often not available even for well-known biologically active peptides. A few values are assembled in Table 2 for a number of established neuropeptides, peptide hormones and peptide drugs. Some of these literature data have been obtained in 1-octanol/aqueous buffer mixtures at a defined pH (around 7) by the direct shake-flask method [39] and detection of the peptide in one of the two phases by UV (without or after HPLC) or by radioactive counting. The values in this solvent system are often used as reference data. However, other systems have been considered such as several alkane/water(buffer) mixtures [40] which may be useful in membrane permeability studies. Chromatographic retention parameters obtained by reverse phase HPLC often replace the 1-octanol/water system [41–44]. In these cases, calibration with a few directly measured values of P in 1-octanol/water allows one to estimate log Ps (1-octanol/water) for comparative purposes within the calibration range [27]. More recently, centrifugal counter-current chromatography has been successfully used for the determination of log P of amino acids and peptides [19]. This liquid/liquid partition chromatography, which eliminates the need for a solid support, has been shown to produce highly correlated values with the shake-flask technique also for zwitterionic amino acids and peptides.

Due to the different pK values of the amino terminal and carboxy terminal functions, and also of other ionizable side chains, several species are present during partitioning and the obtained parameter is better described as a distribution coefficient (log D instead of log P). At pH 7.0 $\pm$ 0.2, the carboxylic groups are negatively charged and the amine and guanidine groups positively charged to a very large extent. The side chain of histidine is a particular case since the pK_a of the imidazole function (pK_a 6.1) indicates that both the protonated and free ionic forms will be significantly represented at pH 7, while the pK_a of the phenolate group of tyrosine, being more than two log units above the pH of measurement, is mostly uncharged. It is therefore not surprising that a very different chromatographic behavior was observed at pH 2 and pH 7 for 20 model octapeptides [45], which obviously reflects hydrophobicity (and possibly conformational) changes.

It can be observed (Table 2) that linear peptides are hydrophilic compounds with negative values of log P at a physiologically relevant pH. Cyclic peptides are more lipophilic, by an increment of about 2.5 log units (1-octanol/water, [19]), due to the lack of the terminal ionic carboxylic and ammonium functions. Monoiodinated peptides such as a number of compounds in Table 2, display higher log P values than the origi-

Table 2. Partition coefficients in octanol/water of some biologically active peptides

Peptide	n^a	Analog	$\log D_{o/w}{}^b$	Reference
Angiotensin	8	^{125}I[Sar^1, D-Phe^8]angiotensin II	$-1.02^{c)}$	[22]
		[Leu^8]angiotensin II	$0.18^{d)}$	[23]
Bradykinin	10	DArg-Arg-Pro-Hyp-Gly-Thi-Ser-DTic-Oic-Arg^e	-0.65	f
	9	Gba-Arg-Pro-Hyp-Gly-Thi-Ser-DTic-Oic-Arg^g	-1.25	f
Cyclosporin	11	c[MeBmt-Abu-Sar-MeLeu-Val-MeLeu-Ala-DAla-MeLeu-MeLeu-MeVal]	2.92 / 2.99	[24] / [25]
DSIP^h	9	Trp-Ala-Gly-Gly-Asp-Ala Ser-Gly-Glu	-2.03	[26]
	8	desTrp-DSIP	-2.20	[26]
Endothelin	5	c[DVal-Leu-DTrp-DGlu-Pro] (BQ123)	1.55	[27]
	3	Azc-Leu-DTrp-DPya (FR139317)^i	1.75	[27]
	6	Ac-DDip-Leu-Asp-Ile-Ile-Trp (PD142 893)^j	0.36	[28]
		Ac-DBhg-Leu-Asp-Ile-Ile-Trp (PD145 065)^j	0.35	[28]
Enkephalin	5	^{125}I-[Leu^5]enkephalin	0.05	[29]
		^{125}I-[Met^5]enkephalin	-1.52	[29]
		Tyr-DAla-Gly-Phe-DLeu-NHEt	1.41	[30]
Gastrin releasing	27	^{125}I-APVSVGGGTVLAKM-YPRGNHWAVGHLM-NH$_2{}^k$	-1.65	[29]
Glucagon	29	^{125}I-HSQGTFTSDYSKYLDS-RRAQDFVQWLMNT^{k,l}	-1.20	[29]
Gramicidin	10	c[Pro-Val-Orn-Leu-DPhe-Pro-Val-Orn-Leu-DPhe]	2.51	[31]
Luliberin	10	^{125}I-Pyr-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH$_2$	-1.35	[29]
α-Melanotropin	13	^{125}I-Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val-NH$_2$	-1.30	[29]
Perindopril^m	3	N_α-[ethylpentanoate-2-yl]-Ala-Oic	-1.02	f
Secretin	27	HSDGTFTSELSRLRDSA-RLQRLLQGLV^k	-0.72	[32]
Somatostatin	14	^{125}I-[Tyr^1]somatostatin	-1.09	[29]
	8	DPhe-c[Cys-Phe-DTrp-Lys-Thr-Cys]-Thr(ol) (octreotide)	-2.27	[25]
	6	c[MeAla-Tyr-DTrp-Lys-Val-Phe] (minisomatostatin)	0.90	[33]
Substance P	11	^{125}I-Arg-Pro-Lys(Hpp)-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met-NH$_2{}^n$	-0.56	[29]
Thyroliberin	3	^{125}I-Pyr-His-Pro-NH$_2$ (TRH)	-1.42	[29]
		Pyr-His-Pro-NH$_2$	-2.46	[34]
Vasopressin	9	c[deaminoCys-Tyr-Phe-Gln-Asn-Cys]-Pro-DArg-Gly-NH$_2$ (dDAVP)	-3.50	[35]
		^{125}I-c[Cys-Tyr-Phe-Gln-Asn-Cys]-Pro-Arg-Gly-NH$_2$ (AVP)	-2.15	[29]

^{a} Number of residues. ^{b} log P at pH 7.2 $\pm$ 0.2 unless otherwise stated. ^{c} Unknown pH. ^{d} log P i nBuOH/AcOH/H$_2$O (countercurrent distribution). ^{e} Icatibant, HOE140 [36]. ^{f} J. Y. Ginot and J. L. Fauchère, unpublished. ^{g} S16118 [37]. ^{h} Delta sleep-inducing peptide. ^{i} Acz = hexahydroazepyl-1-ylcarbonyl; DPya = 2-pyridyl-(R)-alanine. ^{j} Dip = β,β-diphenyl-alanine, Bhg = α-(10,11-dihydro-5H-dibenzo[a,d]cycloheptenyl-glycine. ^{k} One-letter symbols. ^{l} Monoiodinated. ^{m} Angiotensin I-converting enzyme inhibitor, Oic = octahydroindole-2-yl-carbonyl [38]. ^{n} Hpp = iodinated 3-(4-hydroxyphenyl)-proprionyl.

nal peptide, by approximately one log unit, as can be seen, e.g., for unlabeled and io-dinated TRH. Introduction of nonnatural amino acids is another way to modulate the hydrophobicity of peptides. The range of partition coefficients between both cyclic and extensively N-methylated cyclosporin (log P = 2.9) and extremely hydrophilic vaso-pressin analog dDAVP (log P = $-$ 3.5) extends over more than six log units!

These experimental values are of considerable interest to obtain insight into the ab-sorption, transport, distribution and crossing of the physiological barriers of these bio-logically active compounds.

20.3.2 Calculated Values of log P (log D) for Peptides

A calculated prediction of the log D value from the primary sequence of a peptide would be a useful information for design purposes. At first sight, the presence of adja-cent amino acid building blocks should be well adapted to the generally observed addi-tivity of the fragmental hydrophobicity contributions. However, experience has shown that additivity is often violated and that many correcting factors have to be introduced to fit the measured values [12]. The reasons for this anomalous behavior are side chain-side chain interactions (due to possible hydrophobic collapse or charge-charge ef-fects), end group interactions (charged amino and carboxyl functions), bound water molecules, hydrogen bonds and other long-range interactions due to conformational flexibility. Hence, the substituent constants obtained for the amino acid side chains (Fauchère and Pliška, [13]) differ significantly from the uncorrected fragmental con-stants (Leo- or Rekker-type) for the same substituents. On the same line, Akamatsu and Fujita [46] proposed an *effective* hydrophobicity for amino acid side chains and Val-lat et al. [19] suggest that a methylene group cannot fully express its hydrophobicity in a side chain shorter than that of norleucine (see [17]). Since the presence of charges se-verely increases complexity, efforts towards the prediction of log D have concentrated first on nonionizable *model peptides*.

Several empirical equations have been proposed for the prediction of the hydro-phobicity (log D) of di- to pentapeptides with unionizable side chains based on their primary structure: 59 di- and tripeptides [47], 53 N-acetyl-di- and tripeptide amides [48], and 124 di- to pentapeptides [49]. These complex equations contain a large num-ber of indicator variables and lose their validity when the training set is extended. They are restricted to nonionizable side chains, as is also the equation of Gao et al. [50]. However, the work of Fujita's groups has provided experimental log Ds for large sets of model compounds from which many useful quantitative features can be extracted.

Using simply the substituent constants determined for the side chains of N-acetyl-amino acid amides [13] and an empirically determined fragment constant for the pep-tide backbone (f_{bb}), log D values for the peptide series described above can easily be calculated, with standard deviations comparable with the experimental errors (Tab-le 3), without any other correcting factor. The best fitting to the experimental values is obtained for N-acetyl amides of both dipeptides and tripeptides and also for a set of cyclodipeptides described by Vallat et al. [19]. The deviations are larger for free pep-tides and increase with chain length from di- to pentapeptides. The (negative) incre-ment Δf_{bb} of the backbone fragment constant is found to decrease with increasing

Table 3. Calculated distribution coefficients at pH 7 ± 0.2 (log D) of seven di- to pentapeptide series

Peptide series	log D	backbone (bb)	n^a	f_{bb}^b	$\Delta \log D^c$	sd^d
1. Cyclodipeptides (diketopiperazines)	$f_{bb} + \pi_{Xa1} + \pi_{Xa2}$	cyclo[-NH-CH-CO-]$_2$	11	-2.01	$-0.35 < \Delta \log D < 0.40$	0.22
2. Dipeptides Ac-Xa1-Xa2-NH$_2$	$f_{bb} + \pi_{Xa1} + \pi_{Xa2}$	Ac-[NH-CH-CO]$_2$-NH$_2$	31	-2.60	$-0.32 < \Delta \log D < 0.38$	0.18
3. Tripeptides Ac-Xa1-Xa2-Xa3-NH$_2$	$f_{bb} + \pi_{Xa1} + \pi_{Xa2} + \pi_{Xa3}$	Ac-(NH-CH-CO)$_3$-NH$_2$	22	-3.25	$-0.50 < \Delta \log D < 0.45$	0.25
4. Free dipeptides	$f_{bb} + \pi_{Xa1} + \pi_{Xa2}$	$^+$H-(NH-CH-CO)$_2$-O$^-$	46 / 45^e	-4.51	$-0.91 < \Delta \log D < 1.59$ / $-0.91 < \Delta \log D < 0.72$	0.52 / 0.47
5. Free tripeptides	$f_{bb} + \pi_{Xa1} + \pi_{Xa2} + \pi_{Xa3}$	$^+$H-(NH-CH-CO)$_3$-O$^-$	48	-5.50	$-0.91 < \Delta \log D < 0.98$	0.45
6. Free tetrapeptides	$f_{bb} + \pi_{Xa1} + \pi_{Xa2} + \pi_{Xa3} + \pi_{Xa4}$	$^+$H-(NH-CH-CO)$_4$-O$^-$	33	-6.27	$-1.18 < \Delta \log D < 1.14$	0.68
7. Free pentapeptides	$f_{bb} + \pi_{Xa1} + \pi_{Xa2} + \pi_{Xa3} + \pi_{Xa4} + \pi_{Xa5}$	$^+$H-(NH-CH-CO)$_5$-O$^-$	32^e	-6.39	$-1.78 < \Delta \log D < 1.14$	0.79

[a] No. of substances used for analysis
[b] fragment constant for the fragment bb
[c] maximum deviation between calculated and observed log D
[d] standard deviation of $\Delta \log D$ within individual series
[e] Outliers omitted

number of residues. While deviations can be large for individual values, the standard deviations are not exceedingly higher than the experimental errors for any of the investigated series. However, compound sampling in the available series is far from random and mainly restricted to nonionizable side chain residues, as is therefore also the validity of the simple additivity model.

The next step would be to introduce refined correcting factors such as described by Leo [51] using a computerized fragmental method for CLOGP of peptides. The method works reasonably well up to pentapeptides, although it does not yet handle peptides with formally charged residues.

Another simple way to calculate log D in the same series has been based on the separation of volume-dependent from polar contributions to the partition coefficient, according to the equation log $D = aV + b - \Lambda$ (Λ = polarity of the peptide; a, b = constants obtained from alkane behavior [19]). When calculating the molecular volume V from the atomic radii of Gavezotti [52] and the polarity factor according to the additivity rule $\Lambda = \Sigma \Lambda_{SC}$, Λ_{SC} = amino acid side chain polarity, the authors obtained estimated log D with deviations from measured values comparable with experimental error. The same restrictions apply as above as to a general validity of the method for longer peptides with ionizable side chains.

On the whole, it can be stated that we are still far from being able to calculate the log D values of an arbitrary decapeptide including polar side chains and secondary structure-inducing residues. The additive models so far neither consider sequence dependence, nor take into account the chirality of the α-carbon, although theoretical studies on diastereoisomers of *N*-acetyl-dipeptide-methylamides [53] predict significant differences in log P. First reverse-phase HPLC experiments on diastereoisomeric and sequence-inversed dipeptides also confirm this fact [54, 55].

20.3.3 Pseudopeptides

Pseudopeptides are chemically modified peptides which contain at least one surrogate of the peptide bond indicated by the symbol ψ[surrogate] (for a review, see [56]). Among the predictable local effects of the introduction of a ψ-bond in the peptide, a change in hydrophobicity can be expected and approximated by calculation. While no experimental data are available, increments in log P have been calculated for ψ-substitution in the model compounds *N*-methylacetamide [57] and alanylalanine [55] showing that the values are strongly dependent on the environment of the surrogate bond (Table 4). A reliable prediction of the overall change in log D of the pseudopeptide due to the ψ-substitution is impossible yet and there is a need for reference experimental data. However, even if other marked structural changes are associated with ψ-substitutions, such as loss of hydrogen bond(s), of planar geometry and *trans* substitution, the local and overall changes in lipophilicity of pseudopeptides should also be taken into account in design.

Oligocarbamates H-[HN-CHR-CH$_2$OCO]$_n$-OH with the side chains R of natural amino acids are also pseudopeptides with the substitution ψ[CH$_2$OCONH]. Comparison of the log D in octanol/water (pH 7.4) of the two dipeptides Ac-YKFLG-OH and Ac-YIFLG-OH ($- 1.95$ and $- 1.0$, respectively) with the corresponding two carbamates (log D 0.30 and 0.40, respectively) reveal the latter to be significantly more hydrophobic [59].

Table 4. Changes of hydrophobicity due to peptide bond surrogates[a]

Surrogate	ψ	$\Delta \log P^b$	$\Delta \log P^c$
Amide	-CONH-	0	0
Retroamide	-NHCO-	0	0
Methylenamino	-CH$_2$NH-	0.45	1.09
Carbamate	-NHCOO-	0.24	0.92
Thiocarbamate	-NHCOS-	0.26	1.01
Urea	-NHCONH-	0.14	0.53
Ester	-COO-	0.32	1.22
N-Methylamide	-CONCH$_3$-	0.63	0.28
Amidomethylene	-CONHCH$_2$-	− 0.08	0.53
Methylenurea	-CH$_2$NHCONH-	0.44	1.06
Ketomethylene	-COCH$_2$-	0.26	1.34
Thioamide	-CSNH-	0.19	0.71
Ethene	-CH=CH-	0.88	3.34
Ethylene	-CH$_2$CH$_2$-	0.92	3.89
Thiolester	-COS-	0.42	1.61
Sulfonamide	-SO$_2$NH-	0.06	0.21
Hydroxyethylene	-CH(OH)CH$_2$-	− 0.46	1.71
Cyanomethylenamino	-CH(CN)NH-	− 0.40	0.35
Methylenesulfoxide	-SOCH$_2$-	− 0.21	0.23
Methylenoxy	-CH$_2$-O-	0.56	1.42
Methylenethio	-CH$_2$-S-	0.89	2.45
Carbamate	-CH$_2$OCONH	0.60	0.71

[a] Calculated with CLOGP [58].
[b] $\log P$ (Ala ψ Ala) − $\log P$ (Ala-Ala).
[c] $\log P$ (CH$_3$ ψ CH$_3$) − $\log P$ (CH$_3$CONHCH$_3$).

20.3.4 Peptidomimetics

Peptidomimetics are nonpeptide structures, which retain the biological activity of the original peptide [58]. The same experimental (partitioning or chromatography) as well as theoretical (Rekker's or Leo's fragmental calculation) methods apply for the determination of P of peptidomimetics as for other organic compounds.

Typical peptide mimetics obtained by chemical modification of peptide leads, are the ACE inhibitor cilazapril [60] and the tuftsin antagonist [4-aminopropyl, 2-aminobutyl]-indolizidinone [61] (both tripeptide mimics) or the nonprodrug mimic of RGDX (X = variable amino acid), SB208651 [62]. A number of nonpeptide ligands of peptide receptors discovered by blind screening are also known (for a review, see [63]; see also Table 1 in [2]). Optimized lipophilicity balanced with hydrogen bonding capacity are considered to be the factors which improve bioavailability of these compounds, compared with the natural peptide.

Mimics of architectural elements such as turns [64] or α-helix segments [58], which are means to introduce conformational constraints, also modulate lipophilicity. Among the simplest of these structural elements, the γ-lactam-constrained dipeptides (formed over two additional methylene groups) Gly[ANC-2]-Leu [65] (CLOGP −1.57) and Pro[spiro-γ-lactam]Leu [66] (−0.75), display higher log P values than the original dipeptide Gly-Leu and Pro-Leu (CLOGP −1.96 and −1.64, respectively). Linked pyrrolinones used to mimic β-strands in peptides [67] are also characterized by a building unit cyclo[-CHR-CO-CH-CH$_2$-NH-] with increased hydrophobicity (calculated Δ log P = 0.85) compared with the amino acid residue with the same side chain R. Similarly, the calculated log P values for the γ-turn CH$_3$-CH$_2$-CH(CH$_3$)-CH(NH$_2$)-cyclo[CH-CH$_2$-CH$_2$-N-CO-CH(CH$_3$)-NH]-CH$_2$-COOH [68] (CLOGP −1.07) mimic of Ile-Ala-Gly is higher by 1.2 log units than for the tripeptide. The phenothiazine amino acid scaffold, -NH-CH$_2$-C$_6$H$_4$-[-NCH$_3$-,-S-]-C$_6$H$_4$-CO-, introduced to replace two α-helix turns in vasointestinal peptide [13–19] is a strongly hydrophobic unit (CLOGP 3.58), which probably mimics only the hydrophobic side of the amphipathic helix [69]. These examples show that the rational design of peptide mimetics should not only base on structural complementarity with the target site, but should also consider the lipophilicity changes induced by chemical modification on the way to mimetics.

20.4 Lipophilicity and Peptide Conformation

20.4.1 Log *P* and Conformation

Effortsto relate partition coefficients to chemical structure are illustrated by the determination of fragmental and substituent constants and even more clearly be the identification of intramolecular proximity effects [21, 58]. Since peptides are highly flexible polymers and their conformation of major importance for receptor recognition, the dependence of lipophilicity on 3D-structure should be investigated (for a review, see [70]).

A first clue about effects of 3D-elements on the measured values of partition coefficients was the introduction of a correction term derived from the β-turn-inducing potential of amino acids in the empirical equation for the calculation of log D [46–48, 50]. The theoretical [53] and experimental [54, 55] differences in log D for diastereoisomeric peptides are another indication of these effects. However, we are still far from a quantitative prediction of the conformation-dependent effects on the value of log D for a medium-size arbitrary peptide.

A theoretical study of eight acetyl-amino acid methylamides, taking into account the redistribution of the conformer populations during partitioning into water and octanol, respectively, suggests this to be a reason for the many correcting factors used in CLOGP for computing log P [71]. Other studies demonstrate that peptide hydrophobicity is a property not only of sequence, but also of secondary structure. The model peptide Lys-(Ala)$_{18}$-Lys, predicted to be α-helical, goes through a minimum retention time when a proline residue replaces one alanine along the whole sequence, indicating that disruption of the α-helix lowers the apparent peptide hydrophobicity in the RP-HPLC system [72].

20.4.2 Amphipathic Secondary Structures

Medium size linear peptides (> 10 amino acid residues) tend to fold into well-defined secondary structure motifs, such as the α-helix, the β-pleated sheet or the 3_{10}-helix (a helix with 3 residues per turn and 10 atoms in the ring joined by the intramolecular hydrogen bond, in contrast to the ideal 3.6_{13} helix), depending on the sequence. In naturally occurring peptides, hydrophobicity is unequally distributed on the outer surface of these structures. Hence, the commonly encountered amphipathic α-helix displays two strikingly different sides, the hydrophilic one providing water solubility and the hydrophobic one allowing it to bind to complementary structures. Both graphical methods [73] (now replaced by computer graphics) and computational techniques [74] based on amino acid hydrophobicity scales have been devised to represent and predict these amphipathic helices. Optimization of these structures has been used in peptide drug design. Hence, Kaiser [75] synthesized chimeric analogs of calcitonin, glucagon, and β-endorphin, in which amphiphilicity of the naturally helical segment was increased by amino acid substitutions. Despite minimal sequence homology with the parent peptide (in the modified segment), the resulting analogs displayed high potencies, thus confirming the validity of the concept. It is believed that the amphiphilic helix (address) aids the positioning of the active site of the hormone (message) for productive interaction with the receptor, or helps the peptide to find its receptor by better absorption and diffusion at the cell surface.

Four interconnected amphipathic α-helical structures were also designed to mimic the structure of chymotrypsin, leading to chymohelizyme, a 73-residue peptide with high catalytic activity [76]. As studied by DeGrado et al. [77], hydrophobic interactions play a dominant role in secondary structure formation and may be used as a central driving force in the *de novo* design of synthetic proteins.

Finally, molecular models of the 3D-structure of the G-protein-coupled receptors, on which most of the polypeptide hormones act and which could not be crystallized so far, are also obtained on the basis of hydrophobicity profiles [18] and of sequence alignment in order to identify the seven hydrophobic transmembrane fragments [78]. Despite some controversy about the validity of the presently available models, these molecular graphics are unique tools for the design of antagonists of peptide hormones and neuropeptides [79].

20.4.3 Hydrophobic Collapse

Hydrophobic collapse is the clustering of hydrophobic aromatic or lipophilic groups in flexible structures, typically peptide-like compounds, which dramatically distorts the aqueous conformation compared with that observed in an organic solvent or lipophilic environment [2]. This process has been observed for cyclosporin A, which displays a distinct conformation in the free or receptor-bound state [80]. Major implications of this hydrophobic effect, which should be either exploited or prevented, are expected for the rational design of peptidomimetics. It is postulated that many bioactive ligands described in recent years [81], display an overall restriction of the conformational freedom compared with the original peptide, thus preventing hydrophobic collapse and maintaining individual side chain interactions with the binding site. Another conse-

quence of hydrophobic collapse is that different conformational populations are likely to exist in water and octanol (or another apolar phase) during partitioning [71].

20.4.4 Molecular Lipophilicity Potential

The solvent-exposed groups of a molecule create not only an electrostatic but also a lipophilic potential at the surface of and around any molecule. This potential, which varies with conformation, is increasingly "felt" by another approaching species and certainly participates to the ordering of the neighbor molecules in solution. The lipophilic potential originating from a given molecule can be estimated at each point in space as the sum of the hydrophobic contributions of each individual atom or group of atoms in the molecule (see Chapters 12 and 13).

Molecular lipophilicity potentials for enalapril (a modified tripeptide) [82] and for cyclosporin (a cyclic undecapeptide) [83] have been presented, which help to visualize possible hydrophobic interactions and to distinguish them from electrostatic effects. In another application, the log *P* of 13 cyclodipeptides has been estimated from the sum of the hydrophobic (positive) and hydrophilic (negative) parts of the molecular lipophilicity potential derived from a set of conformers of each molecule [83]. General applicability of the latter approach has still to be assessed.

20.5 Lipophilicity and Peptide Transport

20.5.1 Pharmacokinetic Properties

These properties are required for any drug candidate and are related with absorption, distribution, metabolism, and elimination. Among other factors, molecular weight, hydrophobicity, and hydrogen bonding are crucial for the crossing of cellular and physiological barriers. For peptide drugs, a competition between proteolysis and absorption is of special relevance. Favorable hydrophobicity ranges have been delineated, for the crossing of a given barrier. Hence, absorption from the gastrointestinal tract is optimal for log *P* values between 0.5 and 2, while maximal CNS penetration is obtained for log *P* around 2 (for a review, see [84]). It is therefore of importance to avoid large hydrophobicity increase on the basis of *in vitro* tests only, which will shift the values of log *P* to ranges no longer compatible with *in vivo* absorption. Early monitoring during the design of analogs, of enteral absorption or CNS penetration has also proven useful, e.g., for the discovery of potent renin inhibitors [85]. However, lipophilicity is often overestimated as the major controlling factor of oral activity or transport properties, which can be shown to be strongly modulated by the hydrogen bonding capacity of the active compound.

20.5.2 Hydrogen Bonding and Hydrophobicity

The amide bonds in polypeptides, beside linking adjacent amino acids, also frequently form intramolecular hydrogen bonds, e.g., in cyclic structures and for the induction of turns and of other well-defined secondary structures such as the β-pleated sheet and the

α-helix. Intermolecular hydrogen bonding certainly occurs also with the solvent (e.g., with water and octanol during partitioning) as well as with other solutes.

The relative effect of hydrophobicity and hydrogen bonding has been investigated in several series of homologous model peptides for the transport across the intestinal mucosa and the blood-brain barrier (see Chapter 14). The end-group-protected peptides (N_α-acetyl-amides) consisting either of free or N-methylated D-phenylalanine residues, were: Ac-DPhe-NH$_2$, Ac-(DPhe)$_2$-NH$_2$, Ac-(DPhe)$_3$-NH$_2$, Ac-(DPhe)$_2$-N(Me)DPhe-NH$_2$, Ac-DPhe-(N(Me)DPhe)$_2$-NH$_2$, Ac-(N(Me)DPhe)$_3$-NH$_2$, Ac-(N(Me)DPhe)$_3$-NHMe. It was found first that the permeability of Caco-2 cell monolayers (as a model of the intestinal mucosa) was substantially increased by N-methylation, while this chemical modification had only minor effects on the partition coefficient in octanol/water [86]. This observation was also made for bovine brain microvessel endothelial cell monolayers (as a model of the blood-brain barrier) and in rat brain perfusion experiments *in situ*. The permeabilities of these peptides did not correlate with the octanol/aqueous buffer partition coefficients, while they were correlated with the number of potential hydrogen bonds of the peptides with water. Consistent with this, correlations of the permeability were found with the partition coefficient in heptane/ethylene glycol or with the differences $\Delta\log P$ of the partition coefficients in octanol/water and isooctane/water [40, 87].

These results suggest that: 1) in contrast to many other compound classes, the octanol/water partition coefficient may not be a good indicator of membrane permeability and barrier crossing for peptides, due to possible hydrogen bonds of the solutes with octanol; 2) that it should be replaced either by heptane/ethylene glycol [88] or by the difference $\Delta\log P = \log P$ (octanol/water) $- \log P$(alkane/water), in order to better estimate hydrogen bonding [89, 90]; and 3) N-methylation may be a convenient way to provide a peptide analog, with increased membrane permeability.

Table 5. Blood levels (μg ml) of pseudopeptide endothelin and neurokinin receptor antagonists after oral administration (25 mg kg [27])

Compound		Rat portal vein			Rat jugular vein			$\log P^b$	K_I^c (nmol l^{-1})
		30 min	60 min	90 min	30 min	60 min	90 min		
BQ 123	1	1.50	1.14	0.47	0.78	0.66	0.32	1.55	–
FR 139317	2	0.55	0.26	0.19	0.46	0.20	0.14	1.75	–
S16764	3	0.83	0.71	0.47	0.49	0.40	0.27	1.86	–
± CP96345	4^d	1.79	0.77	0.50	1.24	0.55	0.36	1.44	0.25
S 18523	5	0.24	0.14	0.10	0.19	0.10	0.08	2.61	1.5
S 16375	6	0.25	0.15	0.10	0.14	0.11	0.08	2.48	11
S 16474	7	0.19	0.20	0.20	0.07	0.09	0.06	3.09	81
S 15890	8	0.18	0.26	0.11	0.11	0.15	0.04	3.17	1400
S 16265	9	0.09	0.12	0.16	0.06	0.09	0.10	5.66	1400

[a] For structure, see [27].

[b] log P in octanol/water, pH 7.1.

[c] Binding affinity on human lymphoblast cell line IM9 using the radioligand[^{3}H]Sar^9, Met(O$_2$) substance P.

[d] Nonpeptide reference neurokinin antagonist [92].

As a matter of fact, the exceptional bioavailability of cyclosporin A [91], is related mainly to the extensive natural N-methylation and to the cyclic structure, in addition to the conformation-dependent orientation of the hydrogen bonds in the free and bound state [80]. However, the latter statement is limited by the known drastic structural changes conferred on peptides by N-methylation [56], such as tendency to *cis* amide bond configuration and reduced conformational freedom (partly compensated by better resistance to proteolysis).

The blood levels obtained after oral administration to rats of three endothelin receptor antagonists and of several neurokinin antagonists, including a non peptide reference were measured as a first estimate of their bioavailability [27]. An *inverse* correlation between the blood concentrations and the log P of the peptide analogs was observed (Table 5), not only within each series, but also over the two series. This confirmed the high lipophilicity of the pseudopeptide analogs and suggested an optimized log P value for proper absorption and barrier crossing lower than 1.5, in line with observations made for nonpeptide series [68]. Another reason for this correlation may be that intermolecular hydrogen bonds were not dominant in these strongly modified (N-methylated, cyclized) peptides, leaving hydrophobicity as the crucial factor for absorption. As can be expected, the *in vitro* binding affinity in the neurokinin series does not follow the absorption levels. While the nonpeptide antagonist CP 96345 [93] is both better absorbed and more potent *in vitro*, the affinity of the peptide analogs, as measured by the dissociation constant in the competitive binding model, cannot be rationalized on the basis of the reached blood levels. However, selection of S18523 as the best peptide analog is justified, since increase of the *in vitro* affinity was not achieved at the cost of reduced bioavailability. The potency order of the oral antinociceptive activity of the neurokinin antagonist paralleled the plasma levels at the jugular vein [27]. No oral bioactivity data are available so far for the other endothelin and neurokinin antagonists. On the whole, these results ([27] and Table 5) stress again the importance of monitoring oral absorption early in the process of peptide drug design and confirm the importance of hydrophobicity and hydrogen bonding for peptide drug action.

20.5.3 Prodrugs

Following the classical prodrug approach, bioreversible derivatization of peptide-derived drugs has been successfully undertaken in several cases. Hence, pseudotripeptide ACE inhibitors such as enalapril or perindopril are administered as hydrophobic ethyl ester derivatives (log $P = -1.0$; Table 2), which are converted into the free acid active form by plasma esterase-catalyzed hydrolysis in the circulation. The pivalate ester (on tyrosine-2, log $P = -2.2$) of dDAVP (log $P = -3.5$) showed a marked higher flux across monolayers of Caco cells than the parent drug [35]. The hydrophobicity of TRH was also modulated by N-acylation of the histidine imidazole, leading not only to better CNS penetration but also to better protection against proteolysis [34]. With the same purpose of transmembrane delivery of peptides, N_α-lauroyl-TRH was prepared and shown to increase the log P from -1.17 to 0.28 (in octanol/water, pH 6.5) and to only weakly reduce the biological activity [94]. The covalent conjugate of octreotide

with D(+)-maltose may use a specific sugar transport system, which could explain the increased oral somatostatin activity [95]. In contrast, the doubly substituted [Leu^5]enkephalin apparently benefits from alternating hydrophobic and hydrophilic properties imparted to the peptide by conjugation with cholesterin and nicotinic acid residues: passive diffusion of the nonionic conjugate into the CNS would be followed by confinement of the membrane-impermeable oxidized cationic species [96] within the CNS. The number of examples of peptide-derived prodrugs with optimized hydrophobicity and improved transport properties will certainly increase in future.

20.6 Conclusion and Outlook

This short overview leads to the following main statements:

1. The lipophilicity of amino acids, as expressed by their log P (or log D), or their substituent constants π, is well defined for neutral side chain residues, increasingly better established for those with ionizable side chains and still controversial for Cys and Pro. This fact fits well to the decomposition of lipophilicity in a volume-dependent hydrophobicity and an electrostatic polarity contributions [19].

2. The log P values for peptides are both difficult to estimate experimentally or to calculate by semiempirical methods, the resulting data depend upon the prevailing conformation. No satisfactory algorithm is today available for the reliable calculation, starting from the structure, of the partition coefficients of an arbitrary decapeptide in water/octanol. The hope that clues on the peptide secondary structure could be gained from the measurement of partition coefficients is even more elusive.

3. The log P values of peptides can be strongly modulated by chemical modification of the natural sequence, such as the introduction of peptide bond surrogates or of non-coded amino acid residues. However, the correct prediction of the resulting value is still a major challenge, mainly due to the intramolecular interactions.

4. The secondary structure of medium-size peptides can be predicted and hydrophobic domains optimized. Hydrophobicity profiles [18], although very often used, are questionable descriptions since only adjacent residues in the primary sequence are considered.

5. More than 80 % of the QSAR studies of peptide-derived drugs let hydrophobicity appear as a major variable. Principal component analyses ([96] and review [3]) for amino acids also results in hydrophobicity being one of the three main components. *In vivo* studies show even more marked effects of hydrophobicity on the absorption and distribution of peptide-derived drugs.

6. Optimization of the lipophilicity of a peptide or peptide-derived drug candidate with respect to absorption (e.g., after peroral administration) is a reasonable goal which has been achieved in a few cases, either by chemical modification, vectorization or synthesis of prodrugs. The hydrophobicity range adjusted in *in vitro* tests may not be compatible with absorption requirements.

These statements demonstrate the progress achieved in the estimation and characterization of the lipophilicity of amino and peptides. The complexity of the methods needed is a strong indication that lipophilicity is encoding a lot of information, reflecting both the repetitive and flexible structure of peptides. Molecular recognition processes as well as transport properties are likely to be better elucidated in future, when taking into account a differential estimation of the lipophilicity of peptide drug candidates.

References

[1] Tanford, C., *The Hydrophobic Effect*. Wiley: London, 1973
[2] Rich, D., Effect of hydrophobic collapse on enzyme-inhibitor interactions: Implications for the design of peptidomimetics. In: *Perspectives in Medicinal Chemistry*. Testa, B., Kyburz, E., Fuhrer, W., and Giger, R. (Eds.). Helvetica Chimica Acta: Basel 15–25 (1993)
[3] Van de Waterbeemd, H., Karajiannis, H., and El Tayar, N., *Amino Acids* **7**, 129–145 (1994)
[4] Fauchère, J. L., Do, K. Q., Jow, P. Y., and Hansch, C., *Experientia* **36**, 1203–1204 (1980)
[5] Yunger, L. M., and Cramer, R. D., *Mol. Pharmacol.* **20**, 602–608 (1981)
[6] Tsai, R. S., Testa, B., El Tayar, N., and Carrupt, P. A., *J. Chem. Soc. Perkin Trans.* **2**, 1797–1802 (1991)
[7] Pliška, V., Schmidt, M., and Fauchère, J. L., *J. Chromatogr.* **216**, 79–92 (1981)
[8] Molnar, I., and Horvath, C., *J. Chromatogr.* **142**, 623–640 (1977)
[9] Janin, J., *Nature*, **277**, 491–492 (1979)
[10] Ooi, T., Oobatake, M., Nemethy, G., and Scheraga, H. A., *Proc. Natl. Acad. Sci. USA* **84**, 3086–3090 (1987)
[11] Eisenberg, D., and McLachlan, A. D., *Nature* **319**, 199–203 (1986)
[12] Abraham, D. J., and Leo, A. J., *Proteins, Struct. Funct. Gen.* **2**, 130–152 (1987)
[13] Fauchère, J. L., and Pliška, V., *Eur. J. Med. Chem.* **18**, 369–375 (1983)
[14] Wynne, H. J., Van Buuren, K. H., and Wakelkamp, W., *Experientia* **38**, 655–656 (1982)
[15] Kim, A., and Szoka, F. C., *Pharmaceut. Res.* **9**, 504–514 (1992)
[16] Parker, J. M., Guo, D., and Hodges, R. S., *Biochemistry* **25**, 5425–5432 (1986)
[17] Roseman, M. A., *J. Mol. Biol.* **200**, 513–522 (1988)
[18] Kyte, J., and Doolittle, R. F., *J. Mol. Biol.* **157**, 105–132 (1982)
[19] Vallat, P., Gaillard, P., Carrupt, P. A., Tsai, R. S., and Testa, B., *Helv. Chim. Acta* **78**, 471–485 (1995)
[20] Purcell, W., Bass, G., and Clayton, J., *Strategy of Drug Design, Appendix I*, Wiley, New York, 1973
[21] Rekker, R. F., and Mannhold, R., *Calculation of Drug Lipophilicity*, VCH: Weinheim, 1992
[22] Bossé, R., Gerold, M., Fischli, W., Holck, M., and Escher, E., *J. Cardiovasc. Pharmacol.* **16** (Suppl. 4), S50–S55 (1990)
[23] Paiva, A. C., Nouailhetes, V. L., and Paiva, T. B., *J. Med. Chem.* **20**, 898–901 (1977)
[24] El Tayar, N., Mark, A. E., Vallat, P., Brunne, R. M., Testa, B., and Van Gunsteren, W. F., *J. Med. Chem.* **36**, 3757–3764 (1993)
[25] Begley, D. J., *Prog. Brain Res.* **91**, 163–169 (1992)
[26] Banks, W. A., Kastin, A. J., Coy, D. H., and Angulo, E., *Brain Res. Bull.* **17**, 155–158 (1986)
[27] Paladino, J., Kucharczyk, N., Morris, A. D., Thibault, M., Mahieu, J. P., Serkiz, B., Volland, J. P., Autissier, C., and Fauchère, J. L., *Drug Design Discov.* **12**, 121–128 (1994)
[28] Clarke, F. H., and Cahoon, N. M., *J. Pharm. Sci.* **84**, 53–54 (1995)
[29] Banks, W. A., and Kastin, A. J., *Brain Res. Bull.* **15**, 287–292 (1985)

[30] Dodson, B. A., and Miller, K. W., *Anesthesiology* **62**, 615–620 (1985)

[31] Katayama, T., Nakao, K., Akamatsu, M., Ueno, T., and Fujita, T., *J. Pharm. Sci.* **83**, 1357–1362 (1994)

[32] Ohwaki, T., Ishii, M., Aoki, S., Tatsuishi, K., and Kayano, M., *Chem. Pharm. Bull.* **37**, 3359–3362 (1989)

[33] Veber, D. F., Saperstein, R., Nutt, R. F., Freidinger, R. M., Brady, S. F., Curley, P., Perlow, D. S., Paleveda, W. J., Colton, C. D., Zacchei, A. G., Tocco, D. J., Hoff, D. R., Gerich, G. E., and Hirschmann, R., *Life Sci.* **34**, 1371–1378 (1984)

[34] Bundgaard, H., and Moos, J., *Pharmaceut. Res.* **7**, 885–892 (1990)

[35] Kahns, A. H., Buur, A., and Bundgaard, H., *Pharmaceut. Res.* **10**, 68–74 (1993)

[36] Hock, F. J., Wirth, K., Albus, U., Linz, W., Gerhards, H. J., Wiemer, G., Henke, S., Breipohl, G., König, W., Knolle, J., and Schölkens, B. A., *Br. J. Pharmacol.* **102**, 769–773 (1991)

[37] Thurieau, C., Félétou, M., Canet, E., and Fauchère, J. L., *Bioorg. Med. Chem. Lett.* **4**, 781–784 (1994)

[38] Scalbert, E., Narcisse, G., Abdon, D., Fauchère, J. L., and Santoni, J. P., *J. Cardiovasc. Pharmacol.* **15**, 676–683 (1990)

[39] Black, S. D., and Mould, D. R., *Anal. Biochem.* **193**, 72–82 (1991)

[40] Chikhale, E. G., Ka-Yun Ng, Burton, P. S., and Borchardt, R. T., *Pharmaceut. Res.* **11**, 412–419 (1994)

[41] Meek, J. L., and Rossetti, Z. L., *J. Chromatogr.* **211**, 15–28 (1981)

[42] Wilson, K. J., Honegger, A., Stoetzel, R. P., and Hughes, G. J., *Biochem. J.* **199**, 31–41 (1981)

[43] Mant, C. T., Burke, T. W., Black, J. A., and Hodges, R. S., *J. Chromatogr.* **458**, 193–205 (1988)

[44] Wilce, M. C., Aguilar, M. I., and Hearn, M. T., *J. Chromatogr.* **536**, 165–183 (1991)

[45] Guo, D., Mant, C. T., Taneja, A. K., Parker, J. M., and Hodges, R. S., *J. Chromatogr.* **359**, 499–517 (1986)

[46] Akamatsu, M., and Fujita, T., A new hydrophobicity index for amino acid side chains and its applications. In: *Peptide Chemistry 1992*, Yanainara, N. ed., Escom, Leiden (1993) p. 446–448

[47] Akamatsu, M., Yoshida, Y., Nakamura, H., Asao, M., Iwamura, H. and Fujita, T., *Quant. Struct.-Act. Relat.* **8**, 195–203 (1989)

[48] Akamatsu, M., Okutani, S., Nakao, K., Hong, N. J., and Fujita, T., *Quant. Struct.-Act. Relat.* **9**, 189–194 (1990)

[49] Akamatsu, M., and Fujita, T., *J. Pharm. Sci.* **81**, 164–174 (1992)

[50] Gao, H., Lien, E. J., and Wang, F., *J. Drug Target.* **1**, 59–66 (1993)

[51] Leo, A. J., Computer calculation of peptide hydrophobicity. In: *QSAR: Rational Approaches to the Design of Bioactive Compounds,* Silipo, C., and Vittoria, A. eds., Elsevier, Amsterdam (1991) p. 349–352

[52] Gavezzotti, A., *J. Am. Chem. Soc.* **105**, 5220–5225 (1983)

[53] Bodor, N., and Huang, M. J., *J. Comput. Chem.* **12**, 1182–1186 (1991)

[54] Funasaki, N., Hada, S., and Neya, S., *Anal. Chem.* **65**, 1861–1867 (1993)

[55] Van de Waterbeemd, H., Karajiannis, H., Kansy, M., Obrecht, D., Müller, K., and Lehmann, C., Conformation-lipophilicity relationships of peptides and peptide mimetics. In: *QSAR and Computer Modeling in Drug Design.* Sanz, F. (Ed.). Prous: Barcelona: 1995

[56] Spatola, A. F., Peptide backbone modifications: A structure-activity analysis of peptides containing amide bond surrogates. In: *Chemistry and Biochemistry of Amino Acids, Peptides and Proteins,* Vol. **7**, Weinstein, B. ed., Dekker, New York (1983) p. 267–357

[57] Fincham, I. F., Higginbottom, M., Hill, D. R., Horwell, D. C., O'Toole, J. C., Ratcliffe, G. S., Rees, D. C., and Roberts, E., *J. Med. Chem.* **35**, 1472–1484 (1992)

[58] Olson, G. L., Bolin, D. R., Bonner, M. P., Bös, M., Cook, C. M., Fry, D. C., Graves, B. J., Hatada, M., Hill, D. E., Kahn, M., Madison, V. S., Rusiecki, V. K., Sarabu, R., Sepinwall, J., Vincent, J. P., and Voss, M. E., *J. Med. Chem.* **36**, 3039–3049 (1993)

[59] Cho, C. Y., Moran, E. J., Cherry, S. R., Stephans, J. C., Fodor, S. P., Adams, C. L., Sundaram, A., Jacobs, J. W., and Schultz, P. G., *Science* **261**, 1303–1305 (1993)

[60] Attwood, M. R., Francis, R. J., Hassall, C. H., Krohn, A., Lawton, G., Natoff, I. L., Nixon, J., Redshaw, S., and Thomas, W. A., *FEBS Lett.* **165**, 201–205 (1984)

[61] Kahn, M., and Devens, B., *Tetrahedron Lett.* **27**, 4841–4844 (1986)

[62] Bondinell, W. E., Keenan, R. M., Miller, W. H., Ali, F. E., Allen, A. C., De Brosse, C. W., Eggleston, D. S., Erhard, K. F., Haltiwanger, R. C., and Yuan, C. K., *Bioorg. Med. Chem.* **2**, 897–908 (1994)

[63] Gante, J., *Angew. Chem. Int. Ed. Engl.* **33**, 1699–1720 (1994)

[64] Rose, G. D., Gierasch, L. M., and Smith, J. A., *Adv. Protein Chem.* **37**, 1–109 (1985)

[65] Freidinger, R. M., Schwenk, D., and Veber, D. F., *J. Org. Chem.* **47**, 104–109 (1982)

[66] Ward, P., Ewan, G. B., Jordan, C. C., Ireland, S. J., Hagan, R. M., and Brown, J. R., *J. Med. Chem.* **33**, 1848–1851 (1990)

[67] Smith, A. B., Hirschmann, R., Pasternak, A., Akaishi, R., Guzman, M. C., Jones, D. R., Keenan, T. P., and Sprengeler, P. A., *J. Med. Chem.* **37**, 215–218 (1994)

[68] Ma, S., Passaro, G., Rolland-Fulcrand, V., Fauchère, J. L., Lazaro, R., Roumestant, M. L., and Viallefont, P., Synthesis of γ-turn mimetics, In: *Peptides 94*, Proc. 23rd Eur. Peptide Symposium, Braga, 1995, Maia, H. L. (Ed.). ESCOM: Leiden 1996

[69] Leo, A. J., *Chem. Rev.* **93**, 1281–1306 (1993)

[70] El Tayar, N., Karajiannis, H., and van de Waterbeemd, H., *Amino Acids* **8**, 125–139 (1995)

[71] Richards, N. G., Williams, P. B., and Tute, M. S., *Int. J. Quantum Chem.* **44**, 219–233 (1992)

[72] Büttner, K., Ostresh, J. M., and Houghten, R. A., Induced confromational effects during HPLC. In: *Peptides, Chemistry, Structure and Biology.* Rivier, J. E., and Marshall, G. (Eds.). Escom: Leiden; 423–425 (1990)

[73] Schiffer, M., and Edmundson, A. B., *Biophys. J.* **7**, 121–135 (1967)

[74] Cornette, J. L., Cease, K. B., Margalit, H., Spouge, J. L., Berzofsky, J. A., and DeLisi, C., *J. Mol. Biol.* **195**, 659–685 (1987)

[75] Kaiser, E. T., *Trends Biochem. Sci.* **12**, 305–309 (1987)

[76] Stewart, J. M., Cann, J. R., Corey, M. J., Hahn, K. W., and Klis, W. A., Solid-phase synthesis of a peptide designed to have enzymatic activity. In: *Solid Phase Synthesis 1992.* Epton, R. (Ed.). Intercept: Andover; 23–27 (1992)

[77] DeGrado, W. F., Wasserman, Z. R., and Lear, J. D., *Science* **243**, 622–628 (1989)

[78] Hibert, M., Trumpp-Kallmeyer, S., Bruinvels, A., and Hoflack, J., *Mol. Pharmacol.* **40**, 8–15 (1991)

[79] Trumpp-Kallmeyer, S., Hoflack, J., and Hibert, M., Modeling of G-protein-coupled receptors: Application to the NK1 receptor: In: *The Tachykinin Receptors.* Buck, S. H. (Ed.). Humana Press: Totowa, NJ; 237–255 (1994)

[80] Fesik, S. W., Gampe, R. T., Holzman, T. F., Egan, D. A., Edalji, R., Lulu, J. R., Simmer, R., Helfrich, R., Kishore, V., and Rich, D. H., *Science* **250**, 1406–1409 (1990)

[81] Wiley, R. A., and Rich, D. H., *Med. Res. Revs.* **13**, 327–384 (1993)

[82] Fauchère, J. L., Quarendon, P., and Kaetterer, L., *J. Mol. Graphics* **6**, 202–206 (1988)

[83] Gaillard, P., Carrupt, P. A., Testa, B., and Boudon, A., *J. Comput.-Aid. Mol. Des.* **8**, 83–96 (1994)

[84] Topliss, J. G., *Perspect. Drug. Discov. Des.* **1**, 253–268 (1993)

[85] Rosenberg, S. H., Spina, K. P., Woods, K. W., Polakowski, J., Martin, D. L., Yao, Z., Stein, H. H., Cohen, J., Barlow, J. R., Egan, D. A., Tricarico, K. A., Baker, W. R., and Kleinert, H. D., *J. Med. Chem.* **36**, 449–459 (1993)

[86] Conradi, R. A., Hilgers, A. R., Ho, N. F., and Burton, P. S., *Pharmaceut. Res.* **9,** 435–439 (1992)

[87] Burton, P. S., Conradi, R. A., and Hilgers, A. R., *Adv. Drug Deliv. Rev.* **7,** 365–386 (1991)

[88] Paterson, D. A., Conradi, R. A., Hilgers, A. R., Vidmar, T. J., and Burton, P. S., *Quant. Struct.-Act. Relat.* **13,** 4–10 (1994)

[89] 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)

[90] El Tayar, N., Testa, B., and Carrupt, P. A., *J. Phys. Chem.* **96,** 1455–1459 (1992)

[91] Grevel, J., *Transplant. Proc. (Suppl.)* **18,** 9–15 (1986)

[92] Snider, R. M., Constantine, J. W., Lowe, J. A., Longo, K. P., Lebel, W. S., Woody, H. A., Drozda, S. D., Desai, M. C., Vinick, F. J., Spencer, R. W., and Hess, H. J., *Science* **251,** 435–437 (1991)

[93] Fauchère, J. L., Charton, M., Kier, L. B., Verloop, A., and Pliška, V., *Int. J. Peptide Protein Res.* **32,** 269–278 (1988)

[94] Muranishi, S., Sakai, A., Yamada, K., Murakami, M., Takada, K., and Kiso, Y., *Pharmaceut. Res.* **8,** 649–652 (1991)

[95] Albert, R., Marbach, P., Bauer, W., Briner, U., Fricker, G., Bruns, C. and Pless, J., *Life Sci.* **53,** 517–525 (1993)

[96] Bodor, N., Prokai, L., Wu, W. M., Farag, H., Jonalagadda, S., Kawamura, M., and Simkins, J., *Science* **257,** 1698–1700 (1992)

[97] Nomenclature and symbolism for amino acids and peptides, *Eur. J. Biochem.* **138,** 1–37 (1984)

21 Side Chain Lipophilicity of Noncoded α-Amino Acids: π-Values

Vladimir Pliška and Emanuel Escher

Abbreviations

DOPA   3,4-Dihydroxyphenylalanine
GABA   γ-Aminobutyric acid
HPLC   High-pressure liquid chromatography
IUPAC   International Union of Pure and Applied Chemistry
IUB   International Union of Biochemistry
s.e.   Standard error (standard deviation of mean)
TL/TLC   Thin-layer/thin-layer chromatography

Symbols for amino acids, if not suggested by the IUPAC-IUB nomenclature rules (cf. Section 21.3), are taken from [1–7].

Symbols

P Partition coefficient n-octanol/water
R Side chain
R_f Relative mobility on a TLC chromatographic plate
π Hansch hydrophobic constant for an amino acid side chain (R)

21.1 Introduction

Synthetic peptides used as drugs frequently contain amino acid side chains not coded in the human DNA. Only very few of them occur in living organisms (GABA, DOPA, kynurenine, etc.), and even less, as products of postribosomal processes, in native proteins (pyroglutamic acid, γ-carboxyglutamic acid, tyrosinesulfate, etc.). Their properties can be very different from those of coded amino acids: some of the physicochemical descriptors reach extreme values and may therefore considerably change interactions of their bearers with enzymes, receptors, and other biologically important macromulecules. Obviously, this is the reason why they are considered essential for rational drug design. In addition, they can serve efficiently as enzyme or receptor "probes" in peptide pharmacology [8–11], since they enable to investigate effects of structural and physico-chemical variations of the interacting substance on a particular biological response. Due to highly developed tools in peptide synthesis, variations within a broad range of properties can be achieved and can reach extreme levels [1].

Physico-chemical descriptors used in the quantitative structure-activity analysis of biologically active peptides (correlation methods) were several times listed in the past.

Revised values of lipophilicity descriptors for coded amino acids are presented in Chapter 20 of this book. Tables 1–3 represent a collection of the presently available hydrophobic constants π for amino acid side chains which may serve to the purposes mentioned in the foregoing paragraph.

21.2 Lipophilicity Descriptor π

Hansch hydrophobic constants π for amino acid side chains are defined as

$$\pi_R = \log P_R - \log P_H, \tag{1}$$

where P_R, P_H are partition coefficients in the system octanol/water for a side chain R,

$$-NH-CH-CO-, \tag{2}$$
$$\vert$$
$$R$$

and for a standard bearing hydrogen as the side chain (R = H, glycine), respectively. Estimation of Ps may be encumbered by complex dissociation/association equilibria in any amino acid molecule which bring about uncertainties when π values are derived from partition of amino acids or peptides not substituted at both α-amino and α-carboxy groups. Methods described in Chapter 7 may yield more accurate values in the future. Alternatively, π-values were derived from partition data obtained in nondissociable backbone basis (see Chapter 20). These restrictions are valid for any of the experimental path toward obtaining partition coefficients P, independent of the type of compound. As for amino acids, the experimental tools employed include shake flask method, high-pressure liquid chromatography (HPLC, see Chapter 5), and thin-layer chromatography (TLC, see Chapter 8).

21.3 Description of Tables

Table 1 summarizes data for side chains derived from coded non-aromatic amino acids. Tables 2 and 3 list derivatives of phenylalanine (mostly aromatic substitutions) and tryptophan. Nomenclature recommended by the IUPAC-IUB nomenclature rules (Nomenclature and symbolism for amino acids and peptides: *Eur. J. Biochem.* **138**, 1–37 (1984)) is used as a standard whenever applicable. Derivatives are usually assigned by their commonly used names (not always identical with the current nomenclature of organic chemistry). Positions of functional groups, however, are almost systematically assigned by position numbers; Greek letters were kept only exceptionally in a few very common names (β-naphtylalanine, N^εlysine derivatives, etc.), prefixes *o-*, *m-*, *p-* only in cases when numbers might be confusing or cumbersome. Abbreviations (symbols) follow, first, IUPAC-IUB recommendations (see above), and second, abbreviations already used in the literature. (*Comment:* some symbols, particularly for polysubstituted aromatic derivatives, are still very complex and ought to be substituted by more suitable ones as soon as the side chain in question is more frequently used in synthetic peptides). Braces [ ] mark alternative, and in most of the cases no more rec-

ommended nomenclature (this concerns particularly the use of the prefix *nor:* Names like norvaline and norleucine should *not* be utilized any more).

Formulae of side chains are presented in the simplest form, basic structures are indicated in the headings. Formulae of ionized forms (as alternatives) are presented only in cases in which these forms can clearly be defined. Otherwise, conditions ("*acidic*", "*neutral*", etc.) related to the water phase of the partitioning system are given at corresponding π-values (next column). Some less common side chains are shown in Fig. 1.

Reported π-values are arithmetic means of several measurements by the same author or authors' group. When taken from the literature (Ref. column), standard errors (s.e.) are not indicated, even if sometimes available in the original publication(s). Values communicated for the first time in this chapter ("*new*", see below), or those obtained by conversion of TLC data in [5] using a modified Collander equation described earlier (see section 21.4), are presented by arithmetic mean and standard error. Still uncertain parameters π, particularly if the experimental values are strongly deviated from a preliminary prediction based on fractional contributions for amino acids [2], a question mark (?) is attached. In some instances (for potentially important, but not yet measured, side chains), estimates computed as suggested above [2] are listed ("*calc.*").

Reference numbers refer to the reference list. For "*calc.*" and "*new*" see above.

21.4 Newly Reported *π*-values

Recently, we have measured π-values of a number of side chains of unusual amino acids (Fig. 1) by means of TLC method (cellulose plates), as described earlier. Following solvent systems were used:

n-butanol/acetic acid/water	5:2:3
iso-propanol/acetic acid/water	4:1:1
iso-propanol/25 % ammonia soln.	2:1
n-butanol/pyridine/acetic acid/water	5:4:1:2 (pH ca. 5.2)
n-butanol/0.1M acetate buffer, pH 7.2	2:1 (organic phase).

Individual amino acids were synthesized in our laboratory (E.E.) or purchased from various companies. Position of the migrated substances on TLC plates was detected by ninhydrin. R_f-values were converted to log P values using the modified Collander equation [2], and subsequently to π. A set of amino acids with independently estimated P values listed in our earlier communication [2] was used for conversion. Only amino acid derivatives with R_f values within this standard set are listed in Tables 1–3; extrapolated values were excluded.

Each substance was measured in at least 3 solvent systems, 3 to 12 times. Unless otherwise indicated, π-values in Tables 1–3 are arithmetic means of compatible values from all solvent systems and their respective standard errors.

In addition, we have converted TCL data by Wold et al. [5] obtained in 7 solvent systems into π-values. Conversion of R_f into π was performed by the method mentioned above [2]. Arithmetic means and standard errors were computed from 5 to 7 values obtained in individual solvent systems. They are referred to as "[5] *new*" in Tables 1–3.

o-carboranylalanine   adamantylalanine

ferrocenylalanine   cymantrenylalanine   cyclobutadienyl-iron-
                                                   tricarbonylalanine
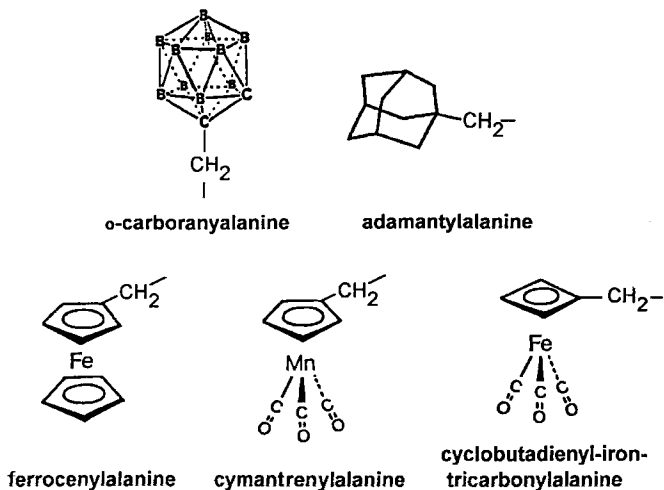
Figure 1. Structures of side chains of some less conventional synthetic amino acids (π-values; cf. Table 1).

21.5 Tables

Table 1. Glycine and alanine derivatives: $-NH-CH-CO-$
$$\overset{|}{R}$$

	Amino acid side chain	Symbol	Side chain structure R	$\pi \pm$ s.e.	Ref.
	2-Aminobutyric acid	Abu	$-CH_2-CH_3$	0.64 ± 0.03	[5] *new*
				0.83	[6]
	2-Aminovaleric acid {norvaline}	Avl {Nva}	$-(CH_2)_2-CH_3$	1.73	[2]
				1.14 ± 0.05	[5] *new*
				1.18 ± 0.03	*new*
	2-Aminohexanoic acid {norleucine}	Ahx {Nle}	$-(CH_2)_3-CH_3$	1.70	[2]
				1.54 ± 0.05	[5] *new*
	tert-Butylglycine	Bug	$CH_3-\overset{\displaystyle CH_3}{\underset{\displaystyle CH_3}{\overset{|}{\underset{|}{C}}}}-CH_3$	1.51	[6]
				1.44	[1]
	Neopentylglycine	Neo	$-CH_2$ $\overset{|}{CH_3-\overset{|}{\underset{|}{C}}-CH_3}$ $\overset{|}{CH_3}$	1.89	[6]
	Allylglycine		$-CH_2-CH=CH_2$	0.89 ± 0.01	*new*
	Propargylglycine	Prg	$-CH_2-C\equiv CH$	0.50	[6]
				0.63 ± 0.03	*new*

(Aliphatic side chains)

Table 1 (continued)

	Amino acid side chain	Symbol	Side chain structure R	$\pi \pm$ s.e.	Ref.
Cyclic side chains	Cyclopentylglycine	Cpg		1.99	[6]
	Cyclohexylglycine	Chg		2.40	[6]
	Cyclohexylalanine	Cha	$-CH_2$	2.72	[6]
	2,5-Dihydrophenylalanine		$-CH_2$	2.72 ± 0.12	*new*
	Adamantylalanine	Ada	$-CH_2-C_{10}H_{15}$ (cf. Fig. 1)	3.63 / 3.24 / 3.73	[1] / [2] / [4]
Aromatic side chains	Phenylglycine	Phg		1.60	[6]
	Homophenylalanine	Hph	$-CH_2-CH_2$	2.10	[6]
	Kynurenine		$-CH_2-CO$ (NH$_2$)	1.59 ± 0.15	*new*
	β-Naphthylalanine	Bnp	$-CH_2-$	3.15	[6]
	1-Pyrenylalanine	Pyr	$-CH_2-$	3.54	*new*
	N^γ-Phenylasparagine	Asn(Φ)	$-CH_2-CO-NH-$	1.40	[6]
Heterocyclic side chains	2-Thienylglycine	Thg		1.25	[6]
	3-Pyridinylalanine		$-CH_2$	1.10	[6]
	2-Pyrazinylalanine	Paa	$-CH_2$	0.42	[4]
	Benzofurylalanine		$-CH_2$	2.65 ± 0.11	*new*
	Benzothienylalanine		$-CH_2$	3.20 ± 0.16	*new*
	o-Carboranylalanine	Car	$-CH_2-C_2B_{10}H_{11}$ (cf. Fig. 1)	4.14 / 4.30	[2] / [4]
Metal-containing side chains	Ferrocenylalanine		(cf. Fig. 1)	2.65 ± 0.15	*new*
	Cymanthrenylalanine		(cf. Fig. 1)	2.61 ± 0.15	*new*
	Cyclobutadienyl-iron-tricarbonyl-alanine		(cf. Fig. 1)	2.93 ± 0.05	*new*

Table 1 (continued)

	Amino acid side chain	Symbol	Side chain structure R	$\pi \pm$ s.e.	Ref.
Side chains containing basic or acidic groups	2,5-Diaminobutyric acid	Dab	$-(CH_2)_2-NH_2$	-1.34	[6]
	Ornithine	Orn	$-(CH_2)_3-NH_2$	-1.11	[2]
				-1.14 ± 0.09	*new*
				-1.17	[6]
	N^ε-Methyllysine	Lys (Me)	$-(CH_2)_4-NH-CH_3$	0.41	[2]
			$-(CH_2)_4-N^+H_2-CH_3$	-0.78	[2]
	N^ε-Formyllysine	Lys (HCO)	$-(CH_2)_4-NH-CHO$	0.41	[2]
	Homoarginine	Har	$-(CH_2)_4-NHC = N^+H_2$ $\mid$ NH_2	-0.44	[2]
	Citrulline	Cit	$-(CH_2)_3-NHCO-NH_2$	-0.02	[2]
				0.02 ± 0.05	[5] *new*
	2-Aminoadipic acid	Aap	$-(CH_2)_3-COOH$	0.77	*calc.*
			$-(CH_2)_3-COO^-$	-0.42	*calc.*
Side chains containing OH and derivatives	N^γ-phenylasparagine	Asn(Φ)	$-CH_2-CO-NH$ ⬡	1.40	[6]
	Homoserine	Hse	$-CH_2-CH_2-OH$	0.36	*calc.*
	O-Benzylserine	Ser(Bzl)	$-CH_2-O-CH_2-$ ⬡	2.34	[6]
	O-(3-Pyridinylmethyl)-serine	Ser(Pym)	$-CH_2-O-CH_2$ ⬡N	0.86	[6]
	O-Methylthreonine	Omt	$-CH-CH_3$ $\mid$ OCH_3	0.83 ± 0.08	[5] *new*
Sulfur-containing side chains	Homocysteine	Hcy	$-(CH_2)_2-SH$	1.23	*calc.*
			$-(CH_2)_2-S^-$	0.02	*calc.*
	Penicillamine	Pen	$CH_3-\overset{\mid}{\underset{\mid}{C}}-SH$ CH_3	1.52	[2]
			$CH_3-\overset{\mid}{\underset{\mid}{C}}-S^-$ CH_3	0.82	[2]
	Methioninesulfone	Met(O$_2$)	$-CH_2-SO_2-CH_3$	0.2	[6]

Table 2. Phenylalanine and tyrosine derivatives: -NH·CH·CO-

-CH₂-⟨benzene ring: positions 2,3,4,5,6⟩

Amino acid side chain	Symbol	Side chain position 2	3	4	5	6	$\pi \pm$ s.e.	Ref.
Phenylalanine	Phe						1.69	[1]
							1.63	[2]
							1.79	[3]
4-Methylphenylalanine	Phe(4-Me)			$-CH_3$			2.04	[4]
							1.64	calc.
4-Ethylphenylalanine	Phe(4-Et)			$-CH_2CH_3$			2.46	calc.
2-Hydroxyphenyl-alanine [o-Tyrosine]	Phe(2-OH)	$-OH$					1.91 ± 0.03	new
3-Hydroxyphenylalanine [m-Tyrosine]	Phe(3-OH)		$-OH$				1.74 ± 0.07	new
Tyrosine	Tyr			$-OH$			0.88	[2]
							0.96	[3]
3-Hydroxymethylphenyl-alanine	Phe(3-CH₂OH)		$-CH_2OH$				1.19 ± 0.05	new
O-Methyltyrosine	Tyr(Me)			$-OCH_3$			1.61	[2]
							1.77	[4]
O-Ethyltyrosine	Tyr(Et)			$-OCH_2CH_3$			2.02	calc.
O-n-Butyltyrosine	Tyr(nBu)			$-O(CH_2)_3CH_3$			2.84	calc.
O-Benzyltyrosine	Tyr(Bzl)			$-OCH_2C_6H_5$			2.72	[6]
2-Fluorophenylalanine	Phe(2-F)	$-F$					1.66 ± 0.09	new
3-Fluorophenylalanine	Phe(3-F)		$-F$				1.89 ± 0.12	new
4-Fluorophenylalanine	Phe(4-F)			$-F$			1.32	[4]
							1.99 ± 0.09	new
4-Chlorophenylalanine	Phe(4-Cl)			$-Cl$			2.73	[4]
							2.10 ± 0.05	new

Monosubstituted

Table 2. (continued)

Monosubstituted

Amino acid side chain	Symbol	Side chain position					$\pi \pm$ s.e.	Ref.
		2	3	4	5	6		
4-Bromophenylalanine	Phe(Br)			$-Br$			2.35	[4]
							2.27 ± 0.11	new
4-Iodophenylalanine	Phe(4-I)			$-I$			2.28 ± 0.06	new
3-Aminophenylalanine	Phe(3-NH$_2$)		$-NH_2$				0.51 ± 0.06	new
			$-N^+H_3$				0.05 ± 0.03	new
4-Aminophenylalanine	Phe(4-NH$_2$)			$-NH_2$			0.99	[4]
							0.96 ± 0.20	new
				$-N^+H_3$			0.03 ± 0.02	new
4-Nitrophenylalanine	Nip			$-NO_2$			1.96	[4]
	Phe(4-NO$_2$)						1.85 ± 0.13	new
3-Azidophenylalanine	Phe(3-N$_3$)		$-N{=}N{\equiv}N$				1.97 ± 0.11	new
4-Azidophenylalanine	Phe(4-N$_3$)			$-N{=}N{\equiv}N$			2.24 ± 0.13	new
4-Diazoniumphenylalanine	Phe(4-N$_2$)			$-N{\equiv}N^+$			acidic: -1.35 ± 0.23	new
				$-N{=}N{-}OH$			neutral: -0.14 ± 0.14	new
3-Carboxyphenylalanine	Phe(3-CO$_2$H)	$-COOH$					acidic: 1.40	new
Phenylalanine-4-sulfonic acid	Phe(4-SO$_3$H)			$-SO_3H$			acidic and neutral: 0.08 ± 0.03	new
3-Cyanophenylalanine	Phe(3-CN)	$-CN$					1.44 ± 0.02	new
4-Dihydroxyboryl-phenylalanine				$HO{-}B{-}OH$			acidic: 1.67 ± 0.01	new
							basic: -0.51	new

Table 2. (continued)

Amino acid side chain	Symbol	Side chain position					$\pi \pm$ s.e.	Ref.
		2	3	4	5	6		
2- and 3-Tyrosine derivatives								
2-Methyltyrosine	Tyr(2-Me)	$-CH_3$			$-OH$		1.82 ± 0.09	*new*
3-Hydroxytyrosine [DOPA]	DOPA		$-OH$	$-OH$			0.73	*new*
3-Methoxytyrosine [3-Methyl-DOPA]	Tyr(3-OMe)		$-OCH_3$	$-OH$			0.83	[2]
3-Chlorotyrosine	Tyr(3-Cl)		$-Cl$	$-OH$			1.32 ± 0.06	*new*
3-Iodotyrosine	Tyr(3-I)		$-I$	$-OH$			1.91 ± 0.78 (?)	*new*
3-Aminotyrosine	Tyr(3-NH$_2$)		$-NH_2$	$-OH$			2.0	*calc.*
			$-N^+H_3$	$-OH$			0.64	*new*
3-Nitrotyrosine	Tyr(3-NO$_2$)		$-NO_2$	$-OH$			-0.03 ± 0.11	*new*
3-Carboxytyrosine	Tyr(3-CO$_2$H)		$-COOH$	$-OH$			1.78 ± 0.10	*new*
							acidic: 0.64 ± 0.05	*new*
Disubstituted								
3,5-Dihydroxyphenylalanine	Dph		$-OH$		$-OH$		0.56	[6]
							acidic: 0.79 ± 0.07	*new*
							neutral: -0.09 ± 0.07	*new*
3,5-Dichlorophenylalanine	Phe(3,5-Cl$_2$)		$-Cl$		$-Cl$		2.69 ± 1.19(?)	*new*
3-Nitro-4-chlorophenylalanine	Phe(3-NO$_2$,4-Cl)		$-NO_2$	$-Cl$			2.25 ± 0.14	*new*
3,5-Diaminophenylalanine	Phe(3,5-(NH$_2$)$_2$)		$-NH_2$		$-NH_2$		*neutral :* 0.21 ± 0.13	*new*
							acidic: -0.33 ± 0.09	*new*
2-Nitro-4-diamino-phenylalanine	Phe(3-NO$_2$,4-NH$_2$)	$-NO_2$		$-NH_2$			*acidic and neutral:* 0.86 ± 0.06	*new*
2,4-Dinitrophenylalanine	Dip	$-NO_2$		$-NO_2$			2.04	[4]

Table 2. (continued)

Amino acid side chain	Symbol	Side chain position 2	3	4	5	6	$\pi \pm$ s.e.	Ref.
Disubstituted								
2-Nitro-4-azidophenylalanine	Phe(2-NO$_2$, 4-N$_3$)	$-$NO$_2$		$-$N=N≡N			1.94 ± 0.08	new
3-Carboxy-4-nitrophenylalanine	(Phe(3-CO$_2$H, 4-NO$_2$))		$-$COOH	$-$NO$_2$			acidic: 0.05 ± 0.03	new
2,5-Dihydroxytyrosine [2-hydroxy-DOPA]	Tyr(2,5-(OH)$_2$)	$-$OH		$-$OH	$-$OH		acidic: 0.79 ± 0.00 neutral: -0.09 ± 0.07	new
Trisubstituted								
3,5-Dibromotyrosine	Tyr(3,5-Br$_2$)		$-$Br	$-$OH	$-$Br		2.57 ± 0.07	new
3,5-Diiodotyrosine	Tyr(3,5-I$_2$)		$-$I	$-$OH	$-$I		2.77 ± 0.07	new
3,5-Dibromo-4-chlorophenylalanine	Phe(3,5-Br$_2$, 4-Cl)		$-$Br	$-$Cl	$-$Br		2.96 ± 0.15	new
3,4,5-Tribromophenylalanine	Phe(3,4,5-Br$_3$)		$-$Br	$-$Br	$-$Br		3.45 ± 0.42	new
3,5-Dibromo-4-aminophenylalanine	Phe(3,5-Br$_2$, 4-NH$_2$)		$-$Br	$-$NH$_2$	$-$Br		acidic and neutral: 2.56 ± 0.10	new
3,5-Diiodo-4-aminophenylalanine	Phe(3,5-I$_2$, 4-NH$_2$)		$-$I	$-$NH$_2$	$-$I		acidic and neutral: 2.70 ± 0.11	new
3,5-Dinitrotyrosine	Tyr(3,5-(NO$_2$)$_2$)		$-$NO$_2$	$-$OH	$-$NO$_2$		1.15 ± 0.08	new
Pentasubstituted								
2,3,4,5,6-Pentamethylphenylalanine	Phe(2,3,4,5,6-Me$_5$)	$-$CH$_3$	$-$CH$_3$	$-$CH$_3$	$-$CH$_3$	$-$CH$_3$	3.14 ± 0.25	new
2,3,4,5,6-Pentaflourophenylalanine	Phe(2,3,4,5,6-F$_5$)	$-$F	$-$F	$-$F	$-$F	$-$F	0.53	[4]
2,3,4,5,6-Pentachlorophenylalanine	Phe(2,3,4,5,6-Cl$_5$)	$-$Cl	$-$Cl	$-$Cl	$-$Cl	$-$Cl	2.74 ± 0.15	new
2,3,4,5,6-Pentabromophenylalanine	Phe(2,3,4,5,6-Br$_5$)	$-$Br	$-$Br	$-$Br	$-$Br	$-$Br	4.42 ± 0.23	new

Table 3. Tryptophan derivatives: $-NH-CH-CO-$ with CH_2 on the side chain.

Amino acid side chain	Symbol	Side chain position 1 (N^{indol})	4	5	6	7	$\pi \pm$ s.e.	Ref.
Tryptophan	Trp						2.15	[1, 7]
							1.41	[2]
							2.25	[3]
5-Hydroxytryptophan	Trp(5-OH)			$-OH$			1.34 ± 0.02	new
5-Methoxytryptophan	Trp(5-OMe)			$-OCH_3$			2.04 ± 0.11	new
5-Methyltryptophan	Trp(5-Me)			$-CH_3$			1.63 ± 0.09	new
6-Methyltryptophan	Trp(6-Me)				$-CH_3$		1.99 ± 0.09	new
7-Methyltryptophan	Trp(7-Me)					$-CH_3$	1.98 ± 0.11	new
4-Fluorotryptophan	Trp(4-F)		$-F$				1.82 ± 0.09	new
5-Fluorotryptophan	Trp(5-F)			$-F$			1.89 ± 0.10	new
6-Fluorotryptophan	Trp(6-F)				$-F$		1.93 ± 0.11	new
4,5,6,7-Tetrafluorotryptophan	Trp(4,5,6,7-F$_4$)		$-F$	$-F$	$-F$	$-F$	2.84 ± 0.20	new
1-Methyltryptophan	Trp(1-Me)	$-CH_3$					1.90 ± 0.04	new

References

[1] Fauchère, J.-L., Do, K. Q., Jow, P. Y. C., and Hansch, C., *Experientia* **36**, 1203–1204 (1980)

[2] Pliška, V., Schmidt, M., and Fauchère, J.-L., *J. Chromatogr.* **216**, 79–91 (1981)

[3] Fauchère, J.-L., and Pliška, V., *Eur. J. Med. Chem.* **18**, 369–375 (1983)

[4] Fauchère, J.-L., In: *QSAR in Design of Bioactive Compounds.* Kuchar, M. (Ed.). Prous: Barcelona; 135–144 (1984)

[5] Wold, S., Erikson, L., Hellberg, S., Jonsson, J., Sjöström, M., Skagerberg, B., and Wikström, C., *Can. J. Chem.* **65**, 1814–1820 (1987)

[6] Nisato, D., Wagnon, J., Callet, G., Mettefeu, D., Assens, J.-L., Plouzane, C., Tonnerre, B., Pliška, V., and Fauchère, J.-L., *J. Med. Chem.* **30**, 2287–2291 (1987)

[7] Fauchère, J.-L., Charton, M., Kier, L. B., Verloop, A., and Pliška, V., *Int. J. Peptide Protein Res.* **32**, 267–278 (1988)

[8] Rudinger, J., Pliška, V., and Krejčí, I., *Rec. Prog. Hormone Res.* **28**, 131–166 (1972)

[9] Pliška, V., and Rudinger, J., *Clin. Endocrinol.* **5** (Suppl.), 73S–84S (1976)

[10] Pliška, V., Jutz, G., and Beck, S., In: *Peptides: Structure and Function. Proceedings of the Ninth American Peptide Symposium.* Deber, C.M., Hruby, V. J., and Kopple, K. D. (Eds.). Pierce Chemical Company: Rockford, IL; 603–606 (1985)

[11] Pliška, V., and Charton, M., *J. Receptor Res.* **11**, 59–78 (1991)

22 The Application of The Intermolecular Force Model to Bioactivity. Peptide and Protein Quantitative Structure-Activity Relationships

Marvin Charton

Abbreviations

Aax	Amino acid with side chain X
hGH	Human growth hormone
IMF	Intermolecular force
MIC	Minimum inhibitory concentration
Mpa(O)	β-Mercaptopropionic acid sulfoxide
Sta	Statine
vdW	van der Waals

Symbols

BA	Biological activity
K_d	Dissociation constant
K_i	Inhibition constant
MR	Molar refractivity
μ_X	Bond moment of the X-C bond
σ_I	"Inductive effect" substituent constant
σ_F	Field effect substituent constant
υ	Steric effect parameter
ψ	Steric effect parametrization

22.1 Introduction

There are many phenomena which are dependent on the difference in intermolecular forces between an initial and a final state. They include partition, distribution, solubility, phase changes such as melting point and boiling point, and chromatographic properties such as retention times in gas chromatography, relative flow rates in paper and thin-layer chromatography, and in capacity factors in high-performance liquid chromatography, they also include bioactivities. In the now classical Hansch-Fujita approach to modeling bioactivity the most important factor a hydrophobicity-lipophilicity parameter such as log P where P is the partition coefficient, or log k' where k' is the high-pressure liquid chromatography capacity factor (see Chapter 5). It has been shown

that these quantities are themselves composite parameters which are a function of in-termolecular force differences as noted above [1–3]. It should be possible therefore when modeling bioactivities and other properties, to use in place of hydrophobicity or lipophilicity parameters as method based on intermolecular forces. Such a method has been applied successfully to the properties and bioactivities of amino acids, peptides and proteins [4]. In this chapter, recent results on the application of an intermolecular force model to the quantitative description of the bioactivities of peptides and proteins will be reviewed.

22.1.1 The Intermolecular Force (IMF) Equation

Consider Q_x, as the measurable quantity of interest which varies with molecular struc-ture. E is the energy due to the intermolecular forces, X denotes the variable structural feature, and i and f indicate the initial and final states respectively. Then:

$$Q_x = E_f - E_i = \Delta E \tag{1}$$

The intermolecular forces and the factors on which they depend are summarized in Table 1 [3].

22.1.1.1 Intermolecular Force Parameterization

Parameterization of the intermolecular forces described in Table 1 gives the inter/intra-molecular force (IMF) equation which in its most general form [3] was written as:

$$Q_X = L\sigma_{lX} + D\sigma_{dX} + R\sigma_{aX} + M\mu_X + A\alpha_X + H_1 n_{HX} + H_2 n_{nX} + Ii_X + \tag{2}$$
$$B_{DX} n_{DX} + B_{AX} n_{AX} + S\psi_X + B^o$$

where: σ_{lX} is the localized electrical effect parameter. It is identical to the σ_I and σ_F con-stants [5]; σ_{dX} is the intrinsic delocalized electrical effect parameter [5]; σ_{eX} is the elec-tronic demand sensitivity electrical effect parameter [5]; and α is a polarizability para-meter [1–3]. It is defined by the equation:

Table 1. Intermolecular forces and the quantities upon which they depend

Intermolecular force	Quantity
Molecule-molecule	
Hydrogen bonding	E_{hb}
Dipole-dipole	Dipole moment
Dipole-induced dipole	Dipole moment, polarizability
Induced dipole-induced dipole	Polarizability
Charge transfer	Ionization potential, electron affinity
Ion-molecule	
Ion-dipole	Ionic charge, dipole moment
Ion-induced dipole	Ionic charge, polarizability

$$\alpha = \frac{MR_X - MR_H}{100} = \frac{MR_X}{100} - 0.0103 \qquad (3)$$

where MR_X and MR_H are the group molar refractivities of X and H respectively. n_H and n_n are hydrogen bonding parameters [4]. n_H is equal to the number of OH or NH bonds in X, while n_n is equal to the number of lone pairs on O or N atoms in X. i is the ionic charge parameter [4] and takes the value 1 when the substituent is ionized and 0 when it is not. n_D and n_A are charge transfer parameters; n_D is 1 when X acts as an electron donor and 0 when it cannot. n_A is 1 when X can function as an electron acceptor and 0 when it cannot. ψ is an appropriate steric effect parameterization.

22.1.1.2 Steric Effect Parameterization

Several possible parameterizations of the steric effect are possible [6–8]. In choosing among them it is necessary to consider that the steric effect is likely to be a function of the position in the side chain. The simplest method, which is monoparametric, makes use of v [8] which is a composite steric parameter based on van der Waals radii that emphasizes the steric effect at the first atom of the side chain:

$$S\psi = Sv \qquad (4)$$

The side chain is numbered starting with the atom which is bonded to the rest of the amino acid residue. Improving the mode so that it will account for steric effects throughout the side chain requires an investment in additional parameters which is affordable only when a sufficiently large data set is available. There are four of these multiparametric methods available to choose from:

1. The simple branching equation:

$$S\psi = \sum_{i=1}^{m} a_i n_i \qquad (5)$$

This method accounts for the steric effect at each position along the side chain by counting the number of branches at each atom of the skeleton (longest chain). It uses the pure parameters n_1, n_2, and n_3 [6, 8]. The model is directly applicable only to atoms with tetrahedral geometry. It assumes that the effect of all branching atoms attached to a skeletal atom is the same. Such an assumption is justified only as a crude first approximation.

2. The extended branching equation:

$$S\psi = \sum_{i=1}^{m} \sum_{j=1}^{3} a_{ij} n_{if} \qquad (6)$$

This method distinguishes between the first, second, and third branches on a tetrahedral atom at the expense of many more parameters. Few data sets are large enough to permit the use of this method [7, 8].

3. A hybrid model which is a combination of the v steric parameter and the simple branching equation [8]:

$$S\psi = Sv + \sum_{i=1}^{m} a_i n_i \qquad (7)$$

4. The segmental model:

$$S\psi = \sum_{i=1}^{m} S_i v_i \tag{8}$$

where v_i is the steric parameter of the smallest face of the i-th segment of the side chain. The i-th segment consists of the i-th atom of the longest chain and all the groups attached to it [8].

22.1.2 The Composition of the Side Chain Effect

In discussing our results it is often convenient to make use of the per cent contribution of each independent variable in the regression equation, C_i [4]. This quantity is given by the expression:

$$C_i = \frac{|100 a_i x_i|}{\sum_{1}^{m} |a_i x_i|} \tag{9}$$

where a_i is the regression coefficient of the i-th independent variable and x_i is its value for the reference residue. The reference residue chosen must be one for which each independent variable has a value not equal to zero. This requirement can be met either by defining a hypothetical reference group or by choosing an existing residue which meets the requirement. We have generally chosen His as the reference side chain in our amino acid, peptide, and protein studies. Comparisons of side chain structural contributions refer therefore to the His side chain. His was chosen because its parameter values are roughly mid range.

22.1.3 The IMF Equation for Peptide and Protein Bioactivity

It is very unusual to find a dependence on charge transfer interactions in modeling the properties and bioactivities of amino acids, peptides, and proteins. As amino acid side chains are bonded to an sp^3 hybridized carbon atom no terms in σ_{dX} or σ_{eX} are necessary. As the amino acid moiety has a large dipole moment the term in μ is unnecessary. The IMF equation therefore takes the form:

$$Q_X = L\sigma_{lX} + + Aa_X + H_1 n_{HX} + H_2 n_{nX} + Ii_X + S\psi_X + B^{\circ} \tag{10}$$

22.2 The IMF Method as a Bioactivity Model

22.2.1 The Hansch-Fujita Model

It was shown that if all the necessary pure parameters are included in the composite parameters used, then a model constructed from composite parameters is completely equivalent to one which uses pure parameters in representing the data [9]. As was noted above, a pure parameter represents a single effect, a composite parameter represents two or more effects. The reason for using pure parameters is the ease of interpretation of the results. In its use of lipophilicity parameters such as log P, log k', or π, the Hansch-Fujita model uses composite parameters [1–3].

The Hansch-Fujita model often requires in addition to transport parameters the use of electrical effect, steric effect, and polarizability parameters and sometimes also dipole moment [10] parameters. These parameters are needed because the mix of pure parameters found in a particular transport parameter may not be that which is required for a particular type of bioactivity. This is no surprise. *The probability that all biomembranes and all receptor sites will require the same mix of pure parameters is extremely small.* This conclusion is supported by the review of Seydel and coworkers [11]. The addition of electrical, steric, and polarizability terms provides the proper parameter composition for mode the bioactivity being studied. To illustrate the point let us consider a typical Hansch-Fujita correlation equation:

$$BA_X = T\tau_X + \varrho\sigma_X + AMR_X + Sv + B^\circ \tag{11}$$

where BA is the bioactivity, σ is a composite electrical effect parameter of the Hammett type, τ is a transport parameter such as $\log P$, π, or $\log k'$, MR is the group molar refractivity, v is a steric parameter, and T, ϱ, A, S, and B°, are coefficients. $\log 1/C$ is often used in place of $\log BA$. The reason for the use of $\log BA$ as the independent variable in quantitative structure-activity relationships (QSAR) results from the use of physicochemical parameters as descriptors. The justification for the use of these parameters is that bioactivities themselves are physicochemical phenomena. All of the steps in the bioactivity model described in section 22.3 may be considered to involve either equilibria or rates. Such bioactivity parameters as k_{cat} and K_M from enzyme kinetics, the binding dissociation constant K_d and the inhibitor constant K_i are clearcut examples. Bioactivities of other type are assumed to be related to equilibrium or rate constants. There is general agreement among physical organic chemists that ΔG, the change in the Gibbs free energy, is the best measure of structural effects on an equilibrium process. At a given temperature ΔG is a linear function of $\log K$. From transition state theory the same argument applies to the log of the rate constant. It follows then that $\log BA$ is the proper choice for the independent variable in QSAR.

As was noted above σ is given by the expression:

$$\sigma_X = l\sigma_{lX} + d\sigma_{dX} + ro_{eX} + h \tag{12}$$

Eq. (3) gives:

$$MR_K = 100(\alpha_X + 0.0103) = 100\alpha_X + 1.03 \tag{13}$$

Eq. (8) gives:

$$Sv = S_1v_1 + S_2v_2 + S_3v_3 + S_o \tag{14}$$

τ is given by the equation:

$$\tau_X = L\sigma_{lX} + D\sigma_{dX} + R\sigma_{aX} + A\alpha_X + H_1n_{HX} + H_2n_{nX} + Ii_X + M\mu_X + \\ + S_1v_{1X} + S_2v_{2X} + S_3v_{3X} + B^\circ \tag{15}$$

Substituting Eqs. (12) through (15) into Eq. (11) results in:

$$BA_X = (L + \varrho l)\sigma_{lX} + (D + \varrho d)\sigma_{dX} + (R + \varrho r)\sigma_{aX} + (A + 100A)\alpha_X + H_1n_{HX} \tag{16} \\ + H_2n_{nX} + Ii_X + M\mu_X + (S_1 + SS_1)\,v_{1X} + (S_2 + SS_2)\,v_{2X} \\ + (S_3 + SS_3)\,v_{3X} + B^\circ + \varrho h_o + 1.03A + SS_o$$

which may be rewritten as:

$$\log BA_X = L'\sigma_{IX} + D'\sigma_{dX} + R'\sigma_{aX} + A'\alpha_X + H_1 n_{HX} + H_2 n_{nX} + Ii_X + M\mu_X \quad (17)$$
$$+ S'_1 v_{1X} + S'_2 v_{2X} + S'_3 v_{3X} + B'_o$$

This is a form of the IMF equation. Then bioactivity is a function of the difference in intermolecular forces between initial and final states. This does not mean that transport parameters should not continue to be used in modelling bioactivities. It simply provides an explanation of why and how they work.

22.2.2 Alternatives to the Use of Lipophilicity Parameters

A very important point must be made here. As any combination of pure and/or composite parameters which has the correct composition will serve to quantitatively describe a phenomenon it is *not necessary* to use log P or π values. Bioactivities can be correlated directly with the IMF equation. Other alternatives to the use of lipophilicity parameters have been proposed (see Chapters 4 and 11).

It is convenient at this point to consider what determines the choice of a model for the obtention of QSAR. This is determined by the reasons for which the QSAR is sought. If the objective is to aid in the design of a bioactive molecule and it is felt that an empirical relationship is sufficient for this purpose, then any method which produces such a relationship will serve. Large numbers of parameters can be generated by quantum chemical or topological methods or selected from a wide range of known physical properties and by means of appropriate statistical methods can be used to obtain empirical equations. This is what might be considered an engineering approach. An alternative is to start with a model of bioactivity, and use physico-chemical parameters whose interpretation is unequivocal because it has been established in well-understood relationships for chemical reactivities and physical properties of simple molecules. This is the approach that has been used here.

22.3 Bioactivity Model

Bioactivity results from the interaction of some chemical compound termed the bioactive substance and a living organism or some component of one. The component may range from a pure enzyme through a whole cell homogenate to a whole tissue. Organisms may range from viruses through single-cell bacteria, algae, yeast, and protozoa to mammals. In order to justify the application of the IMF model to bioactivities, it is necessary to consider the way in which a bioactive substance expresses its bioactivity. The model used in this work is a modification of that proposed by McFarland [4, 12]. The bioactivity is considered to be dependent on one or more of the following sequences.

1. Transport. The substance enters the organism at some point. It then moves to a receptor site with which it is to interact. This movement is through an aqueous phase. It may involve diffusion through the medium or random binding to a biopolymer mole-

cule such as plasma protein which carries it. During its passage, it is likely to cross one or more biomembranes. The crossing of a biomembrane begins with the transfer of the substance from the initial aqueous phase Φ_i to the anterior membrane surface. It then proceeds to the posterior membrane surface either by diffusion, or by complexing with a lipid-soluble membrane carrier molecule which transports it. The substance is then transferred from the surface to a second aqueous phase Φ_f. Each step in this process is equivalent to a transfer from one phase to another. It is therefore a function of the difference between intermolecular forces involving medium and substance in the initial and final phases.

2. Receptor-substance interaction. A receptor can be considered to be a set of functional groups bonded to a molecular framework whose interaction with a bioactive substance results in the observed bioactivity. Receptors may be membrane-bound (G-protein-coupled receptor and ion channels are examples), cytosolic or nuclear, or part of an enzyme or some other type of biopolymer that can act as an enzyme. The interaction between receptor and substance can be divided into two parts, recognition and tight complex formation.

   a) Recognition. It is necessary for the receptor to distinguish the substance from all of the other chemical species present in the medium which surrounds it. The receptor consists of some number of functional groups attached to a framework. These functional groups have a particular orientation in space. To be recognized, a substance must have functional groups that are capable of interacting with those of the receptor and have the proper spatial arrangement to do so. The result of recognition is the formation of a loose substance-receptor complex held together by intermolecular forces. Recognition is therefore a function of the difference between the intermolecular forces involving medium and the substrate in the aqueous phase surrounding the receptor site, plus those involving the receptor and the aqueous phase in which it is immersed; and the intermolecular forces between substrate and receptor in the loose complex.
   b) Tight complex formation. Conformational changes can occur in the substance and/or in the receptor which maximize the intermolecular forces between the two. This results in an increase in binding energy that accompanies the formation of a tight complex. The process is a function of the difference in intermolecular forces between the initial loose complex and the final tight complex.

3. Chemical reaction. This is applicable only in some special cases when the receptor is located on an enzyme or some other biopolymer such as an antibody or RNA that can function as an enzyme. The tight complex forms a transition state that decomposes into a receptor-product complex by the formation and/or cleavage of covalent bonds. The receptor-product complex then dissociates into solvated receptor and solvated product. The key features of model are summarized in Fig. 1.

   Each step in the sequences described above involves a difference in intermolecular forces between an initial and a final state. The IMF equation was designed expressly to model such differences. Clearly then it should be capable of modeling bioactivities.
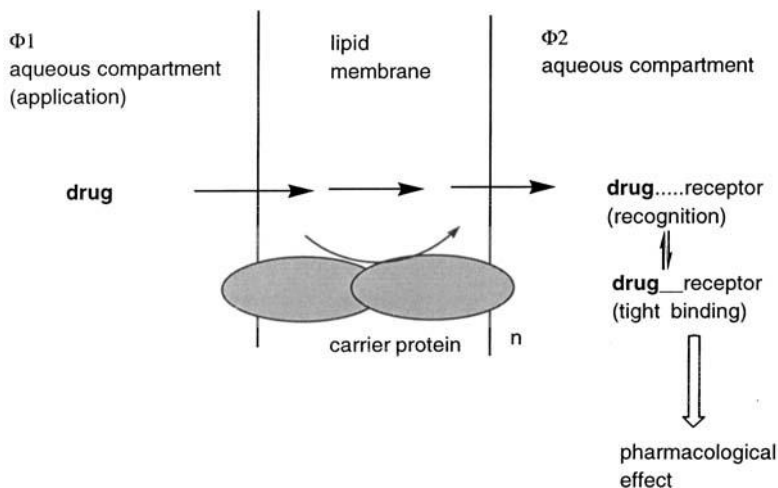
Figure 1. Drug transport to the receptor. From an aqueous phase $\Phi 1$ (e.g., the gastrointestinal tract) a drug enters a membrane, diffuses through it, and reaches another aqueous compartment $\Phi 2$ (e.g., the bloodstream or the intracellular cytosol). Transport through a membrane may occur n times. Binding to the receptor may involve two steps, recognition, and finally tight binding, which leads to a pharmacological effect (cf. Chapters 16, 17).

This is a general model. All receptor types share some form of step 1. There will be variation in detail in step 2, depending on the receptor type. This should not affect the conclusion that bioactivity involves a series of transfers from an initial to a final phase, that these transfers involve equilibria or rates, and that these transfers depend on the difference in intermolecular forces between the initial and final states.

22.4 Peptide Bioactivities

22.4.1 Types of Structural Variation in Peptides

Peptides may undergo substitution in any combination of several different sites [4]. Amino acid residues may be varied at a particular position in the peptide. This is designated as residue substitution and represented by Aax^i where Aax is the residue with side chain X and i is its position in the peptide. Substitution at the amino terminus of a linear peptide is represented by X^N, and at the carboxyl terminus by X^C. Substitution at the nitrogen atom of a peptide bond is represented by X^{Pij} where i and j are the positions of the residues attached to the atom undergoing substitution. Finally, one or more amino acid residues in a peptide may be replaced by groups which are not amino acids. This is represented by X^{Rij} where $i, j,\ldots$ designates the positions of residues which are being replaced. Substitution of a residue in the i-th position by either another amino acid residue or a nonamino acid fragment is called replacement and designated Rpl^i. Transition state analogs, peptide mimetics, and peptidomimetics are examples of this type of substitution. For example, consider the peptide

t-BuOCO$_2$-Ser-Ala-Thr-His(Et)-Asp-Arg-Aax-NHCH2CH=CHCH$_2$CO-Tyr-OPh

X^N $X^{P4,5}$ Aax^7 Rpl^{8,9} X^C
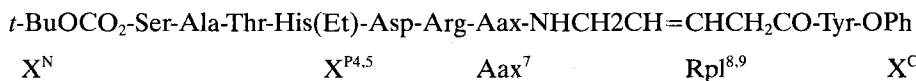
The X^N substitution is the *t*BuOCO$_2$ group, the X^C substituent is the OPh group, the $X^{P4,5}$ substituent is the Et group, the NHCH$_2$CH=CHCH$_2$CO group is Rpl^{8,9}, and the side chain of Aax^7 varies.

22.4.2 Peptide QSARs

All of the peptide and protein QSARs referred to in this work were carried out using the amino acid side chain parameters given in our compilation [4]. A number of peptide QSARs were described by Charton [4] in which examples of all types of peptide substitution were outlined.

1. Bacteriostatic small peptides. Of particular interest was the development of QSAR for the bacteriostatic peptide analogs Aax-Ala(P) where Aax is an amino acid with side chain x and Ala(P) has the structure H$_2$NCHCh$_3$PO$_3$H$_2$. The bacteria studied were two species of Gram-positive cocci and six of Gram-negative facultatively anaerobic rods of which five belong to the family Enterobacteriaceae. The sixth is of uncertain family. An attempt to combine them all into a single data set using the Omega method [4] was unsuccessful. However, combination into one set of Gram-positive species and one of Gram-negative species was successful. The difference in the QSAR for the two types of bacteria was largely due to a different dependence on steric effects.

Also of interest were data for the inhibition of ^{125}I-ristocetin binding to *Micrococcus* cell wall by the acylamino acids Ac-Aax, and the peptides Ac-Aax-Aax, and Ac-Aax-Aax-Aax [13] was studied. A QSAR which included all three types of compound was obtained [4].

2. Inhibitors of oxytocin. Side chain contributions obtained by the application of the Free-Wilson method [14] from pA_2 values (where A_2 is the IC$_{50}$ or ID$_{50}$) for 155 structural analog of oxytocin (Fig. 2) that have an inhibitory effect on oxytocin in isolated rat uterus in the absence of magnesium [15] were studied. Most of these substrates were nonapeptides substituted at all positions except position 5; two of them had X^N substitution as well. The pA values were taken from the literature. Substitution at position 1 of the peptide involved three different sites as the amino acid residue at this position has the form Z-CRX(C=O)− where X and Z are substituents and R may be H or Me. If a substituent is part of a disulfide bridge it is considered to be the X group. If only one substituent was present it is considered to be the X group. Not all of the substitutions at position 1 involve amino acids. As none of the X groups had OH or NH bonds, or with the exception of Mpa(O) had lone pairs on O or N atoms; and no X group was likely to ionize, the X group was parameterized only by σ_{IX}, α_X, and v_X. The Z group was parameterized by a complete set of IMF parameters. The R group was represented by the parameter n_{Me} which took the value 1 when R and Me and 0 otherwise.
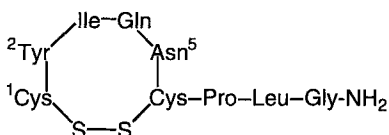
Figure 2.   Structure of oxytocin.

The results obtained [16, 17] suggest that the Z group is involved in binding by induced dipole-induced dipole (dispersion) interactions, and the X group by hydrogen bonding and Van der Waals (vdW) interactions. The R group seems to have no effect on the activity.

Substitution at position 2 involved only the replacement of one amino acid residue by another. Substitution at this position is complicated, however, by the inclusion in this data set of both D- and L-amino acid residues. An attempt to account for configuration by means of an indicator variable that took the value 1 for the D configuration and 0 for the L configuration was totally unsuccessful. The data set was therefore separated into a D and an L set, and these sets were separately correlated successfully [16, 17].

3. Renin inhibitors. Bioactivities of peptide renin inhibitors [18–24] have received considerable attention. They are of interest in the development of drugs for the treatment of hypertension and are peptide analogs of human angiotensinogen and of aspartyl proteinase pepstatin (see Fig. 3). Several of these sets involve residues 8, 9, 10, and 11 of angiotensinogen. The residues Leu^{10}-Val^{11} in these sets are replaced by a nonpeptide structural unit, an example of X^R substitution. Residue 8 may undergo replacement by either another residue or a nonpeptide fragment, residue 9 may vary, and there may also be X^N or X^C substitution. One set consists of derivatives of pepstatin in which residue 1 is Phe or Trp, residue 2 varies, and both X^N and X^C substitution occur.

The sets studied were correlated with an appropriate form of the IMF equation. Also studied were IC_{50} and K_i values determined for peptides having the structure Boc-Phe-His-(3S,4S)-Sta-Leu-X^C, where Sta is the amino acid 3-hydroxy-4-amino-6-methyl-heptanoic acid (see Fig. 3).

Sta is thought by some to be a replacement for two amino acid residues. The fragment NHCH(*i*Bu)CHOH is indeed equivalent to Leu, the remaining fragment, $CH_2(C{=}O)$, needs one more atom other than hydrogen in order to be a valid replacement of an amino acid residue. The substrates in these data sets are therefore roughly equivalent to pentapeptide angiotensinogen analogs in which Sta replaces Leu^{10}-Val^{11}.

H_2N-Asp^1-Arg^2-Val^3-Tyr^4-Ile^5-His^6-Pro^7-Phe^8-His^9-Leu^{10}-Val^{11}-Ile^{12}-His^{13}-*protein*

    human angiotensinogen

Iva-Val^1-Val^2-Sta-Ala-OH   H2N-CH(iBu)CH(OH)CH₂-COOH
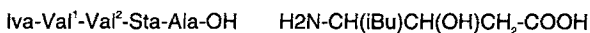
    pepstatin   statin (Sta)

Figure 3.   Structures of angiotensinogen, pepstatin and statin.

The structural variation is in X^C, which can be written in the form NHW where W is CHZ^1Z^2. The structural effects are modeled by $\sum\sigma_{IZ}$, $\sum\alpha_Z$, $\sum n_{HZ}$, n_1, n_2 and n_3. For four of the X^C substituents in the data sets studied the carbon atom bonded to nitrogen is chiral. Values of the bioactivity have been determined for both enantiomers in three cases and for the meso form and the racemate in the fourth case. The configurational isomerism is accounted for by the n_{cf} parameter defined above. The use of this parameter in a data set which contains both chiral and achiral members is based on the assumption that the achiral group will prefer a conformation analogous to that of one of the two chiral isomers and will therefore behave in the same way as that isomer, while the activity of the other chiral isomer will be either greater or less throughout the data set. Although one of the substituents studied, 4-amido-1-benzylpiperidine, can ionize it was not necessary to include parameterization to account for this. The results of the correlations were used to propose guidelines for the design of angiotensin analogs [25].

22.5 Protein Bioactivities

22.5.1 Types of Protein Bioactivity Data Sets

There are two major types of protein bioactivity data sets:

1. Sets in which substitution occurs at a single position in the protein. In this type of data set a residue occupying a given position in the wild-type (native) protein that is known to be involved in the bioactivity is replaced by a number of other residues and the activities of these mutant proteins are determined. The activities can be correlated with an appropriate form of the IMF equation. An example of this type of data set is the determination by Alber and coworkers [26, 27] of the relative activities of Phage T4 lysozomes substituted at position 86. A variation of this type of data set occurs when substitution occurs at two positions. An example of this is the study by Wallace and coworkers of substitution in recombinant hirudins at positions 1 and 2 [28].

2. Residues which are known to be part of a receptor in a bioactive protein are substituted for and the activities of the resulting mutants are determined. In correlation this type of data set it is necessary to assume that only structural effects on the receptor site are of importance and therefore the effect of the side chains of nearby residues on that of the residue that has undergone substitution is negligible. The parameterization of the IMF equation in this case must represent the difference is effect between the side chain in the wild-type and that in the mutant. An example of this type of substitution is the report of Fersht et al. on kinetically determined $\Delta\Delta G$ values for the binding of ATP and of Tyr to tyrosyl-tRNA synthetase [29, 30]. A variation on this type of study is that of Cunningham and Wells [31–33] which involves alanine substitution of each residue.

Correlation of this data set requires a modified form of the IMF equation:

$$Q_{Xi}-X_f = L\sigma^\Delta_1 + A\alpha^\Delta + H_1 n_H{}^\Delta + H_2 n_n{}^\Delta + Ii^\Delta + Sv^\Delta + B_1 n_1{}^\Delta + B_2 n_2{}^\Delta + \qquad (18)$$
$$+ B_3 n_3{}^\Delta + B^o$$

where the superscript Δ indicates the difference between the parameter value of the final side chain and that of the initial side chain. Thus, when the amino acid residue Aax^i is replaced by the residue Aax^f, the value of v^Δ is given by:

$$V^\Delta = V_x^f - V_x^i \tag{19}$$

where v is one of the independent variables in Eq. (3) while v_{Xi} and v_{Xf} are its values for the initial and final side chains respectively.

22.5.2 Protein QSAR

The first type of protein substitution is exemplified by a study of the relative activities reported by Alber et al. [26] that were based on the quantity required to give a clear area of 7.00 mm radius in the Lysoplate assay. Correlation with the IMF equation gave a QSAR which indicated that the dominant effect was hydrogen bonding with polarizability and steric effects also contributing.

An example of the general case of the second type of protein substitution is the study of the $\Delta\Delta G$ values measured by Fersht et al. [29] for the binding of ATP and Tyr to tyrosyl-tRNA synthase. The two sets of binding data were combined into a single data set by means of an indicator variable. Correlation with Eq. (18) gave very good results. Polarizability, ionic charge, and steric effects were the factors which determined activity.

The alanine scanning mutagenesis method has been applied to residues that are within the three regions of human growth hormone hGH that are said to be involved in receptor recognition (residues 2–19, 54–74, and 167–191), [31]. Binding dissociation constants, K_d, were determined for each of the mutant hGH–hGH liver receptor site complexes substituted at residues which are part of the receptor In two cases the alanine-substituted mutant could not be prepared and a different substitution product was used in its place.

In this type of protein bioactivity study only those residues which are either part of or interact with a receptor site need be considered. Residues which are not involved in binding act as a skeletal group to which the active residues are attached. It is assumed that their side chains take no part in the observed bioactivity. Cunningham and Wells have considered all residues for which the ratio of $K_{d,mutant}$ to $K_{d,wild-type}$ is > 4 to be possibly involved in binding. Correlation with Eq. (18) indicated that structural effects are determined by the electrical effect, hydrogen bond H donor capacity and steric effector of the side chain.

22.5.3 Limitation of the Model in Protein QSAR

It is important to review the disadvantages of the IMF model, particularly as applied to proteins [4]. They include:

1. The protein data set almost always is restricted to the 19 structurally similar amino acid residues other than proline normally found in protein (protein basis set). The IMF

equation for proteins requires from six to nine independent variables depending on the choice of steric effect parameterization. This results in two problems.

  a) The possibility of a chance correlation is comparatively large due to the large number of independent variable required and the small number of data points generally available.

  b) There are colinearities among the independent variables which are inherent in the protein basis set. Some workers prefer to avoid the use of multiple linear regression analysis in this case. Others are suspicious of the validity of alternative statistical methods that have been proposed.

2. In general, α-amino substituted residues require two additional independent variables to account for their structural effect. The proline type of residue requires one. As its inclusion in the data set adds only one degree of freedom the model cannot be applied to it.

References

[1] Charton, B., Bulk and Steric Parameters in Binding and Reactivity of Bioactive Compounds. In: *Rational Approaches to the Synthesis of Pesticides.* P. S. Magee, J. J. Menn and G. K. Koan (Eds.). American Chemical Society: Washington, D.C.; 247–278 (1984)

[2] Charton, M., The Nature of Transport Parameters. In: *Trends in Medicinal Chemistry '88,* van der Goot, H., Domany, G., Pallos, L. and Timmerman, H. (Eds.). Elsevier: Amsterdam; 89–108 (1989)

[3] Charton, M., Transport Parameter Dependence on Intermolecular Forces. In: *Classical and 3-D QSAR in Agrochemistry and Toxicology,* C. Hansch and T. Fujita (Eds.). American Chemical Society: Washington, D. C.; 75–95 (1995)

[4] Charton, M., *Prog. Phys. Org. Chem.* **18,** 163–284 (1990)

[5] Charton, M., *Prog. Phys. Org. Chem.* **16,** 287–315 (1987)

[6] Charton, M., The Prediction of Chemical Lability Through Substituent Effects. In: *Design of Biopharmaceutical Properties Through Prodrugs and Analogs,* Roche, E. B. (Ed.). American Pharmaceutical Society: Washington, D. C; 228–280 (1977)

[7] Charton, M., *Top. Current Chem.* **114,** 57–91 (1983)

[8] Charton, M., The quantitative description of steric effects. In: *Similarity Models in Organic Chemistry, Biochemistry and Related Fields.* Zalewski, R. I., Krygowski, T. M., and Shorter, J (Eds.). Elsevier: Amsterdam; 629–687 (1991)

[9] Charton, M., Greenberg, A., and Stevenson, T. A., *J. Org. Chem.* **50,** 2643–2646 (1985)

[10] Lien, E. J., Guo, Z.-R., Li, R.-L., and Su, C.-T., *J. Pharm. Sci.* **71,** 641–655 (1982)

[11] Seydel, J. K., Coats, E. A., Cordes, H. P., and Wiese, M., *Arch. Pharm. (Weinheim)* **327,** 601–610 (1994)

[12] McFarland, J. W., *Prog. Drug Res.* **15,** 123–146 (1971)

[13] Kim, K. H., Martin, Y. C., Otis, E. R., and Mao, J. C., *J. Med. Chem.* **32,** 84–93 (1989)

[14] Free, S. M., and Wilson, J. W., *J. Med. Chem.* **7,** 395–399 (1964)

[15] Pliška, V., and Heininger, J., *Int. J. Peptide Protein Res.,* **31,** 520–536 (1988)

[16] Pliška, V., and Charton, M., *Proc. 11th Am. Peptide Symp.* 290–292 (1990)

[17] Pliška, V., and Charton, M., *J. Receptor Res.* **11,** 59–78 (1991)

[18] Dellaria, J. F., Maki, R. G., Bopp, B. A., Cohen, J., Kleinert, H. D., Luly, J. R., Merits, I., Plattner, J. J., and Stein, H. H., *J. Med. Chem.* **30,** 2137–2144 (1987)

[19] Nisato, D., Wagnon, J., Callet, G., Mettefeu, D., Assens, J.-L., Plouzane, C., Tonnerre, B., Pliška, V., and Fauchère, J.-L., *J. Med. Chem.* **30,** 2287–2291 (1987)

[20] Bock, M. G., DiPardo, R. M., Evans, B. E., Rittle, K. E., Boger, J., Poe, M., LaMont, B. I., Lynch, R. J., Ulm, E. H., Vlasuk, G. P., Greenlee, W. J., and Veber, D. F., *J. Med. Chem.* **30,** 1853–1857 (1987)

[21] Bolis, G., Fung, A. K. L., Greer, J., Kleinert, H. D., Marcotte, P. A., Perun, T. J., Plattner, J. J., and Stein, H. H., *J. Med. Chem.* **30,** 1729–1737 (1987)

[22] Kempf, D. J., de Lara, E., Stein, H. H., Cohen, J., and Plattner, J. J., *J. Med. Chem.* **30,** 1978–1983 (1987)

[23] Luly, J. R., Yi, N., Soderquist, J., Stein, H. H., Cohen, J., Perun, T. J., and Plattner, J. J., *J. Med. Chem.* **30,** 1609–1616 (1987)

[24] Hui, K. Y., Carlson, W. D., Bernatowicz, M. S., and Haber, E., *J. Med. Chem.* **30,** 1287–1295 (1987)

[25] Charton, M., Peptide Renin Inhibitor QSAR. In: *QSAR in Design of Bioactive Compounds*, Vol. **2**, J. Prous, Barcelona (1992) p. 329–346

[26] Alber, T., Bell, J. A., Dao-Pin, S., Nicholson, H., Wozniak, J. A., Cook, S., Matthews, B. W., *Science* **239,** 631–635 (1988)

[27] Charton, M., *Coll. Czech. Chem. Commun.* **55,** 273–281 (1990)

[28] Wallace, A., Dennis, S., Hofsteenge, J., and Stone, S. R., *Biochemistry* **28,** 10079–10084 (1989)

[29] Fersht, A. R., Shi, J.-P., Knill-James, 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)

[30] Charton, M., *Int. J. Peptide Protein Res.* **28,** 201–206 (1986)

[31] Cunningham, B. C., and Wells, J. A., *Science* **244,** 1081–1085 (1989)

[32] Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P., and Estell, D. A., *Proc. Natl. Acad. Sci. USA* **84,** 1219–1223 (1987)

[33] Charton, M., *J. Chim. Phys.* **89,** 1689–1701 (1992)

23 Lipophilicity Descriptors for Structure-Property Correlation Studies: Overview of Experimental and Theoretical Methods and A Benchmark of log *P* Calculations

Han van de Waterbeemd and Raimund Mannhold

Abbreviations

ALOGP	Calculated 1-octanol/water log *P* values, based on atomic contributions
ASCLOGP	Conformation-dependent log *P* values based on approximate surface calculation
BLOGP	Log *P* calculation based on molecular properties, e.g., obtained by molecular orbital calculations
CASE	Computer-automated structure evaluation
CHEMICALC	Computer program to calculate log *P* values based on atomic contributions developed by Suzuki
CLOGP	Calculated 1-octanol/water log *P* values, based on fragmental contributions; or algorithm/computer program developed by Leo
CPC	Centrifugal partition chromatography
HINT	Computer program for the visualization of hydrophobic surfaces and for the calculation of log *P* values developed by Abraham
KLOGP	Log *P* calculations based on computer-identified fragments developed by Klopman
PROLOGP	Computer program to calculate log *P* values based on Rekker's fragmental method
QSAR	Quantitative structure-activity relationships
RP-(HP)LC	Reversed-phase (high-performance) liquid chromatography
RP-TLC	Reversed-phase thin-layer chromatography
SANALOGP	Computer program to calculate log *P* values based on Rekker's method developed by Petelin
SMILOGP	Computer program to calculate log *P* values based on atomic contributions developed by Dubost
SPC	Structure-property correlations

Symbols

f	Rekker fragmental constant
$\log k$	Isocratic capacity factor from RP-HPLC
$\log k_w$	Polycratic or extrapolated RP-HPLC lipophilicity index

$\log P_{oct}$ 1-Octanol/water partition coefficient

π Hansch-Fujita substituent constant

23.1 Introduction

Lipophilicity is well-known as a prime physicochemical descriptor of xenobiotics with relevance to their biological properties. The hydrophobic interactions of drugs with their receptors, the pharamacokinetic behaviour of drug molecules, toxicological properties as well as pharmaceutical aspects like solubility are examples for a steadily increasing number of topics in which lipophilicity plays an important role. Lipophilicity or hydrophobicity may be considered for a complete molecule or part of it. Therefore one may distinguish between global (molecular) and local (substructural) lipophilicity descriptors. The widespread application of lipophilicity values to – in many cases – large numbers of xenobiotics easily explains the need for both valid and quick procedures to quantify molecular or substructural lipophilicity. In this context computational approaches are superior to experimental procedures. Routine application of calculative approaches, however, require a continuous check of their validity by comparing with experimental procedures. The present chapter reviews the key techniques for experimental and computational lipophilicity assessment. Moreover, we have undertaken a benchmark of a selection of log P calculation programs.

23.2 Experimental Lipophilicity Scales

23.2.1 Shake-Flask Partitioning

23.2.1.1 Solvent/Water Systems

Lipophilicity is defined by the partitioning of a solute between aqueous and nonaqueous phases. Its quantitative descriptor, the partition coefficient P, expresses the ratio of monomeric, neutral solute concentrations in the organic and aqueous phase of a two-component system under equilibrium conditions.

The organic solvent/water system of choice is 1-octanol. Advantages of this solvent, and a discussion of its physico-chemical properties, are summarized in [1–4]. In contrast to its simple definition, the determination of P poses quite often practical problems, particularly in case of polar or highly lipophilic solutes. Impurity and instability of the solute can produce unreliable experimental data. Among the various precautions to be observed for an accurate measurement of P, presaturation of the phases, the use of low solute concentrations, centrifugation for a proper separation of the phases, and the determination of solute concentration in both phases deserve mention here. Detailed summaries of the experimental prerequisites for precise measurements of partition coefficients including the aspects of ion-correction are given by e.g., Leo et al. [1], Kubinyi [4] and Taylor and coworkers [5–7].

Although 1-octanol/water partitioning is a widely accepted standard for measuring lipophilicity [1], there is ample evidence that other solvent systems may also be of con-

Table 1. Frequently used solvent systems

Solvent pair	character	Reference
1-Octanol/water	Amphiprotic (standard)	[1,2]
Alkane/water	Inert	
hexane/water		[5–7]
cyclohexane/water		[8]
heptane/water		[9]
isooctane/water		[10]
dodecane/water		[11]
Chloroform/water	H-donor	[5–7]
Di-*n*-butylether/water	H-acceptor	[12]
PDGF/water		[5–7]
Hexane/PE glycol	Total H-bonding	[13]

siderable interest [5–7] (see Table 1). Particularly, e.g., penetration into the brain does not always correlate well to log P_{oct} values. Furthermore, it was demonstrated that the behavior of lipophilic ionized compounds in lipid membrane-water systems differs from octanol/water [14].

23.2.1.2 Aqueous Biphasic Systems

A largely unexplored field is the description of the relative lipophilic properties of biologically active compounds in aqueous polymer biphasic systems [15, 16].

23.2.2 Chromatographic Methods

Since the measurement of accurate log P values is limited to a certain range of log P values (e.g., for octanol/water from -3 to $+3$), alternative methods have been explored to obtain an assessment of lipophilicities. In particular, chromatographic methods appear to be well suited for this purpose. The most important are RPLC, RP-TLC and CPC. These are more extensively discussed in other chapters (see Chapters 5–8) and will only be briefly summarized here.

23.2.2.1 RP-TLC

As substitutes for partition coefficients chromatographic parameters obtained from reversed-phase thin-layer chromatography (RP-TLC) can be used. RP-TLC plates, coated with hydrophobic material, are eluated with aqueous/organic solvent systems. Obtained R_f-values allow the calculation of R_M-values as true measures of lipophilicity according to Bate-Smith and Westall [17]:

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

Accuracy and reproducibility of chromatographic R_M data depends strictly on the application of standardized experimental conditions. Control of temperature and humidity, the use of front markers, the densitometric evaluation of starting and running points, and the extrapolation to modifier-free conditions deserve mention here. Provided that RP-TLC data are obtained under well-controlled experimental conditions as described above, they represent a valuable alternative to the tedious and time-consuming direct measurement of partition coefficients.

Important advantages of RP-TLC are, *inter alia*, that test compounds need not be pure and only trace amounts of test material are necessary. Compounds can be investigated over a broad lipophilicity range, and last but not least, a quantitative determination of their concentration (often posing analytical problems) is not necessary.

Details of a precise measurement of lipophilicity data by RP-TLC are given in Chapter 8 of this volume.

23.2.2.2 RP-HPLC

Reversed-phase high-performance liquid chromatography (RP-HPLC or RPLC) has been widely used to measure lipophilicities of biologically active chemical entities. The advantages include the broad range of lipophilicity values which can be covered, the small amount of substance required–which may even be impure or unstable- and the ease of automation. Since there is a great choice of possible stationary phases, several scales of lipophilicity indices may be obtained. The data obtained with certain columns closely mimic 1-octanol/water log P values. However, it should be realized that there is no general equation to correlate HPLC lipophilicity values to log P_{oct} values, although this is sometimes suggested [18]. Each structural class has its own correlation. (For details see Chapter 5.)

23.2.2.3 CPC

Centrifugal partition chromatography is a rather new chromatographic technique, which can be used for separation as well as for lipophilicity measurements. With this technique log P values in 1-octanol/water or hexane/water can be measured directly, but with practically the same limitation, i.e., a restricted lipophilicity range, as shake-flask measurements. (See Chapter 6.)

23.2.3 Alternative Experimental Methods

Various other techniques have been proposed for the determination of log P values or related lipophilicity indices. Reviews can be found in references [19, 20]. Some examples will be discussed here briefly.

23.2.3.1 Slow Stirring

By monitoring the changes in concentration in a partitioning process using the slow stirring technique, the log P of rather lipophilic compounds can be obtained [21].

23.2.3.2 Filter Probe and Filter Chamber

Both the filter probe [22] and its revised filter chamber technique [20] are more rapid and economical than the traditional shake flask, but limited to the range of ca. -1.3 to $+2.8$.

23.2.3.3 Flow-Injection Extraction

Single-channel coaxial segmenters can be used for the introduction of an aqueous solution into an organic phase [23]. The disadvantages, as reported by the author, are the requirement of a fast-reading detection system and the complicated treatment of the analytical signal.

23.2.3.4 Microscale Partitioning Apparatus

A rapid and reliable method has been developed, using a commercially available mixer-separator device [24]. Relative concentrations are measured by HPLC. The procedure requires only 10 μg of sample.

23.2.3.5 pH-metric log *P* Determination

Although the principles of measuring log *P* values from phase potentiometric titrations go back to early work in the 1960s and 1970s, this method became increasingly popular only very recently. It has now been fully automated [25] (see Chapter 7).

23.3 Calculated log *P* Values

23.3.1 Overview

23.3.1.1 The π-system

The constitutive-additive property of log *P* values has been recognized by Fujita and Hansch in the early 1960s [26]. It is based on the summation of substituent values, now well-known as π values. Briefly, the log *P* of a substituted compound RX (where R is the backbone, and X the substituent) can be obtained from the log *P* of the parent compound RH and the lipophilicity contribution π of the substituent:

$$\log P_{RX} = \log P_{RH} + \pi_X \tag{2}$$

By its definition, $\pi_H = 0$. This is of course not correct, because hydrogen atoms also contribute to the overall lipophilicity. Therefore, Rekker [27–32] and later Leo [33–37], introduced fragmental contributions to the lipophilicity, f, instead of a substituent value, π. Since the definition of a fragment is not unambiguous, Broto et al. [38], and later others, have developed a log *P* calculation system based on atomic contributions. Finally, some approaches are based on molecular properties, which may be derived with molecular orbital calculations. An overview of methods is presented in Table 2.

Table 2. Programs and methods for calculation of log *P* values

Program	Method	Reference
ALOGP		
MOLCAD, TSAR	Atomic values	[38–43]
ATOMIC5	Atomic values	
HINT	MLP	[44]
SMILOGP	Atomic contributions	[45]
AUTOLOGP	Autocorrelation	[46]
BLOGP		
--[a]	Charge densities	[57]
BLOGP	Molecular descriptors	[49, 50]
MLPLOGP	MLP	[53]
ASCLOGP	Approximate Surface	[54, 55]
MOLFESD	Free energy surface densities	[56]
--[a]	Structural parameters	[47, 48]
--[a]	Solvatochromic parameters	[51]
--[a]	Graph theoretical descriptors	[52]
CLOGP		
CLOGP	Hansch-Leo	[33–37]
Σf	Rekker/Mannhold, revised version	[30–32]
PROLOGP (CDR)	Rekker	[27–32, 58]
SANALOGP	Extended Rekker	[59]
KLOGP (CASE)	Computer-identified fragments	[60]
CHEMICALC	Group/atomic values	[61]
LOGKOW	Atom/fragment contributions	[62]

[a] Not available as computer program

Thus, one can now subdivide log *P* calculations into three subtypes: "ALOGP" (atomic contributions); "BLOGP" (molecular properties); and "CLOGP" (fragmental contributions) [63]. The methods of Hansch-Fujita, Leo and Rekker described above are thus "CLOGP" approaches. We refer to the chapters of Tute and Leo in this volume for a historical perspective and discussion on a number of log *P* calculation methods. Below, we shortly review a selection of the elder and newer methods.

23.3.2 ALOGP Methods

23.3.2.1 Calculation Method According to Ghose-Crippen

The principles of the Ghose/Crippen approach [39–42] as applied in log *P* calculations differ largely from the fragmental treatments. In essence, Ghose and Crippen apply fragmentation as well. Functional groups, however, always left intact in the fragmental procedures, are now separated into atomic units. The consequence is that the application of the method must be accompanied by a rather intricate classification procedure; the total number of C-types runs as high as 44. The Ghose-Crippen system was revised three times [40–43], the number of atom classifications increasing from 90 to 120. Calculations are performed according to:

$$\log P_{GC} = \sum n_i a_i \qquad (3)$$

where n_i is the number of atoms of type i and a_i is the contribution of atom type i.

Computerized versions based on the Ghose-Crippen method are implemented in the MOLCAD and TSAR software [43].

23.3.2.2 The HINT Approach of Abraham

The HINT program works with so-called hydrophobic atomic constants (a_i), which are derived from the fragmental constants of Hansch and Leo [44]. For partitioning between atoms within a fragment it is assumed that exposed peripheral atoms of a fragment have a more significant role in hydrophobic interactions than shielded interior atoms. Thus, they are generally maintained at near-atomic values within fragments while the interior atom hydrophobic constants are adjusted to reflect the cumulative bonding effects on hydrophobicity within the fragment. In addition, all Leo system fragments are applied to the involved atoms and divided equally between the atoms in the case of bond-type factors. Polar proximity factors are applied to the central atom of the fragments as serial multiplicative factors tailored to the distance between the fragments and their nature.

23.3.2.3 The SMILOGP Approach of Dubost

The most recent contribution to ALOGP procedures stems from Dubost and coworkers. In 1994 they presented a software that generates an extended connectivity matrix from the SMILES code of a given molecule [45]. This extended connectivity matrix allows the determination of the atomic code for an atomic fragment and then the attribution of its lipophilicity contribution (f_C); log P is calculated by the summation of (f_C):

$$\log P = \sum_{c=1}^{c=n} f_c \qquad (4)$$

23.3.3 BLOGP Methods

Moriguchi and colleagues devised a simple method for calculating octanol/water partition coefficients using 13 structural parameters, which were found by multiple regression analysis of 1230 compounds [47, 48]. Some of the parameters, like the total number of N and O atoms, are numerical, while others, like the presence of ring structures, are indicator (or dummy with value 1 for presence or 0 for absence) variables. The overall correlation to experimental data for the 1230 compounds is excellent $(r = 0.952)$.

Using the semi-empirical AM1 method for the generation of various molecular descriptors, a multiparameter equation to calculate log P values was derived by Bodor and coworkers [49, 50]. Their test set includes 302 compounds and correlates well to experimental values $(r = 0.978)$.

The solvatochromic approach of Taft, Kamlet, Abraham (see Chapter 18) has been applied to calculate log P values based on four basic contributions, namely cavity size, dipolarity, and hydrogen bonding acceptor/donor ability [52].

While these latter parameters are difficult to assess, others explored e.g., graph theoretical descriptors in log P calculations [52]. However, their results were not satisfactory.

23.3.3.1 Conformation-Dependent log P Calculations

In the molecular orbital calculations by Bodor et al., the conformation of a molecule is implicitly part of the calculation. Various other approaches have been proposed in order to take the conformational freedom into account in log P calculations. A first approach is the HINT method developed by Abraham [44]. Another example is the method based on molecular lipophilicity potentials, discussed in detail in Chapters 12 and 13 [54]. Alternatively, conformation-dependent log P values may be obtained by approximate surface calculations (ASCLOGP) [54, 55] or by the MHM approach based on free energy surface densities [56].

23.3.4 CLOGP Methods

23.3.4.1 The Σf System of Rekker

The first hydrophobic fragmental system was developed by Rekker and his collaborators [27–32]. A data set of more than 1000 experimentally derived log P_{oct} values of simple organic compounds was used to derive a list of about 160 fragmental values by means of regression analysis. Correspondingly this approach was labeled "reductionistic". The calculation procedure is given by:

$$\log P = \Sigma f = \sum_{i=1}^{n} a_i f_i + \sum_{i=1}^{m} k_i C_M \tag{5}$$

where f denotes the hydrophobic fragmental constant, the lipophilicity contribution of a constituent part of a structure to the total lipophilicity, and a is a numerical factor indicating the incidence of a given fragment in that structure. In the second right-hand term, C_M denotes a correction factor (the "magic constant" = 0.289) and k_i gives the frequency of this constant in the structure under consideration. The constant proves to be of high importance in restoring imbalancies between experimentally derived log P values and calculations done by a mere addition of fragmental values.

A revised version of the Rekker system became recently available [30, 31]. It covers an increased number of fragmental values and gives a better-founded concept for the constant C_M; its value is now fixed to 0.219. Precise rules for application of the correction factor C_M are given in [32].

Computerized versions based on the Rekker approach are the PROLOGP and SANALOGP program.

23.3.4.2 CLOGP system of Hansch and Leo

In 1975 Leo et al. [33] published a fragmental system, which is based on the principles of "constructionism". This approach started with some basic fragmental values, ob-

tained by experimental measurement of a small set of the simplest possible molecules and then performed the construction of the fragment set by duly applying numerous correction factors in order to maintain the desired adaptation of new material in the system. Also the Leo method necessitates the usage of correction factors [33–37]:

$$\log P = \text{CLOGP} = \sum a_n f_n + \sum b_m F_m \tag{6}$$

where f is the fragmental constant; a is the incidence of fragments, F is a correction factor, and b is the frequency of correction factors.

The reliability of calculated log P values can be appreciated from Eq. (7) [37, 64]:

$$\log P^* = (0.911 \pm 0.004) \text{ CLOGP} + (0.191 \pm 0.011)$$
$$n = 7500, r = 0.978, s = 0.336 \tag{7}$$

where log P^* are reliable experimental log P values. Thus, for the great majority the calculated value is within ± 0.5 log units. However, also errors of up to ca. ± 1.5 units may occur.

23.3.4.3 Calculation Method According to Suzuki and Kudo

In 1990 Suzuki and Kudo [61] published their variant for log P calculation. It uses both atomic, as well as fragmental contributions. A group-contribution model without usage of correction terms is proposed:

$$\log P_{\text{SK}} = \sum_i^N n_i G_i \tag{8}$$

where N is the total number of groups, n_i is the number of the i-th group in the molecule, and G_i is the group contribution of the i-th group to log P. Groups may influence each other and this fact necessitates the application of a refined group classification procedure. A simple group like CH_2 shows up in 51 different types.

A similar method was recently published based on a set of atom and fragment contributions (AFC) [62]. However, this AFC method also requires a set of corrections factors.

23.3.4.4 The CASE KLOGP Method

The Computer Automated Structure Evaluation (CASE) program is able to identify the most important fragments, or sometimes single atoms, required for a good log P estimation model. A test set of 1663 compounds was highly correlated to experimental values ($r^2 = 0.928$) [60].

23.4 Comparison of log *P* Calculation Methods

A number of comparisons between log P calculation methods have been reported [65, 66]. A recent study of the reliability of log P values was performed on a small set of 22 diverse drugs [67, 68]. Methods included were those of Rekker, Leo and Hansch, Su-

zuki and Kudo, and Moriguchi et al. [48]. The Moriguchi method correlates best to experimental log *P* values ($r = 0.901$), while the Suzuki-Kudo approach did not perform well ($r = 0.710$). Although such comparative studies give some indication on the performance of the selected programs, the choice of the test set may have considerable bias on the results. Therefore we have used a larger set, both more compounds and more programs, in the benchmark reported below.

23.4.1 A Benchmark of Simple Organic Compounds and Drugs

Experimental data used for this comparison are octanol/water partition coefficients taken from the literature [69–72]. Two test sets of molecules have been selected. The first set comprises 114 simple organic molecules including, e.g., phenols, benzoic acids, benzophenones, and imidazoles. The second set consists of 70 drug molecules which vary in size and flexibility and contain various and multiple functional groups, since they belong to five pharmacological classes, i.e., class I- and class III-antiarrhythmics, β-blockers, potassium channel openers and neuroleptics, the latter including phenothiazines and benzamides. Although we realize that such a test is never perfect and fully representative, it illustrates some of the differences between the log *P* calculation procedures (data table and a poster copy with the main results can be obtained from H.v.d.W. by email: johannes.van_de_waterbeemd@roche.com).

Our benchmark included most of the currently available methods (see Table 2).

23.4.1.1 General Remarks

The difference in computed log *P* values may be quite large, e.g., for chlorpromazine the calculated log *P* varies from 3.82 (MOLCAD) to 5.31 (ASCLOGP), while its experimental value is 5.19–5.34. The experimental log *P* values of regioisomers are not identical. Some of the programs do consider this, while others predict identical values and require thus some more refinement.

Another point is that not all programs are able to compute each compound, e.g., when certain fragments are not parametrized. This affects also the comparison below, since the number of compounds for each pairwise correlation is varying. It should be noted that in some cases one or a few badly calculated compounds may affect the result. Nevertheless, we report here the main trends. A full paper will be published elsewhere.

23.4.1.2 The Full Data Set

The best predictions were obtained as follows in decreasing order: Rekker $\sum f$ and KLOGP ($r = 0.96$), CLOGP ($r = 0.95$), SANALOGP and PROLOGP (0.94), SMILOGP and TSAR (0.91), MOLCAD (0.90), CHEMICALC-2 and HINT (0.89), BLOGP (0.87), ASCLOGP (0.86). This overview demonstrates that the fragment-based methods score better than the atom-based approaches.

23.4.1.3 Subsets: Drugs and Simple Organic Compounds

In order to see whether the structural complexity of the compounds affects the log P calculations, the data set was divided into two parts. First taking the 70 drugs apart, gives the following results for the predictions: KLOGP and Rekker $\sum f$ ($r = 0.94$), SANALOGP and CLOGP ($r = 0.93$), SMILOGP and ATOMIC5 ($r = 0.91$), TSAR (0.89), MOLCAD (0.87), CHEMICALC-2 ($r = 0.87$), HINT ($r = 0.86$), ASCLOGP ($r = 0.84$), BLOGP (0.79). Looking only at the 114 chemicals the order is: CLOGP, KLOGP and Rekker $\sum f$ ($r = 0.99$), PROLOGP, TSAR, CHEMICALC-2 and SANALOGP ($r = 0.98$), MOLCAD and BLOGP (0.96), HINT and SMILOGP ($r = 0.93$), ASCLOGP ($r = 0.91$). Again, in both cases the fragment methods give the best results. As expected, simple compounds are easier to compute than more complex ones. The conformation-dependent ASCLOGP method suffers from the problem that it is not clear which conformation is most relevant.

23.5 Databases

23.5.1 Log P Databases

Large compilations of experimental log P values can be found in the collections made by Leo and coworkers [34, 73]. The best-known database is the Pomona or MedChem database [72]. Presently, it contains log P values for over 25 000 compounds. A further alternative is LOGKOW, which contains a collection of log P values for more than 13 500 compounds [74].

Table 3. Lipophilicity contribution of a set of widely diverse substituents

Substituent	π_{ar}	π_{al}	f_{ar}	f_{al}	l_1	l_2
H	0.00	0.00	0.182	0.182	0.00	0.00
Br	0.86	0.60	1.116	0.249	1.16	0.78
OH	− 0.67	− 1.12	− 0.314	−1.470	− 2.08	0.64
CN	− 0.57	− 0.84	− 0.174	− 1.041	− 1.62	0.50
COCH$_3$	− 0.55	− 0.62	− 0.075	− 0.942	− 1.34	0.46
CH$_3$	0.56	0.50	0.701	0.701	1.13	0.06
OC$_3$H$_7$	1.05	0.45	1.300	0.144	1.27	0.93
SO$_2$CH$_3$	− 1.63	− 1.50	− 1.169	− 1.430	− 3.16	− 0.05
N(CH$_3$)$_2$	0.18	− 0.30	0.473	− 0.683	− 0.38	0.78
C(CH$_3$)$_3$	1.98	1.17	2.258	2.258	3.85	0.18
C$_6$H$_5$	1.96	2.15	1.840	1.840	3.63	0.55
COC$_6$H$_5$	1.05	0.36	1.064	0.197	0.94	0.84

π, Hansch-Fujita substituent constant [26].
f, Rekker fragment constant [29, 75].
al, substituent to an aliphatic chain.
ar, substituent to an aromatic (benzene) ring.
l, disjoint principal properties (DPPs) for an optimal diverse set of substituents derived by principal component analysis [76].

Table 4. Fragmental constants for aliphatic and aromatic substitution (ordered by increasing lipophilicity)

Substituent	f_{al}	f_{ar}	π_{ar}
$N(CH_3)_3{}^+$			-5.96
COO^-	-5.19		-4.36
$NH_3{}^+$			-4.19
O^-			-3.87
SO_2NH_2	-2.42	-1.59	-1.82
$SOCH_3$	-2.24		-1.58
$CON(CH_3)_2$			-1.51
$CONH_2$	-2.18	-1.26	-1.49
OCH_2CONH_2			-1.37
NH_2	-1.54	-1.00	-1.23
$OCONH_2$	-1.58	-0.82	-1.05
1-Tetrazolyl			-1.04
CH_2OH	-1.10	-0.98	-1.03
OSO_2CH_3	-1.34		-0.88
OCH_2COOH	-1.16	-0.46	-0.79
CH_2OCH_3			-0.78
$SO_2N(CH_3)_2$			-0.78
CH_2CH_2OH			-0.77
$CH_2CH_2NH_2$			-0.76
CH_2COOH	-0.45		-0.72
CH_2COCH_3			-0.69
OH	-1.64	-0.44	-0.67
CHO	-1.10	-0.42	-0.65
$OCOCH_3$	-0.72	-0.36	-0.64
CN	-1.27	-0.34	-0.57
$B(OH)_2$			-0.55
$COCH_3$	-1.13	-0.20	-0.55
2-Pyrrolidinyl			-0.48
$NHCH_3$	-1.38	-0.14	-0.47
COOH	-1.11	-0.03	-0.32
CH_2COOCH_3			-0.30
NO_2	-1.16	-0.03	-0.28
2-Imidazolyl			-0.25
OCH_3	-1.54	0.28	-0.02
H	0.23	0.23	0.00
NHEt	-0.84		0.08
F	-0.38	0.37	0.14
$N(CH_3)_2$	-0.64	0.53	0.18
$SO_2C_6H_5$			0.27
2-Piperidinyl			0.32
OCH_2CH_3	-0.39	0.82	0.38
SH	-0.23	0.62	0.39
SCN	-0.48	0.64	0.41
$SO_2NHC_6H_5$			0.45
2-Pyridyl			0.50

Table 4 (continued)

Substituent	f_{al}	f_{ar}	π_{ar}
COOEt	− 0.59	0.87	0.51
SO$_2$CF$_3$	0.40	0.90	0.55
CH$_3$	0.77	0.89	0.56
SCH$_3$	− 0.02	0.92	0.61
Cl	0.06	0.94	0.71
Br	0.20	1.09	0.86
CF$_3$	0.29	1.11	0.88
OSO$_2$C$_6$H$_5$			0.93
CH$_2$CH$_3$	1.43	1.55	1.02
OCF$_3$	− 1.10	0.30	1.04
I	0.59	1.35	1.12
c-Propyl	1.49	1.57	1.14
N(Et)$_2$	0.16	1.57	1.18
NHC$_6$H$_5$	0.75	0.87	1.37
SCF$_3$	0.16	0.93	1.44
COOC$_6$H$_5$	0.60		1.46
C$_3$H$_7$	1.97	2.09	1.55
C$_6$H$_5$	1.90	1.90	1.96
C(CH$_3$)$_3$	2.22	2.61	1.98
OC$_6$H$_5$	1.22	1.29	2.08
C$_4$H$_9$	2.61	2.63	2.13
c-Pentyl			2.14
1-Adamantyl			3.37

Data taken from [34, 75, 77]
al, substituent to an aliphatic chain
ar, substituent to an aromatic (benzene) ring.

23.5.2 Substituent Values for Aliphatic and Aromatic Substituents

The design of new compounds involves the structural variation of parts of a lead compound. Typically, certain positions in a molecule will be varied to a certain extent. A rational choice with respect to the physico-chemical properties of the compounds can be made using tabulated descriptor values for candidate substituents. Large compilations of such substituent descriptors are available [e.g., 73, 75]. Of particular interest are contributions of a substituent to the lipophilicity of the compound. It is important to notice the difference between aromatic ring and aliphatic side chain substitution. The lipophilicity contribution to an aromatic ring is always larger than to an aliphatic chain or ring (see Tables 3 and 4).

23.5.3 Lipophilicity Scales for Amino Acids

The characterization and classification of the 20 natural amino acids is subject of many publications. A manifold of lipophilicity scales can be found [e.g., 78–81] (see Table 5 and Chapter 20). Many synthetic peptides contain unusual building blocks. Informa-

Table 5.   Lipophilicity contribution of the side chains of 20 coded amino acids

		CLOGP[a] v. 3.55	Levitt (1971)	Fauchère Pliška (1983) [78]	Parker et al. (1986)	Radzicka Wolfenden (1988)	Cowan, Whit- taker (1990)	Chmelik et al. (1991)	El Tayar et al. (1992)
1.	Ala	− 2.999	0.5	0.31	− 2.1	1.81	0.35	0.30	− 2.77
2.	Arg	− 7.938	− 3.0	− 1.01	− 4.2	− 14.92	− 1.50	− 0.91	− 3.79
3.	Asn	− 4.563	− 0.2	− 0.60	− 7.0	− 6.64	− 0.99	− 0.48	− 3.48
4.	Asp	− 3.823	− 2.5	− 0.77	− 10.0	− 8.72	− 2.15	− 0.55	− 3.61
5.	Cys	− 3.172	1.0	1.54	− 1.4	1.28	0.76	0.86	− 2.55
6.	Gln	− 4.927	− 0.2	− 0.22	− 6.0	− 5.54	− 0.93	− 0.30	− 3.11
7.	Glu	− 3.927	− 2.5	− 0.64	− 7.8	− 6.81	− 1.95	− 0.32	− 3.51
8.	Gly	− 3.308	0.0	0.0	− 5.7	0.94	0.0	0.0	− 3.00
9.	His	− 4.000	0.5	0.13	− 2.1	− 4.66	− 0.65	0.02	− 2.85
10.	Ile	− 1.542	1.8	1.80	8.0	4.92	1.83	1.53	− 1.80
11.	Leu	− 1.542	1.8	1.70	9.2	4.92	1.80	1.50	− 1.72
12.	Lys	− 3.299	− 3.0	− 0.99	− 5.7	− 5.55	− 1.54	− 0.74	− 3.77
13.	Met	− 2.851	1.3	1.23	4.2	2.35	1.10	1.14	− 2.10
14.	Phe	− 1.581	2.5	1.79	9.2	2.98	1.69	1.91	− 1.44
15.	Pro	− 2.655	1.4	0.72	− 2.1	− 0.52[b]	0.84	0.53	− 2.62
16.	Ser	− 4.215	− 0.3	− 0.04	− 6.5	− 3.40	− 0.63	− 0.19	− 3.00
17.	Thr	− 3.906	0.4	0.26	− 5.2	− 2.57	− 0.27	0.09	− 2.83
18.	Trp	− 1.581	3.4	2.25	10.0	2.33	1.35	2.01	− 1.15
19.	Tyr	− 2.248	2.3	0.96	1.9	− 0.14	0.39	0.85	− 2.11
20.	Val	− 2.071	1.5	1.22	3.7	4.04	1.32	1.07	− 2.29

[a] CLOGP: calculated 1-octanol/water partition coefficients for the amino acid, included for comparison.
[b] Estimated value [78].

tion on the properties of these fragments is only recently appearing. Lipophilic characteristics of a number of selected noncoded amino acids can be found in Chapter 21.

23.6 Perspectives

The recent literature demonstrates that lively interest exists in the calculation of reliable log *P* values. The success of various methods is difficult to judge unequivocally. Certain classes of compounds with new functional groups will remain difficult to predict; thus, there is still a need for reliable experimental approaches.

The present benchmark of log *P* calculations indicates that the predictive value of a calculation procedure for molecular lipophilicity is significantly better for not-too-complex organic molecules than for structurally more complex bioactive compounds. The calculation procedures compared here can be arranged into three groups with significantly differing predictive power: fragmental-based > atom-based > property-based. This benchmark study reveals that the fragmental method is at present the best

approach. We think that all tested calculation procedures have their own limitations; for the future development we would advise a thorough reconsideration of structural effects not fully (or even not at all) incorporated in the parametrizations. Particularly, attention will have to be paid to conformational aspects of partitioning processes.

A simple approach to calculate log P values is based on the factorization of log P according to:

$$\log P = \text{size} - \text{hydrogen bonding} \tag{9}$$

Size descriptors, like molar volume or molecular weight, can be obtained by simple computation. Various hydrogen bonding scales have been proposed in the literature and should be further developed. Thus, using size and appropriate hydrogen bonding scales, log P values can be easily estimated. Two posters on a recent symposium demonstrated applications of this approach [64, 82]. Applications of this approach will be reported [83, 84].

Acknowledgements

References

[1] Leo, A., Hansch, C., and Elkins, D., *Chem. Rev.* **71**, 525–616 (1971)

[2] Smith, R. N., Hansch, C., and Ames, M. M., *J. Pharm. Sci.* **64**, 599–605 (1975)

[3] Dearden, J., *Environ. Health Perspect.* **61**, 203–228 (1985)

[4] Kubinyi, H., *Prog. Drug Res.* **23**, 97–198 (1979)

[5] Leahy, D. E., Taylor, P. J., and Wait, A. R., *Quant. Struct.-Act. Relat.* **8**, 17–31 (1989)

[6] Taylor, P. J., Hydrophobic Properties of Drugs. In: *Comprehensive Medicinal Chemistry*, Vol. **4**. Hansch, C. (Ed.). Pergamon Press: Oxford; 241–294 (1990)

[7] Leahy, D. E., Morris, J. J., Taylor, P. J., and Wait, A. R., Membranes and their models – towards a rational choice of partitioning system. In: *QSAR: Rational Approaches to the Design of Bioactive Compounds*, Silipo, C. and Vittoria, A (Eds.). Elsevier: Amsterdam; 75–82 (1991)

[8] Young, R. C., Mitchell, R. C., Brown, Th. 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)

[9] El Tayar, N., Tsai, R.-S., Testa, B., Carrupt, P.-A., Hansch, C., and Leo, A., *J. Pharm. Sci.* **80**, 744–749 (1991)

[10] Kim, D.-C., Burton, P. S., and Borchardt, R. T., *Pharmaceut Res.* **10**, 1710–1714 (1993)

[11] Ter Laak, A. M., Tsai, R. S., Donné-Op den Kelder, G. M., Carrupt, P. A., Testa, B., and Timmerman, H., *Eur. J. Pharm. Sci.* **2**, 373–384 (1994)

[12] Van de Waterbeemd, J. Th. M., Van Boekel, C. A. A., De Sévaux, R. L. F. M., Jansen, A. C. A., and Gerritsma, K. W., *Pharm. Weekbl. Sci. Ed.* **3**, 12–25 (1981)

[13] Paterson, D. A., Conradi, R. A., Hilgers, A. R., Vidmar, T. J., and Burton, P. S., *Quant. Struct.-Act. Relat.* **13**, 4–10 (1994)

[14] Smejtek, P., and Wang, S., *Arch. Environ. Contam. Toxicol.* **25**, 394–404 (1993)

[15] Zaslavsky, B. Y., and Masimov, E. A., *Top. Curr. Chem.* **146**, 171–202 (1988)

[16] Zaslavsky, B. Y., *Aqueous Two-Phase Partitioning. Physical Chemistry and Bioanalytical Applications.* Marcel Dekker: New York, 1995

[17] Bate-Smith, E. C., and Westall, R. G., *Biochim. Biophys. Acta* **4**, 427–440 (1950)

[18] Kubinyi, H., The quantitative analysis of structure-activity relationships. In: *Burger's Medicinal Chemistry and Drug Discovery,* 5th ed. Wolff, M. E. (Ed.). John Wiley: New York; 497–571 (1995)

[19] Dearden, J. C., and Bresnen, G. M., *Quant Struct.-Act. Relat.* **7**, 133–144 (1988)

[20] Hersey, A., Hill, A. P., Hyde, R. M., and Livingstone, D. J., *Quant. Struct.-Act. Relat.* **8**, 288–296 (1989)

[21] De Bruijn, J., Busser, F., Seinen, W., and Hermens, J., *Method. Environ. Tox. Chem.* **8**, 499–512 (1989)

[22] Tomlinson, E., *J. Pharm. Sci.* **71**, 602–604 (1982)

[23] Kuban, V., *Anal. Chim. Acta* **248**, 493–499 (1991)

[24] Ford, H. Jr., Merski, Chr. L., and Kelley, J. A., *J. Liq. Chromatogr.* **14**, 3365–3386 (1991)

[25] Avdeef, A., *Quant. Struct.-Act. Relat.* **11**, 510–517 (1992)

[26] Fujita, T., Iwasa, J., and Hansch, C., *J. Am. Chem. Soc.* **86**, 5175–5180 (1964)

[27] Nys, G. G., and Rekker, R. F., *Eur. J. Med. Chem.* **8**, 521–535 (1973)

[28] Rekker, R. F., and De Kort, H. M., *Eur. J. Med. Chem.* **14**, 479–488 (1979)

[29] Rekker, R. F., *The hydrophobic Fragmental Constant,* Pharmacochemistry Library, Vol. **1**. Elsevier: Amsterdam, 1977

[30] Rekker, R. F., and Mannhold, R., *Calculation of Drug Lipophilicity.* VCH; Weinheim, 1992

[31] Rekker, R. F., and Bijloo, G. J., in preparation (1995)

[32] Mannhold, R., Rekker, R. F., Sonntag, Ch., Ter Laak, A. M., Dross, K. and Polymeropoulos, E. E., *J. Pharm. Sci.,* in press

[33] Leo, A., Jow, P. Y. C., Silipo, C., and Hansch, C., *J. Med. Chem.* **18**, 865–868 (1975)

[34] Hansch, C., and Leo, A. J., *Substituent Constants for Correlation Analysis in Chemistry and Biology.* Wiley: New York, 1979

[35] Leo, A. J., *J. Pharm. Sci.* **76**, 166–168 (1987)

[36] Leo, A. J., *Methods Enzymol.* **202**, 544–591 (1991)

[37] Leo, A. J., *Chem. Rev.* **93**, 1281–1306 (1993)

[38] Broto, P., Moreau, G., and Vandyke, C., *Eur. J. Med. Chem.* **19**, 71–78 (1984)

[39] Ghose, A. K., and Crippen, G. M., *J. Med. Chem.* **28**, 333–346 (1985)

[40] Ghose, A. K., and Crippen, G. M., *J. Chem. Inf. Comput. Sci.* **1**, 21–35 (1987)

[41] Ghose, A. K., and Crippen, G. M., *J. Comp. Chem.* **7**, 565–577 (1986)

[42] Ghose, A. K., Pritchett, A., and Crippen, G. M., *J. Comput. Chem.* **9**, 80–90 (1988)

[43] Program based on the Ghose-Crippen method

[44] Kellogg, G. E., and Abraham, D. J., *J. Comput.-Aided Mol. Des.* **5**, 545–552 (1991)

[45] Convard, T., Dubost, J.-P., Le Solleu, H., and Kummer, E., *Quant. Struct.-Act. Relat.* **13**, 34–37 (1994)

[46] Devillers, J., Domine, D., and Karcher, W., 10th European Symposium on Structure-Activity Relationships (1994), Abstract A114

[47] Moriguchi, I., Hirono, S., Nakagome, I., and Hirano, H., *Chem. Pharm. Bull.* **42**, 976–978 (1984)

[48] Moriguchi, I., Hirono, S., Liu, Q., Nakagome, I., and Matsushita, Y., *Chem. Pharm. Bull.* **40**, 127–130 (1992)

[49] Bodor, N., and Huang, M. J., *J. Pharm. Sci.* **81**, 272–281 (1992)

[50] Bodor, N., Gabanyi, Z., and Wong, C. K., *J. Am. Chem. Soc.* **111**, 3783–3796 (1989)

[51] Leahy, D. E., *J. Pharm. Sci.* **75**, 629–636 (1986)

[52] Niemi, G. J., Basak, S. C., Veith, G. D., and Grunwald, G., *Environ. Tox. Chem.* **11**, 893–900 (1992)

[53] Gaillard, P., Carrupt, P.-A., Testa, B., and Boudon, A., *J. Comput.-Aided Mol. Des.* **8**, 83–96 (1994)

[54] Ulmschneider, M., *Analytical Model for the Calculation of van der Waals and Solvent Accessible Surface Areas. Contribution to the Calculation of Free Enthalpies of Hydration and Octanol/Water Partition Coefficients.* Ph. D. Thesis, University of Haute-Alsace, Mulhouse, France 1993

[55] Van de Waterbeemd, H., Karajiannis, H., Kansy, M., Obrecht, D., Müller, K., and Lehmann, Chr., Conformation-lipophilicity relationships of peptides and peptide mimetics. In: *Trends in QSAR and Molecular Modeling 94.* Sanz, F. (ed.). Prous: Barcelona (1995) in press

[56] Pixner, P., Heiden, W., Merx, H., Moeckel, G., Möller, A., and Brickmann, J., *J. Chem. Inf. Comput. Sci.* **34**, 1309–1319 (1994)

[57] Klopman, G., and Iroff, L. D., *J. Comput. Chem.* **2**, 157–160 (1981)

[58] Darvas, F., Erdoz, I., and Teglas, G., Influence of the fragmental systems on calculated log P values: a theoretical investigation using a logic-based expert system. In: *QSAR in Drug Design and Toxicology.* Hadzi, E. and Jerman-Blazic, B. (Eds.). Elsevier: Amsterdam 70–73 (1987)

[59] Petelin, D. E., Arslanov, N. A., Paulyulin, V. A. and Zefirov, N. S., 10th European Symposium on Structure-Activity Relationships (1994), Abstract B263

[60] Klopman, G., Li, J.-Y., Wang, S., and Dimayuga, M., *J. Chem. Inf. Comput. Sci.* **34**, 752–781 (1994)

[61] Suzuki, T., and Kudo, Y., *J. Comput.-Aid. Mol. Design* **4**, 155–198 (1990)

[62] Meylan, W. M., and Howard, P. H., *J. Pharm. Sci.* **84**, 83–92 (1995)

[63] Viswanadhan, V. N., Reddy, M. R., Bacquet, R. J., and Erion, M. D., *J. Comput. Chem* **14**, 1019–1026 (1993)

[64] Camenisch, G., and Van de Waterbeemd, H., *The prediction of transport properties using molecular descriptors,* Symposium on Lipophilicity in Drug Research and Toxicology, Lausanne, 21–24 March (1995) p. P-02

[65] Mayer, J. M., Van de Waterbeemd, H., and Testa, B., *Eur. J. Med. Chem.* **17**, 17–25 (1982)

[66] Viswanadhan, V. N., Ghose, A. K., Revankar, G. R., and Robins, R. K., *J. Chem. Inf. Comput. Sci.* **29**, 163–172 (1989)

[67] Rekker, R. F., Mannhold, R., and Ter Laak, A. M., *Quant. Struct.-Act. Relat.* **12**, 152–157 (1993)

[68] Moriguchi, I., Hirono, S., Nakagome, I., and Hirano, H., *Chem. Pharm. Bull.* **42**, 976–978 (1994)

[69] Mannhold, R., Dross, K. P., and Rekker, R. F., *Quant. Struct.-Act. Relat.* **9**, 21–28 (1990)

[70] Dross, K. P., Mannhold, R., and Rekker, R. F., *Quant. Struct.-Act. Relat.* **11**, 36–44 (1992)

[71] Dross, K., Sonntag, Ch., and Mannhold, R., *J. Chromatogr.* **673**, 113–124 (1994)

[72] Pomona or MedChem log P database. Available from BioByte or Daylight CIS

[73] Hansch, C., Leo, A., and Hoekman, D., *Exploring QSAR. Hydrophobic, Electronic, and Steric Constants.* ACS: Washington, 1995

[74] LOGKOW log P database. Available from Sangster Research Laboratories. Montreal, Canada

[75] Van de Waterbeemd, H., and Testa, B., *Adv. Drug Res.* **16**, 85–225 (1987)

[76] Van de Waterbeemd, H., Costantino, G., Clementi, S., Cruciani, G., and Valigi, R., Disjoint principal properties of organic substituents. In: *Chemometric Methods in Molecular Design,* Van de Waterbeemd, H. (Ed.). VCH: Weinheim 103–112 (1995)

[77] Hansch, C., and Leo, A., *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology.* ACS: Washington, 1995

[78] Fauchère, J. L., and Pliška, V., *Eur. J. Med. Chem.* **18,** 369–375 (1983)

[79] Van de Waterbeemd, H., Karajiannis, H., and El Tayar, N., *Amino Acids* **7,** 129–145 (1994)

[80] Naray-Szabo, G., and Balogh, T., *J. Mol. Struct. (Theochem)* **264,** 243–248 (1993)

[81] Wilce, M. C. J., Aguilar, M.-I., and Hearn, M. T. W., *Anal. Chem.* **67,** 1210–1219 (1995)

[82] Raevsky, O. A., and Chrétien, J. R, *Hydrogen bond contribution in lipophilicity: example of amines and pyridines.* Symposium on Lipophilicity in Drug Research and Toxicology, Lausanne, 21–24 March (1995) p. P-45

[83] Raevsky, O. A., Schaper, K.-J., and Seydel, J. K., *Quant. Struct.-Act. Relat.,* **14,** 433–436 (1995)

[84] Camenisch, G., Van de Waterbeemd, H., and Folkers, G., in preparation

Index