CHAPTER ONE

History of Quantitative Structure-Activity Relationships

C. D. SELASSIE
Chemistry Department
Pomona College
Claremont, California

Contents

1 Introduction, 2
1.1 Historical Development of QSAR, 3
1.2 Development of Receptor Theory, 4
2 Tools and Techniques of QSAR, 7
2.1 Biological Parameters, 7
2.2 Statistical Methods: Linear Regression Analysis, 8
2.3 Compound Selection, 11
3 Parameters Used in QSAR, 11
3.1 Electronic Parameters, 11
3.2 Hydrophobicity Parameters, 15
3.2.1 Determination of Hydrophobicity by Chromatography, 17
3.2.2 Calculation Methods, 18
3.3 Steric Parameters, 23
3.4 Other Variables and Variable Selection, 25
3.5 Molecular Structure Descriptors, 26
4 Quantitative Models, 26
4.1 Linear Models, 26
4.1.1 Penetration of ROH into Phosphatidylcholine Monolayers (1841, 27
4.1.2 Changes in EPR Signal of Labeled Ghost Membranes by ROH (185), 27
4.1.3 Induction of Narcosis in Rabbits by ROH (184), 27
4.1.4 Inhibition of Bacterial Luminescence by ROH (185), 27
4.1.5 Inhibition of Growth of Tetrahymena pyriformis by ROH (76, 1861, 27
4.2 Nonlinear Models, 28
4.2.1 Narcotic Action of ROH on Tadpoles, 28
4.2.2 Induction of Ataxia in Rats by ROH, 29
4.3 Free-Wilson Approach, 29
4.4 Other QSAR Approaches, 30
5 Applications of QSAR, 30
5.1 Isolated Receptor Interactions, 31

Burger's Medicinal Chemistry and Drug Discovery
Sixth Edition, Volume 1: Drug Discovery
Edited by Donald J. Abraham
ISBN 0-471-27090-3 © 2003 John Wiley & Sons, Inc.

5.1.1 Inhibition of Crude Pigeon Liver DHFR by Triazines (202), 31
5.1.2 Inhibition of Chicken Liver DHFR by 3-X-Triazines (207), 31
5.1.3 Inhibition of Human DHFR by 3-X-Triazines (208), 32
5.1.4 Inhibition of L1210 DHFR by 3-X-Triazines (209), 32
5.1.5 Inhibition of P. *carinii* DHFR by 3-X-Triazines (210), 32
5.1.6 Inhibition of L. major DHFR by 3-X-Triazines (211), 33
5.1.7 Inhibition of T. gondii DHFR by 3-X-Triazines, 33
5.1.8 Inhibition of Rat Liver DHFR by 2,4-Diamino, 5-Y, 6-Z-quinazolines (213), 34
5.1.9 Inhibition of Human Liver DHFR by 2,4-Diamino, 5-Y, 6-Z-quinazolines (214), 34
5.1.10 Inhibition of Murine L1210 DHFR by 2,4-Diamino, 5-Y, 6-Z-quinazolines (214), 34
5.1.11 Inhibition of Bovine Liver DHFR by 2,4-Diamino, 5-Y, 6-Z-quinazolines (215), 34
5.1.12 Binding of X-Phenyl, N-Benzoyl-L-alaninates to α-Chymotrypsin in Phosphate Buffer, pH 7.4 (203), 35
5.1.13 Binding of X-Phenyl, N-Benzoyl-L-alaninates to a-Chymotrypsin in Pentanol (203), 35
5.1.14 Binding of X-Phenyl, *N*-Benzoyl-L-alaninates in Aqueous Phosphate Buffer (218), 35
5.1.15 Binding of X-Phenyl, N-Benzoyl-L-alaninates in Pentanol (218), 35
5.1.16 Inhibition of 5-a-Reductase by 4-X, N-Y-6-azaandrost-17-CO-Z-4-ene-3-ones, I, 36
5.1.17 Inhibition of 5-a-Reductase by 17β-(*N*-(X-phenyl)carbamoyl)-6-azaandrost-4-ene-3-ones, II, 36
5.1.18 Inhibition of 5-a-Reductase by 17β-(*N*-(1-X-phenyl-cycloalkyl)carbamoyl)-6-azaandrost-4-ene-3-ones, III, 36
5.2 Interactions at the Cellular Level, 37
5.2.1 Inhibition of Growth of L1210/S by 3-X-Triazines (209), 37
5.2.2 Inhibition of Growth of L1210/R by 3-X-Triazines (209), 37
5.2.3 Inhibition of Growth of Tetrahymena pyriformis (40 h), 37
5.2.4 Inhibition of Growth of T. pyriformis by Phenols (using σ) (227), 38
5.2.5 Inhibition of Growth of T. pyriformis by Electron-Releasing Phenols (227), 38
5.2.6 Inhibition of Growth of T. pyriformis by Electron-Attracting Phenols (227), 38
5.2.7 Inhibition of Growth of T. pyriformis by Aromatic Compounds (229), 38
5.3 Interactions In *Vivo*, 38
5.3.1 Renal Clearance of β-Adrenoreceptor Antagonists, 38
5.3.2 Nonrenal Clearance of β-Adrenoreceptor Antagonists, 39
6 Comparative QSAR, 39
6.1 Database Development, 39
6.2 Database: Mining for Models, 39
6.2.1 Incidence of Tail Defects of Embryos (235), 40
6.2.2 Inhibition of DNA Synthesis in CHO Cells by X-Phenols (236), 40
6.2.3 Inhibition of Growth of L1210 by X-Phenols, 40
6.2.4 Inhibition of Growth of L1210 by Electron-Withdrawing Substituents ($\sigma^+ > 0$), 41
6.2.5 Inhibition of Growth of L1210 by Electron-Donating Substituents ($\sigma^+ < 0$), 41
6.3 Progress in QSAR, 41
7 Summary, 42

1 INTRODUCTION

It has been nearly 40 years since the quantitative structure-activity relationship (QSAR) paradigm first found its way into the practice of agrochemistry, pharmaceutical chemistry, toxicology, and eventually most facets of chemistry (1)Its stayingpower may be attributed to the strength of its initial postulate that activity was a function of structure as described by electronic attributes, hydrophobicity, and steric properties as well as the rapid and extensive development in methodologies and computational techniques that have ensued to delineate and refine the many variables and approaches that define the paradigm. The overall goals of QSAR retain their original essence and remain focused on the predictive ability of the approach and its receptiveness to mechanistic interpretation.

Rigorous analysis and fine-tuning of independent variables has led to an expansion in development of molecular and atom-based descriptors, as well as descriptors derived from quantum chemical calculations and spectroscopy (2). The improvement in high-throughput screening procedures allows for rapid screening of large numbers of compounds under similar test conditions and thus minimizes the risk of combining variable test data from many sources.

The formulation of thousands of equations using QSAR methodology attests to a validation of its concepts and its utility in the elucidation of the mechanism of action of drugs at the molecular level and a more complete understanding of physicochemical phenomena such as hydrophobicity. It is now possible not only to develop a model for a system but also to compare models from a biological database and to draw analogies with models from a physical organic database (3). This process is dubbed ***model mining*** and it provides a sophisticated approach to the study of chemical-biological interactions. QSAR has clearly matured, although it still has a way to go. The previous review by Kubinyi has relevant sections covering portions of this chapter as well as an extensive bibliography recommended for a more complete overview (4).

1.1 Historical Development of QSAR

More than a century ago, Crum-Brown and Fraser expressed the idea that the physiological action of a substance was a function of its chemical composition and constitution (5). A few decades later, in 1893, Richet showed that the cytotoxicities of a diverse set of simple organic molecules were inversely related to their corresponding water solubilities (6). At the turn of the 20th century, Meyer and Overton independently suggested that the narcotic (depressant) action of a group of organic compounds paralleled their olive oil/water partition coefficients (7, 8). In 1939 Ferguson introduced a thermodynamic generalization to the correlation of depressant action with the relative saturation of volatile compounds in the vehicle in which they were administered (9). The extensive work of Albert, and Bell and Roblin established the importance of ionization of bases and weak acids in bacteriostatic activity (10–12). Meanwhile on the physical organic front, great strides were being made in the delineation of substituent effects on organic reactions, led by the seminal work of Hammett, which gave rise to the "sigma-rho" culture (13, 14). Taft devised a way for separating polar, steric, and resonance effects and introducing the first steric parameter, E_S (15). The contributions of Hammett and Taft together laid the mechanistic basis for the development of the QSAR paradigm by Hansch and Fujita. In 1962 Hansch and Muir published their brilliant study on the structure-activity relationships of plant growth regulators and their dependency on Hammett constants and hydrophobicity (16). Using the octanol/water system, a whole series of partition coefficients were measured, and thus a new hydrophobic scale was introduced (17). The parameter **a,** which is the relative hydrophobicity of a substituent, was defined in a manner analogous to the definition of sigma (18).

$$\pi_X = \log P_X - \log P_H \qquad (1.1)$$

P_X and P_H represent the partition coefficients of a derivative and the parent molecule, respectively. Fujita and Hansch then combined these hydrophobic constants with Hammett's electronic constants to yield the linear Hansch equation and its many extended forms (19).

$$\text{Log } 1/C = a\sigma + b\pi + ck \qquad (1.2)$$

Hundreds of equations later, the failure of linear equations in cases with extended hydrophobicity ranges led to the development of the Hansch parabolic equation (20):

$$\text{Log } 1/C = a \cdot \log \text{P} - b(\log P)^2 + c\sigma + k \qquad (1.3)$$

The delineation of these models led to explosive development in QSAR analysis and related approaches. The Kubinyi bilinear model is a refinement of the parabolic model and, in many cases, it has proved to be superior (21).

$$\text{Log } 1/C = a \cdot \log P - b \cdot \log(\beta \cdot P + 1) + k \quad (1.4)$$

Besides the Hansch approach, other methodologies were also developed to tackle structure-activity questions. The Free-Wilson approach addresses structure-activity studies in a congeneric series as described in Equation 1.5 (22).

$$\text{BA} = \sum a_i x_i + u \quad (1.5)$$

BA is the biological activity, u is the average contribution of the parent molecule, and a_i is the contribution of each structural feature; x_i denotes the presence $X_i = 1$ or absence $X_i = 0$ of a particular structural fragment. Limitations in this approach led to the more sophisticated Fujita-Ban equation that used the logarithm of activity, which brought the activity parameter in line with other free energy-related terms (23).

$$\text{Log } BA = \sum G_i X_i + u \quad (1.6)$$

In Equation 1.6, u is defined as the calculated biological activity value of the unsubstituted parent compound of a particular series. G_i represents the biological activity contribution of the substituents, whereas X_i is ascribed with a value of one when the substituent is present or zero when it is absent. Variations on this activity-based approach have been extended by Klopman et al. (24) and Enslein et al. (25). Topological methods have also been used to address the relationships between molecular structure and physical/biological activity. The minimum topological difference (MTD) method of Simon and the extensive studies on molecular connectivity by Kier and Hall have contributed to the development of quantitative structure property/activity relationships (26, 27). Connectivity indices based on hydrogen-suppressed molecular structures are rich in information on branching, 3-atom fragments, the degree of substitution, proximity of substituents and length, and heteroatom of substituted rings. A method in its embryonic state of development uses both graph bond distances and Euclidean distances among atoms to calculate E-state values for each atom in a molecule that is sensitive to conformational structure. Recently, these electrotopological indices that encode significant structured information on the topological state of atoms and fragments as well as their valence electron content have been applied to biological and toxicity data (28). Other recent developments in QSAR include approaches such as HQSAR, Inverse QSAR, and Binary QSAR (29–32). Improved statistical tools such as partial least square (PLS) can handle situations where the number of variables overwhelms the number of molecules in a data set, which may have collinear X-variables (33).

1.2 Development of Receptor Theory

The central theme of molecular pharmacology, and the underlying basis of SAR studies, has focused on the elucidation of the structure and function of drug receptors. It is an endeavor that proceeds with unparalleled vigor, fueled by the developments in genomics. It is generally accepted that endogenous and exogenous chemicals interact with a binding site on a specific macromolecular receptor. This interaction, which is determined by intermolecular forces, may or may not elicit a pharmacological response depending on its eventual site of action.

The idea that drugs interacted with specific receptors began with Langley, who studied the mutually antagonistic action of the alkaloids, pilocorpine and atropine. He realized that both these chemicals interacted with some receptive substance in the nerve endings of the gland cells (34). Paul Ehrlich defined the receptor as the "binding group of the protoplasmic molecule to which a foreign newly introduced group binds" (35). In 1905 Langley's studies on the effects of curare on muscular contraction led to the first delineation of critical characteristics of a receptor: recognition capacity for certain ligands and an amplification component that results in a pharmacological response (36).

Receptors are mostly integral proteins embedded in the phospholipid bilayer of cell membranes. Rigorous treatment with detergents is needed to dissociate the proteins from the membrane, which often results in loss of

integrity and activity. Pure proteins such as enzymes also act as drug receptors. Their relative ease of isolation and amplification have made enzymes desirable targets in structure-based ligand design and QSAR studies. Nucleic acids comprise an important category of drug receptors. Nucleic acid receptors (aptamers), which interact with a diverse number of small organic molecules, have been isolated by *in vitro* selection techniques and studied (37). Recent binary complexes provide insight into the molecular recognition process in these biopolymers and also establish the importance of the architecture of tertiary motifs in nucleic acid folding (38). Groove-binding ligands such as lexitropsins hold promise as potential drugs and are thus suitable subjects for focused QSAR studies (39).

Over the last 20 years, extensive QSAR studies on ligand-receptor interactions have been carried out with most of them focusing on enzymes. Two recent developments have augmented QSAR studies and established an attractive approach to the elucidation of the mechanistic underpinnings of ligand-receptor interactions: the advent of molecular graphics and the ready availability of X-ray crystallography coordinates of various binary and ternary complexes of enzymes with diverse ligands and cofactors. Early studies with serine and thiol proteases (chymotrypsin, trypsin, and papain), alcohol dehydrogenase, and numerous dihydrofolate reductases (DHFR) not only established molecular modeling as a powerful tool, but also helped clarify the extent of the role of hydrophobicity in enzyme-ligand interactions (40–44). Empirical evidence indicated that the coefficients with the hydrophobic term could be related to the degree of desolvation of the ligand by critical amino acid residues in the binding site of an enzyme. Total desolvation, as characterized by binding in a deep crevice/pocket, resulted in coefficients of approximately 1.0 (0.9–1.1) (44). An extension of this agreement between the mathematical expression and structure as determined by X-ray crystallography led to the expectation that the binding of a set of substituents on the surface of an enzyme would yield a coefficient of about 0.5 (0.4–0.6) in the regression equation, indicative of partial desolvation.

Probing of various enzymes by different ligands also aided in dispelling the notion of Fischer's rigid lock-and-key concept, in which the ligand (key) fits precisely into a receptor (lock). Thus, a "negative" impression of the substrate was considered to exist on the enzyme surface (geometric complementarity). Unfortunately, this rigid model fails to account for the effects of allosteric ligands, and this encouraged the evolution of the induced-fit model. Thus, "deformable" lock-and-key models have gained acceptance on the basis of structural studies, especially NMR (45).

It is now possible to isolate membrane-bound receptors, although it is still a challenge to delineate their chemistry, given that separation from the membrane usually ensures loss of reactivity. Nevertheless, great advances have been made in this arena, and the three-dimensional structures of some membrane-bound proteins have recently been elucidated. To gain an appreciation for mechanisms of ligand-receptor interactions, it is necessary to consider the intermolecular forces at play. Considering the low concentration of drugs and receptors in the human body, the law of mass action cannot account for the ability of a minute amount of a drug to elicit a pronounced pharmacological effect. The driving force for such an interaction may be attributed to the low energy state of the drug-receptor complex: $K_D = [Drug][Receptor]/[Drug\text{-}ReceptorComplex]$. Thus, the biological activity of a drug is determined by its affinity for the receptor, which is measured by its K_D, the dissociation constant at equilibrium. A smaller K_D implies a large concentration of the drug-receptor complex and thus a greater affinity of the drug for the receptor. The latter property is promoted and stabilized by mostly noncovalent interactions sometimes augmented by a few covalent bonds. The spontaneous formation of a bond between atoms results in a decrease in free energy; that is, AG is negative. The change in free energy AG is related to the equilibrium constant K_{eq}.

$$\Delta G^\circ = -RT \ln K_{eq} \qquad (1.7)$$

Thus, small changes in ΔG° can have a profound effect on equilibrium constants.

Table 1.1 Types of Intermolecular Forces

Bond Type	Bond Strength (kcal/mol)	Example
1. Covalent	40-140	CH_3CH_2O-H
2. Ionic (Electrostatic)	5	$R_4\overset{+}{N}\cdots\overset{-}{O}-\overset{O}{\overset{\|\|}{C}}-$
3. Hydrogen	1–10	$-NH\cdots O(H)-H$
4. Dipole–dipole	1	$R_3N:\cdots C=O$
5. van der Waals	0.5–1	$C\cdots C$
6. Hydrophobic	1	

In the broadest sense, these "bonds" would include covalent, ionic, hydrogen, dipole-dipole, van der Waals, and hydrophobic interactions. Most drug-receptor interactions constitute a combination of the bond types listed in Table 1.1, most of which are reversible under physiological conditions.

Covalent bonds are not as important in drug-receptor binding as noncovalent interactions. Alkylating agents in chemotherapy tend to react and form an immonium ion, which then alkylates proteins, preventing their normal participation in cell divisions. Baker's concept of active site directed irreversible inhibitors was well established by covalent formation of Baker's antifolate and dihydrofolate reductase (46).

Ionic (electrostatic) interactions are formed between ions of opposite charge with energies that are nominal and that tend to fall off with distance. They are ubiquitous and because they act across long distances, they play a prominent role in the actions of ionizable drugs. The strength of an electrostatic force is directly dependent on the charge of each ion and inversely dependent on the dielectric constant of the solvent and the distance between the charges.

Hydrogen bonds are ubiquitous in nature: their multiple presence contributes to the stability of the (αhelix and base-pairing in DNA. Hydrogen bonding is based on an electrostatic interaction between the nonbonding electrons of a heteroatom (e.g., N, O, S) and the electron-deficient hydrogen atom of an −OH, SH, or NH group. Hydrogen bonds are strongly directional, highly dependent on the net degree of solvation, and rather weak, having energies ranging from 1 to 10 kcal/mol (47, 48). Bonds with this type of strength are of critical importance because they are stable enough to provide significant binding energy but weak enough to allow for quick dissociation. The greater electronegativity of atoms such as oxygen, nitrogen, sulfur, and halogen, compared to that of carbon, causes bonds between these atoms to have an asymmetric distribution of electrons, which results in the generation of electronic dipoles. Given that so many functional groups have dipole moments, ion-dipole and dipole-dipole interactions are frequent. The energy of dipole-dipole interactions can be described by Equation 1.8, where μ is the dipole moment, θ is the angle between the two poles of the dipole, D is the dielectric constant of the medium and r is the distance between the charges involved in the dipole.

$$E = 2\mu_1\mu_2\cos\theta_1\cos\theta_2/Dr^3 \qquad (1.8)$$

Although electrostatic interactions are generally restricted to polar molecules, there are also strong interactions between nonpolar molecules over small intermolecular distances. Dispersion or London/van der Waals forces are the universal attractive forces between atoms that hold nonpolar molecules together in the liquid phase. They are based on polarizability and these fluctuating dipoles or shifts in electron clouds of the atoms tend to induce opposite dipoles in adjacent molecules, resulting in a net overall attraction. The energy of this interaction decreases very rapidly in proportion to $1/r^6$, where r is the distance separating the two molecules. These van der Waals forces operate at a distance of about 0.4–0.6 nm and exert an attraction force of less than 0.5 kcal/mol. Yet, although individual van der Waals forces make a low energy contribution to an event, they become significant and additive when summed up over a large area with close surface contact of the atoms.

Hydrophobicity refers to the tendency of nonpolar compounds to transfer from an aqueous phase to an organic phase (49, 50). When a nonpolar molecule is placed in water, it gets solvated by a "sweater" of water molecules ordered in a somewhat icelike manner. This increased order in the water molecules surrounding the solute results in a loss of entropy. Association of hydrocarbon molecules leads to a "squeezing out" of the structured water molecules. The displaced water becomes bulk water, less ordered, resulting in a gain in entropy, which provides the driving force for what has been referred to as a hydrophobic bond. Although this is a generally accepted view of hydrophobicity, the hydration of apolar molecules and the noncovalent interactions between these molecules in water are still poorly understood and thus the source of continued examination (51–53).

Because noncovalent interactions are generally weak, cooperativity by several types of interactions is essential for overall activity. Enthalpy terms will be additive, but once the first interaction occurs, translational entropy is lost. This results in a reduced entropy loss in the second interaction. The net result is that eventually several weak interactions combine to produce a strong interaction. One can safely state that it is the involvement of myriad interactions that contribute to the overall selectivity of drug-receptor interactions.

2 TOOLS AND TECHNIQUES OF QSAR

2.1 Biological Parameters

In QSAR analysis, it is imperative that the biological data be both accurate and precise to develop a meaningful model. It must be realized that any resulting QSAR model that is developed is only as valid statistically as the data that led to its development. The equilibrium constants and rate constants that are used extensively in physical organic chemistry and medicinal chemistry are related to free energy values AG. Thus for use in QSAR, standard biological equilibrium constants such as K_i or K_m should be used in QSAR studies. Likewise only standard rate constants should be deemed appropriate for a QSAR analysis. Percentage activities (e.g., % inhibition of growth at certain concentrations) are not appropriate biological endpoints because of the nonlinear characteristic of dose-response relationships. These types of endpoints may be transformed to equieffective molar doses. Only equilibrium and rate constants pass muster in terms of the free-energy relationships or influence on QSAR studies. Biological data are usually expressed on a logarithmic scale because of the linear relationship between response and log dose in the midregion of the log dose-response curve. Inverse logarithms for activity ($\log 1/C$) are used so that higher values are obtained for more effective analogs. Various types of biological data have been used in QSAR analysis. A few common endpoints are outlined in Table 1.2.

Biological data should pertain to an aspect of biological/biochemical function that can be measured. The events could be occurring in enzymes, isolated or bound receptors, in cellular systems, or whole animals. Because there is considerable variation in biological responses, test samples should be run in duplicate or preferably triplicate, except in whole animal studies where assay conditions (e.g., plasma concentrations of a drug) preclude such measurements.

Table 1.2 Types of Biological Data Utilized in QSAR Analysis

Source of Activity	Biological Parameters
1. Isolated receptors	
Rate constants	Log k_{cat}; Log k_{uncat}; Log k
Michaelis–Menten constants	Log $1/K_m$
Inhibition constants	Log $1/K_i$
Affinity data	pA_2; pA_1
2. Cellular systems	
Inhibition constants	Log $1/IC_{50}$
Cross resistance	Log *CR*
In vitro biological data	Log $1/C$
Mutagenicity states	Log TA_{98}
3. "In vivo" systems	
Biocencentration factor	Log *BCF*
In vivo reaction rates	Log I (Induction)
Pharmacodynamic rates	Log T (total clearance)

It is also important to design a set of molecules that will yield a range of values in terms of biological activities. It is understandable that most medicinal chemists are reluctant to synthesize molecules with poor activity, even though these data points are important in developing a meaningful QSAR. Generally, the larger the range (>2 log units) in activity, the easier it is to generate a predictive QSAR. This kind of equation is more forgiving in terms of errors of measurement. A narrow range in biological activity is less forgiving in terms of accuracy of data. Another factor that merits consideration is the time structure. Should a particular reading be taken after 48 or 72 h? Knowledge of cell cycles in cellular systems or biorhythms in animals would be advantageous.

Each single step of drug transport, binding, and metabolism involves some form of partitioning between an aqueous compartment and a nonaqueous phase, which could be a membrane, serum protein, receptor, or enzyme. In the case of isolated receptors, the endpoint is clear-cut and the critical step is evident. But in more complex systems, such as cellular systems or whole animals, many localized steps could be involved in the random-walk process and the eventual interaction with a target. Usually the observed biological activity is reflective of the slow step or the rate-determining step.

To determine a defined biological response (e.g., IC_{50}), a dose-response curve is first established. Usually six to eight concentrations are tested to yield percentages of activity or inhibition between 20 and 80%, the linear portion of the curve. Using the curves, the dose responsible for an established effect can easily be determined. This procedure is meaningful if, at the time the response is measured, the system is at equilibrium, or at least under steady-state conditions.

Other approaches have been used to apply the additivity concept and ascertain the binding energy contributions of various substituent (R) groups. Fersht et al. have measured the binding energies of various alkyl groups to aminoacyl-tRNA synthetases (54). Thus the AG values for methyl, ethyl, isopropyl, and thio substituents were determined to be 3.2, 6.5, 9.6, and 5.4 kcal/mol, respectively.

An alternative, generalized approach to determining the energies of various drug-receptor interactions was developed by Andrews et al. (55), who statistically examined the drug-receptor interactions of a diverse set of molecules in aqueous solution. Using Equation 1.9, a relationship was established between AG and E_X (intrinsic binding energy), E_{DOF} (energy of average entropy loss), and the AS,, (energy of rotational and translational entropy loss).

$$\Delta G = T\,\Delta S_{r,t} + n_{DOF}E_{DOF} + n_X E_X \quad (1.9)$$

E_X denotes the sum of the intrinsic binding energy of each functional group of which n_X are present in each drug in the set. Using Equation 1.9, the average binding energies for various functional groups were calculated. These energies followed a particular trend with charged groups showing stronger interactions and nonpolar entities, such as sp^2, sp^3 carbons, contributing very little. The applicability of this approach to specific drug-receptor interactions remains to be seen.

2.2 Statistical Methods: Linear Regression Analysis

The most widely used mathematical technique in QSAR analysis is multiple regression

analysis (MRA). We will consider some of the basic tenets of this approach to gain a firm understanding of the statistical procedures that define a QSAR. Regression analysis is a powerful means for establishing a correlation between independent variables and a dependent variable such as biological activity (56).

$$Y_i = b + aX_i + E_i \qquad (1.10)$$

Certain assumptions are made with regard to this procedure (57):

1. The independent variables, which in this case usually include the physicochemical parameters, are measured without error. Unfortunately, this is not always the case, although the error in these variables is small compared to that in the dependent variable.
2. For any given value of X, the Y values are independent and follow a normal distribution. The error term E_i possesses a normal distribution with a mean of zero.
3. The expected mean value for the variable Y, for all values of X, lies on a straight line.
4. The variance around the regression line is constant. The "best" straight line for model $Y_i = b + aZ_i + E$ is drawn through the data points, such that the sum of the squares of the vertical distances from the points to the line is minimized. Y represents the value of the observed data point and Y_{calc} is the predicted value on the line. The sum of squares $SS = \Sigma (Y_{obs} - Y_{calc})^2$.

$$Y_{obs} = aX_i + b + E_i \qquad (1.11)$$

$$Y_{calc} = aX_i + b \qquad (1.12)$$

$$E = Y_{obs} - aX_i - b \qquad (1.13)$$

$$\sum_{i=1}^{n} E_i^2 = \sum A^2 = SS = \sum (Y_{obs} - Y_{calc})^2 \qquad (1.14)$$

$$\text{Thus, } SS = \sum_{i=1}^{n} (Y_{obs} - aX_i - b)^2 \qquad (1.15)$$

Expanding Equation 1.15, we obtain

$$SS = \sum_{i=1}^{n} (Y_{obs}^2 - Y_{obs}aX_i - Y_{obs}b - Y_{obs}aX_i + a^2X_i^2 + aX_ib - bY_{obs} + abX_i + b^2) \qquad (1.16)$$

Taking the partial derivative of Equation 1.14 with respect to b and then with respect to a, results in Equations 1.17 and 1.18.

$$\frac{dSS}{db} = \sum_{i=1}^{n} -2(Y_{obs} - b - aX_i) \qquad (1.17)$$

$$\frac{dSS}{da} = \sum_{i=1}^{n} -2X_i(Y_{obs} - b - aX_i) \qquad (1.18)$$

SS can be minimized with respect to b and a and divided by -2 to yield the normal Equations 1.19 and 1.20.

$$\sum_{i=1}^{n} (Y_{obs} - b - aX_i) = 0 \qquad (1.19)$$

$$\sum_{i=1}^{n} X_i(Y_{obs} - b - aX_i) = 0 \qquad (1.20)$$

These "normal equations" can be rewritten as follows:

$$b \sum_{i=1}^{n} X_i + a \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_iY_{obs} \qquad (1.21)$$

$$b + a \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_{obs} \qquad (1.22)$$

The solution of these simultaneous equations yields a and b. More thorough analyses of these procedures have been examined in detail (19, 58–60). The following simple example, illustrated by Table 1.3, will illustrate the nuances of a linear regression analysis.

Table 1.3 Antibacterial Activity of *N'*-(*R*-phenyl)sulfanilamides

Compound	$\sigma(X)$	Observed BA (Y)
1. 4-CH_3	−0.17	4.66
2. 4-H	0	4.80
3. 4-Cl	0.23	4.89
4. 2-Cl	0.23	5.55
5. 2-NO_2	0.78	6.00
6. 4-NO_2	0.78	6.00

k = no. of variables = 1
n = no. of data points = 6
$\Sigma X = 1.85$
$\Sigma Y = 31.90$
$\Sigma X^2 = 1.352$
$\Sigma Y^2 = 171.45$
$\Sigma XY = 10.968$

For linear regression analysis, Y = $ax + b$

$$a = (n \cdot \sum xy - \sum x \cdot \sum y)/n \cdot \sum x^2 - (\sum x)^2 = 1.45$$

$$b = (\sum y - a \cdot \sum x)/n = 4.869$$

$$r^2 = \sum xy - \sum x \cdot \sum y/n)^2/(\sum x^2 - (\sum x)^2/n) \cdot (\sum y^2 - (\sum y)^2/n) = 0.875 \quad \therefore r = 0.935$$

$$s^2 = (1 - r^2) \times (\sum y^2 - (\sum y)^2/n)/(n - k - 1) = 0.058 \quad \therefore s = 0.240$$

$$F = r^2 \cdot (n - k - 1)/k(1 - r^2) = 28.52$$

The correlation coefficient r, the total variance SS_T, the unexplained variance SSQ, and the standard deviation, are defined as follows:

$$r^2 = 1 - \frac{\Sigma \Delta^2}{SS_T} \tag{1.23}$$

$$SS_T = \sum (Y_{obs} - Y_{mean})^2 = \sum y^2 - (\sum y)^2/n \tag{1.24}$$

$$\sum A^2 = SSQ = \sum (Y_{obs} - Y_{calc})^2 \tag{1.25}$$

$$s = \sqrt{\frac{\Sigma \Delta^2}{n - k - 1}} = \sqrt{\frac{SSQ}{n - k - 1}} \tag{1.26}$$

The correlation coefficient r is a measure of quality of fit of the model. It constitutes the variance in the data. In an ideal situation one would want the correlation coefficient to be equal to or approach 1, but in reality because of the complexity of biological data, any value above 0.90 is adequate. The standard deviation is an absolute measure of the quality of fit. Ideally s should approach zero, but in experimental situations, this is not so. It should be small but it cannot have a value lower than the standard deviation of the experimental data. The magnitude of s may be attributed to some experimental error in the data as well as imperfections in the biological model. A larger data set and a smaller number of variables generally lead to lower values of s. The F value is often used as a measure of the level of statistical significance of the regression model. It is defined as denoted in Equation 1.27.

$$F_{k_2-k_1,n-k_2} = \frac{(SS_1 - SS_2)}{SS_2} \frac{(n - k_2 - 1)}{k_2 - k_1} \tag{1.27}$$

A larger value of F implies a more significant correlation has been reached. The confidence intervals of the coefficients in the equation reveal the significance of each regression term in the equation.

To obtain a statistically sound QSAR, it is important that certain caveats be kept in mind. One needs to be cognizant about collinearity between variables and chance correlations. Use of a correlation matrix ensures that variables of significance and/or interest are orthogonal to each other. With the rapid proliferation of parameters, caution must be exercised in amassing too many variables for a QSAR analysis. Topliss has elegantly demonstrated that there is a high risk of ending up with a chance correlation when too many variables are tested (62).

Outliers in QSAR model generation present their own problems. If they are badly fit by the model (off by more than 2 standard deviations), they should be dropped from the data set, although their elimination should be noted and addressed. Their aberrant behavior

may be attributed to inaccuracies in the testing procedure (usually dilution errors) or unusual behavior. They often provide valuable information in terms of the mechanistic interpretation of a QSAR model. They could be participating in some intermolecular interaction that is not available to other members of the data set or have a drastic change in mechanism.

2.3 Compound Selection

In setting up to run a QSAR analysis, compound selection is an important angle that needs to be addressed. One of the earliest manual methods was an approach devised by Craig, which involves two-dimensional plots of important physicochemical properties. Care is taken to select substituents from all four quadrants of the plot (63). The Topliss operational scheme allows one to start with two compounds and construct a potency tree that grows branches as the substituent set is expanded in a stepwise fashion (64). Topliss later proposed a batchwise scheme including certain substituents such as the 3,4-Cl_2, 4-Cl, 4-CH_3, 4-OCH_3, and 4-H analogs (65). Other methods of manual substituent selection include the Fibonacci search method, sequential simplex strategy, and parameter focusing by Magee (66–68).

One of the earliest computer-based and statistical selection methods, cluster analysis was devised by Hansch to accelerate the process and diversity of the substituents (1). Newer methodologies include D-optimal designs, which focus on the use of det $(X'X)$, the variance-covariance matrix. The determinant of this matrix yields a single number, which is maximized for compounds expressing maximum variance and minimum covariance (69–71). A combination of fractional factorial design in tandem with a principal property approach has proven useful in QSAR (72). Extensions of this approach using multivariate design have shown promise in environmental QSAR with nonspecific responses, where the clusters overlap and a cluster-based design approach has to be used (73). With strongly clustered data containing several classes of compounds, a new strategy involving local multivariate designs within each cluster is described. The chosen compounds from the local designs are grouped together in the overall training set that is representative of all clusters (74).

3 PARAMETERS USED IN QSAR

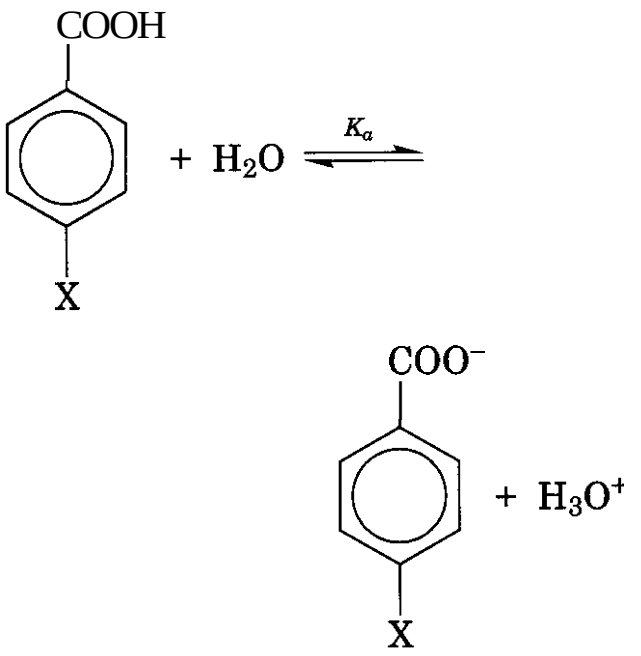
3.1 Electronic Parameters

Parameters are of critical importance in determining the types of intermolecular forces that underly drug-receptor interactions. The three major types of parameters that were initially suggested and still hold sway are electronic, hydrophobic, and steric in nature (20, 75). Extensive studies using electronic parameters reveal that electronic attributes of molecules are intimately related to their chemical reactivities and biological activities. A search of a computerized QSAR database reveals the following: the common Hammett constants (σ, σ^+, σ^-) account for 700018500 equations in the Physical organic chemistry (PHYS) database and nearly 1600/8000 in the Biology (BIO) database, whereas quantum chemical indices such as HOMO, LUMO, BDE, and polarizability appear in 100 equations in the BIO database (76).

The extent to which a given reaction responds to electronic perturbation constitutes a measure of the electronic demands of that reaction, which is determined by its mechanism. The introduction of substituent groups into the framework and the subsequent alteration of reaction rates helps delineate the overall mechanism of reaction. Early work examining the electronic role of substituents on rate constants was first tackled by Burckhardt and firmly established by Hammett (13, 14, 77, 78). Hammett employed, as a model reaction, the ionization in water of substituted benzoic acids and determined their equilibrium constants K_a. See Equation 1.28. This led to an operational definition of σ, the substituent constant. It is a measure of the size of the electronic effect for a given substituent and represents a measure of electronic charge distribution in the benzene nucleus.

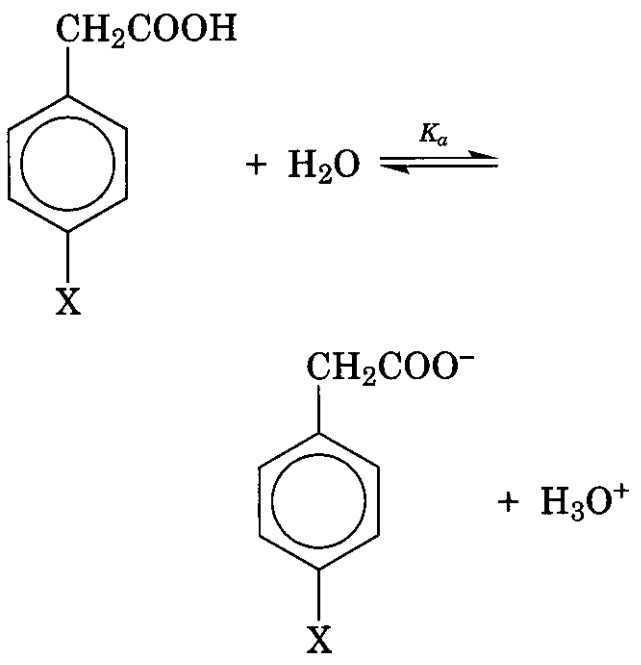
$$\sigma_X = \log K_X - \log K_H \quad or \quad \log(K_X/K_H) = -pK_X + pK_H \tag{1.29}$$

Electron-withdrawing substituents are thus



(1.28)

characterized by positive values, whereas electron-donating ones have negative values. In an extension of this approach, the ionization of substituted phenylacetic acids was measured.



(1.30)

The effect of the 4-C1 substituent on the ionization of 4-Cl phenylacetic acid (PA) was found to be proportional to its effect on the ionization of 4-C1 benzoic acid (BA).

$$\log(K'_{\mathrm{Cl(PA)}}/K'_{\mathrm{H(PA)}}) \propto \log(K_{\mathrm{Cl(BA)}}/K_{\mathrm{H(BA)}})$$

Since $\log(K_{\mathrm{Cl(BA)}}/K_{\mathrm{H(BA)}}) = \sigma$,

$$\text{then } \log \frac{K'_{\mathrm{Cl}}}{K'_{\mathrm{H}}} = \rho \cdot \sigma \qquad (1.31)$$

ρ (rho) is defined as a proportionality or reaction constant, which is a measure of the susceptibility of a reaction to substituent effects. A positive rho value suggests that a reaction is aided by electron withdrawal from the reaction site, whereas a negative rho value implies that the reaction is assisted by electron donation at the reaction site. Hammett also drew attention to the fact that a plot of log K_{A} for benzoic acids versus log k for ester hydrolysis of a series of molecules is linear, which suggests that substituents exert a similar effect in dissimilar reactions.

$$\log \frac{k_{\mathrm{X}}}{k_{\mathrm{H}}} \propto \log \frac{K_{\mathrm{X}}}{K_{\mathrm{H}}} = \rho \cdot a \qquad (1.32)$$

Although this expression is empirical in nature, it has been validated by the sheer volume of positive results. It is remarkable because four different energy states must be related.

A correlation of this type is clearly meaningful; it suggests that changes in structure produce proportional changes in the activation energy AG* for such reactions. Hence, the derivation of the name for which the Hammett equation is universally known: linear free energy relationship (LFER). Equation 1.32 has become known as the Hammett equation and has been applied to thousands of reactions that take place at or near the benzene ring bearing substituents at the meta and para positions. Because of proximity and steric effects, ortho-substituted molecules do not always follow this maxim and are subject to different parameterizations. Thus, an expanded approach was established by Charton (79) and Fujita and Nishioka (80). Charton partitioned the ortho electronic effect into its inductive, resonance, and steric contributions; the factors a, β, and X are susceptibility or reaction constants and h is the intercept.

$$\mathrm{Log\ k} = \alpha\sigma_I + \beta\sigma_{\mathrm{R}} + Xr, + h \qquad (1.33)$$

Fujita and Nishioka used an integrated approach to deal with ortho substituents in data sets including meta and para substituents.

$$\mathrm{Log\ k} = \rho a + \delta E_{\mathrm{S}}^{\mathrm{ortho}} + fF_{\mathrm{ortho}} + C \qquad (1.34)$$

For ortho substituents, para sigma values

were used in addition to Taft's E_S values and Swain-Lupton field constants F_{ortho}.

The reason for employing alternative treatments to ortho-substituted aromatic molecules is that changes in rate or ionization constants mediated by meta or para substituents are mostly changes in ($H^{\ddagger}$ or ΔH° because substitution does not affect AS* or AS". Ortho substituents affect both enthalpy and entropy; the effect on entropy is noteworthy because entropy is highly sensitive to changes in the size of reagents and substituents as well as degree of solvation. Bolton et al. examined the ionization of substituted benzoic acids and measured accurate values for AG, AH, and A S (81). A hierarchy of different scenarios, under which an LFER operates, was established:

1. ΔH° is constant and A S varies for a series.
2. AS" is constant and AH varies.
3. ΔH° and AS" vary and are shown to be linearly related.
4. Precise measurements indicated that category 3 was the prevalent behavior in benzoic acids.

Despite the extensive and successful use in QSAR studies, there are some limitations to the Hammett equation.

1. Primary a values are obtained from the thermodynamic ionizations of the appropriate benzoic acids at 25°C; these are reliable and easily available. Secondary values are obtained by comparison with another series of compounds and are thus subject to error because they are dependent on the accuracy of a measured series and the development of a regression line using statistical methods.
2. In some multisubstituted compounds, the lack of additivity needs to be noted. Proximal effects are operative and tend to distort electronic contributions. For example,

$$\sum \sigma_{calc}(3,4,5\text{-trichlorobenzoic acid}) = 0.97;$$

that is, $2\sigma_M + \sigma_P$ *or* $2(0.37) + 0.23$

$$\sum \sigma_{obs}(3,4,5\text{-trichlorobenzoic acid}) = 0.95$$

Sigma values for smaller substituents are more likely to be additive. However, in the case of 3-methyl, 4-dimethylaminobenzoic acid, the discrepancy is high. For example,

$$\sum \sigma_{calc}(3\text{-}CH_3,\ 4\text{-}N(CH_3)_2 \text{ benzoic acid}) = -0.90$$

$$\sum \sigma_{obs}(3\text{-}CH_3,\ 4\text{-}N(CH_3)_2 \text{ benzoic acid}) = -0.30$$

The large discrepancy may be attributed to the twisting of the dimethylamino substituent out of the plane of the benzene ring, resulting in a decrease in resonance. Exner and his colleagues have critically examined the use of additivity in the determination of a constants (82).

3. Changes in mechanism or transition state cause discontinuities in Hammett plots. Nonlinear plots are often found in reactions that proceed by two concurrent pathways (83, 84).
4. Changes in solvent may lead to dissimilarities in reaction mechanisms. Thus extrapolation of σ values from a polar solvent (e.g., CH_3CN) to a nonpolar solvent such as benzene has to be approached cautiously. Solvation properties will differ considerably, particularly if the transition state is polar and/or the substituents are able to interact with the solvent.
5. A strong positional dependency of sigma makes it imperative to use appropriate values for positional, isomeric substituents. Substituents ortho to the reaction center are difficult to describe and thus one must resort to a Fujita-Nishioka analysis (80).
6. Thorough resonance or direct conjugation effects cause a breakdown in the Hammett equation. When coupling occurs between the substituent and the reaction center through the pi-electron system, reactivity is enhanced, diminished, or mitigated by separation. In a study of X-cumyl chlorides, Brown and Okamoto noticed the strong conjugative interaction between lone-pair,

para substituents and the vacant p-orbital in the transition state, which led to deviations in the Hammett plot (85). They defined a modified LFER applicable to this situation.

$$\text{Log}\frac{k_Y}{k_H} = (\rho^+)(\sigma^+) \qquad (1.35)$$

σ^+ was a new substituent constant that expressed enhanced resonance attributes. A similar situation was noticed when a strong donor center was present as a reactant or formed as a product (e.g., phenols and anilines). In this case, strong resonance interactions were possible with electron-withdrawing groups (e.g., NO, or CN). A scale for such substituents was constructed such that

$$\text{Log}\frac{k_Y}{k_H} = (\rho^-)(\sigma^-) \qquad (1.36)$$

One shortcoming of the benzoic acid system is the extent of coupling between the carboxyl group and certain lone-pair donors. Insertion of a methylene group between the core (benzene ring) and the functional group (COOH moiety) leads to phenylacetic acids and the establishment of σ^0 scale from the ionization of X-phenylacetic acids. A flexible method of dealing with the variability of the resonance contribution to the overall electronic demand of a reaction is embodied in the Yukawa-Tsuno equation (86). It includes normal and enhanced resonance contributions to an LFER.

$$\text{Log}\frac{k_Y}{k_H} = \rho[\sigma + r(\sigma^+ - a)] \qquad (1.37)$$

where r is a measure of the degree of enhanced resonance interaction in relation to benzoic acid dissociations ($r = 0$) and cumyl chloride hydrolysis ($r = 1$).

Most of the Hammett-type constants pertain to aromatic systems. In evaluating an electronic parameter for use in aliphatic systems, Taft used the relative acid and base hydrolysis rates for esters. He developed equation 1.38 as a measure of the inductive effect (a*) of a substituent R' in the ester R' COOR, where B and A refer to basic and acidic hydrolysis, respectively.

$$\sigma^* = \frac{1}{2.48}[\log(k/k_O)_B - \log(k/k_O)_A] \qquad (1.38)$$

The factor of 2.48 was used to make a* equiscalar with Hammett a values. Later, a σ_I scale derived from the ionization of 4-X-bicyclo[2.2.2]octane-1-carboxylic acids was shown to be related to a* (87, 88). It is now more widely used than a*.

$$\sigma_I(X) = 0.45\sigma^*(CH_2X) \qquad (1.39)$$

Ionization is a function of the electronic structure of an organic drug molecule. Albert was the first to clearly delineate the relationship between ionization and biological activity (89). Now, pK_a values are widely used as the independent variable in physical organic reactions and in biological systems, particularly when dealing with transport phenomena. However, caution must be exercised in interpreting the dependency of biological activity on pK_a values because pK_a values are inherently composites of electronic factors that are used directly in QSAR analysis.

In recent years, there has been a rapid growth in the application of quantum chemical methodology to QSAR, by direct derivation of electronic descriptors from the molecular wave functions (90). The two most popular methods used for the calculation of quantum chemical descriptors are *ab initio* (Hartree-Fock) and semiempirical methods. As in other electronic parameters, QSAR models incorporating quantum chemical descriptors will include information on the nature of the intermolecular forces involved in the biological response. Unlike other electronic descriptors, there is no statistical error in quantum chemical computations. The errors are usually made in the assumptions that are established to facilitate calculation (91). Quantum chemical descriptors such as net atomic changes, highest occupied molecular orbital/lowest unoccupied molecular orbital (HOMO-LUMO) energies, frontier orbital electron densities, and superdelocalizabilities have been shown

to correlate well with various biological activities (92). A mixed approach using frontier orbital theory and topological parameters have been used to calculate Hammett-like substituent constants (93).

$$\sigma = -2.480\Delta N - 7.894\Delta E - 0.605 D_X/D_H \cdot (EA_H/EA_X) + 0.009\, \emptyset S_X + 0.028 \sum \pi + 0.279 \quad (1.40)$$

$$n = 150, \quad r^2 = 0.947, \quad s = 0.079, \quad F = 789.9$$

In Equation 1.40, AN represents the extent of electron transfer between interacting acid-base systems; AE is the energy decrease in bimolecular systems underlying electron transfer; $D_X/D_H \cdot (EA_H/EA_X)$ corresponds to electron affinity and distance terms; and $\emptyset S_X$ factors the electrotopological state index, whereas Σ a is the number of all π-electrons in the functional group. Observed principal component analysis (PCA) clustering of 66 descriptors derived from AM1 calculations was similar to that previously reported for monosubstituted benzenes (94, 95). The advantages of quantum chemical descriptors are that they have definite meaning and are useful in the elucidation of intra- and intermolecular interactions and can easily be derived from the theoretical structure of the molecule.

3.2 Hydrophobicity Parameters

More than a hundred years ago, Meyer and Overton made their seminal discovery on the correlation between oil/water partition coefficients and the narcotic potencies of small organic molecules (7, 8). Ferguson extended this analysis by placing the relationship between depressant action and hydrophobicity in a thermodynamic context; the relative saturation of the depressant in the biophase was a critical determinant of its narcotic potency (9). At this time, the success of the Hammett equation began to permeate structure-activity studies and hydrophobicity as a determinant was relegated to the background. In a landmark study, Hansch and his colleagues devised and used a multiparameter approach that included both electronic and hydrophobic terms, to establish a QSAR for a series of plant growth regulators (16). This study laid the basis for the development of the QSAR paradigm and also firmly established the importance of lipophilicity in biosystems. Over the last 40 years, no other parameter used in QSAR has generated more interest, excitement, and controversy than hydrophobicity (96). Hydrophobic interactions are of critical importance in many areas of chemistry. These include enzyme-ligand interactions, the assembly of lipids in biomembranes, aggregation of surfactants, coagulation, and detergency (97–100). The integrity of biomembranes and the tertiary structure of proteins in solution are determined by apolar-type interactions.

Molecular recognition depends strongly on hydrophobic interactions between ligands and receptors. Excellent treatises on this subject have been written by Taylor (101) and Blokzijl and Engerts (51). Despite extensive usage of the term hydrophobic bond, it is well known that there is no strong attractive force between apolar molecules (102). Frank and Evans were the first to apply a thermodynamic treatment to the solvation of apolar molecules in water at room temperature (103). Their "iceberg" model suggested that a large entropic loss ensued after the dissolution of apolar compounds and the increased structure of water molecules in the surrounding apolar solute. The quantitation of this model led to the development of the "flickering" cluster model of Némethy and Scheraga, which emphasized the formation of hydrogen bonds in liquid water (104). The classical model for hydrophobic interactions was delineated by Kauzmann to describe the van der Waals attractions between the nonpolar parts of two molecules immersed in water. Given that van der Waals forces operate over short distances, the water molecules are squeezed out in the vicinity of the mutually bound apolar surfaces (49). The driving force for this behavior is not that alkanes "hate" water, but rather water that "hates" alkanes (105, 106). Thus, the gain in entropy appears as the critical driving force for hydrophobic interactions that are primarily governed by the repulsion of hydrophobic

solutes from the solvent water and the limited but important capacity of water to maintain its network of hydrogen bonds.

Hydrophobicities of solutes can readily be determined by measuring partition coefficients designated as P. Partition coefficients deal with neutral species, whereas distribution ratios incorporate concentrations of charged and/or polymeric species as well. By convention, P is defined as the ratio of concentration of the solute in octanol to its concentration in water.

$$P = [\text{conc}]_{\text{octanol}}/[\text{conc}]_{\text{aqueous}} \qquad (1.41)$$

It was fortuitous that octanol was chosen as the solvent most likely to mimic the biomembrane. Extensive studies over the last 35 years (40,000 experimental P-values in 400 different solvent systems) have failed to dislodge octanol from its secure perch (107,108).

Octanol is a suitable solvent for the measurement of partition coefficients for many reasons (109, 110). It is cheap, relatively nontoxic, and chemically unreactive. The hydroxyl group has both hydrogen bond acceptor and hydrogen bond donor features capable of interacting with a large variety of polar groups. Despite its hydrophobic attributes, it is able to dissolve many more organic compounds than can alkanes, cycloalkanes, or aromatic hydrocarbons. It is UV transparent over a large range and has a vapor pressure low enough to allow for reproducible measurements. It is also elevated enough to allow for its removal under mild conditions. In addition, water saturated with octanol contains only 10^{-3} M octanol at equilibrium, whereas octanol saturated with water contains 2.3 M of water. Thus, polar groups need not be totally dehydrated in transfer from the aqueous phase to the organic phase. Likewise, hydrophobic solutes are not appreciably solvated by the 10^{-3} M octanol in the water phase unless their intrinsic log P is above 6.0. Octanol begins to absorb light below 220 nm and thus solute concentration determinations can be monitored by UV spectroscopy. More important, octanol acts as an excellent mimic for biomembranes because it shares the traits of amphiphilicity and hydrogen-bonding capability with phospholipids and proteins found in biological membranes.

The choice of the octanol/water partitioning system as a standard reference for assessing the compartmental distribution of molecules of biological interest was recently investigated by molecular dynamics simulations (111). It was determined that pure 1-octanol contains a mix of hydrogen-bonded "polymeric" species, mostly four-, five-, and six-membered ring clusters at 40°C. These small ring clusters form a central hydroxyl core from which their corresponding alkyl chains radiate outward. On the other hand, water-saturated octanol tends to form well-defined, inverted, micellar aggregates. Long hydrogen-bonded chains are absent and water molecules congregate around the octanol hydroxyls. "Hydrophilic channels" are formed by cylindrical formation of water and octanol hydroxyls with the alkyl chains extending outward. Thus, water-saturated octanol has centralized polar cores where polar solutes can localize. Hydrophobic solutes would migrate to the alkyl-rich regions. This is an elegant study that provides insight into the partitioning of benzene and phenol by analyzing the structure of the octanol/water solvation shell and delineating octanol's capability to serve as a surrogate for biomembranes.

The shake-flask method, so-called, is most commonly used to measure partition coefficients with great accuracy and precision and with a log P range that extends from -3 to $+6$ (112, 113). The procedure calls for the use of pure, distilled, deionized water, high-purity octanol, and pure solutes. At least three concentration levels of solute should be analyzed and the volumes of octanol and water should be varied according to a rough estimate of the log P value. Care should be exercised to ensure that the eventual amounts of the solute in each phase are about the same after equilibrium. Standard concentration curves using three to four known concentrations in water saturated with octanol are usually established. Generally, most methods employ a UV-based procedure, although GC and HPLC may also be used to quantitate the concentration of the solute.

Generally, 10-mL stopped centrifuge tubes or 200-mL centrifuge bottles are used. They are inverted gently for 2–3 min and then centrifuged at 1000–2000 *g* for 20 min before the phases are analyzed. Analysis of both phases is highly recommended, to minimize errors incurred by adsorption to glass walls at low solute concentration. For highly hydrophobic compounds, the slow stirring procedure of de Bruijn and Hermens is recommended (114). The filler probe extractor system of Tomlinson et al. is a modified, automated, shake-flask method, which is efficient, fast, reliable, and flexible (115).

Partition coefficients from different solvent systems can also be compared and converted to the octanol/water scale, as was suggested by Collander (116). He stressed the importance of the following linear relationship: $\log P_2 = a \log P_1 + b$. This type of relationship works well when the two solvents are both alkanols. However, when two solvent systems have varying hydrogen bond donor and acceptor capabilities, the relationship tends to fray. A classical example involves the relationship between log P values in chloroform and octanol (117, 118).

$$\text{Log } P_{CHCl_3} = 1.012 \log P_{oct} - 0.513 \quad (1.42)$$
$$n = 72, \quad r^2 = 0.811, \quad s = 0.733$$

Only 66% of the variance in the data is explained by this equation. However, a separation of the various solutes into OH bond donors, acceptors, and neutrals helped account for 94% of the variance in the data. These restrictions led Seiler to extend the Collander equation by incorporating a corrective term for H-bonding in the cyclohexane system (119). Fujita generalized this approach and formulated Equation 1.43 as shown below (120).

$$\log P_2 = a \log P_1 + \sum b_i \cdot HB_i + C \quad (1.43)$$

P_1 is the reference solvent and HB_i is an H-bonding parameter. Leahy et al. suggested that a more sophisticated approach incorporating four model systems would be needed to adequately address issues of solute partitioning in membranes (121). Thus, four distinct solvent types were chosen—apolar, amphiprotic, proton donor, and proton acceptor—and they were represented by alkanes, octanol, chloroform, and propyleneglycol dipelargonate (PGDP), respectively. The demands of measuring four partition coefficients for each solute has slowed progress in this particular area.

3.2.1 Determination of Hydrophobicity by Chromatography. Chromatography provides an alternate tool for the estimation of hydrophobicity parameters. R_m values derived from thin-layer chromatography provide a simple, rapid, and easy way to ascertain approximate values of hydrophobicity (122,123).

$$R_m = \log(1/R_f - 1) \quad (1.44)$$

Other recent developments in chromatography techniques have led to the development of powerful tools to rapidly and accurately measure octanol/water partition coefficients. Countercurrent chromatography is one of these methods. The stationary and mobile phases include two nonmiscible solvents (water and octanol) and the total volume of the liquid stationary phase is used for solute partitioning (124,125). Log P_{app} values of several diuretics including ionizable drugs have been measured at different pH values using countercurrent chromatography; the log *P* values ranged from −1.3 to 2.7 and were consistknt with literature values (126).

Recently, a rapid method for the determination of partition coefficients using gradient reversed phase/high pressure liquid chromatography (RP-HPLC) was developed. This method is touted as a high-throughput hydrophobicity screen for combinatorial libraries (127, 128). A chromatography hydrophobicity index (CHI) was established for a diverse set of compounds. Acetonitrile was used as the modifier and 50 mm ammonium acetate as the mobile phase (127). A linear relationship was established between Clog *P* and CHIN for neutral molecules.

$$\text{Clog P} = 0.057 \text{ CHIN} - 1.107 \quad (1.45)$$
$$n = 52, \quad r^2 = 0.724, \quad s = 0.82, \quad F = 131$$

A more recent study using RP-HPLC for the determination of log *P* (octanol) values for

neutral and weakly acidic and basic drugs, revealed an excellent correlation between log P_{oct} and log K_W values (129). Log P_{oct} values determined in this system are referred to as Elog P_{oct}. They were expressed in terms of solvation parameters.

$$\text{Elog } P_{oct} = 0.204 + 0.452R_2 - 1.053\pi_2^H - 0.041 \sum \alpha_2^H - 3.410 \sum \beta_2^O + 3.842V_X \quad (1.46)$$

$$n = 35, \quad r^2 = 0.960, \quad s = 0.244$$

In this equation, R, is the excess molar refraction; π_2^H is the dipolarity/polarizability; $\Sigma \alpha_2^H$ and $\Sigma \beta_2^O$ are the summation of hydrogen bond acidity and basicity values, respectively; and V_X is McGowan's volume.

3.2.2 Calculation Methods. Partition coefficients are additive-constitutive, free energy-related properties. Log P represents the overall hydrophobicity of a molecule, which includes the sum of the hydrophobic contributions of the "parent" molecule and its substituent. Thus, the π value for a substituent may be defined as

$$\pi_X = \log P_{R-X} - \log P_{R-H} \quad (1.47)$$

π_H is set to zero. The π-value for a nitro substituent is calculated from the log P of nitrobenzene and benzene.

$$\pi_{NO_2} = \log P_{\text{nitrobenzene}} - \log P_{\text{benzene}}$$
$$= 1.85 - 2.13 = -0.28$$

An extensive list of π-values for aromatic substituents appears in Table 1.4. Pi values for side chains of amino acids in peptides have been well characterized and are easily available (130–132). Aliphatic fragments values were developed a few years later. For a more extensive list of substituent value constants, refer to the extensive compilation by Hansch et al. (133). Initially, the π-system was applied only to substitution on aromatic rings and when the hydrogen being replaced was of innocuous character. It was apparent from the beginning that not all hydrogens on aromatic systems could be substituted without correction factors because of strong electronic interactions. It became necessary to determine π values in various electron-rich and -deficient systems (e.g., X-phenols and X-nitrobenzenes). Correction factors were introduced for special features such as unsaturation, branching, and ring fusion. The proliferation of π-scales made it difficult to ascertain which system was more appropriate for usage, particularly with complex structures.

The shortcomings of this approach provided the impetus for Nys and Rekker to design the fragmental method, a "reductionist" approach, which was based on the statistical analysis of a large number of measured partition coefficients and the subsequent assignment of appropriate values for particular molecular fragments (118, 134). Hansch and Leo took a "constructionist" approach and developed a fragmental system that included correction factors for bonds and proximity effects (1, 135). Labor-intensive efforts and inconsistency in manual calculations were eliminated with the debut of the automated system CLOGP and its powerful SMILES notation (136–138). Recent analysis of the accuracy of CLOGP yielded Equation 1.48 (139).

$$\text{MLOGP} = 0.959\ \text{CLOGP} + 0.08 \quad (1.48)$$

$$n = 12{,}107, \quad r^2 = 0.973, \quad s = 0.299$$

The Clog P values of 228 structures (1.8% of the data set) were not well predicted. It must be noted that Starlist (most accurate values in the database) contains almost 300 charged nitrogen solutes (ammonium, pyridinium, imidazolium, etc.) and over 2200 in all, which amounts to 5% of Masterfile (database of measured values). CLOGP adequately handles these molecules within the 0.30 standard deviation limit. Most other programs make no attempt to calculate them. For more details on calculating log P_{oct} from structures, see excellent reviews by Leo (140, 141).

The proliferation of methodologies and programs to calculate partition coefficients continues unabated. These programs are based on substructure approaches or whole-molecule approaches (142, 143). Substructure

Table 1.4 Substituent Constants for QSAR Analysis

No.	Substituent	Pi	MR	L	B1	B5	S-P	S-M
1	$+N(CH_3)_3$	−5.96	1.94	4.02	2.57	3.11	0.82	0.88
2	$EtN(CH_3)_3+$	−5.44	2.87	5.58	1.52	4.53	0.13	0.16
3	$CH_2N(CH_3)_3+$	−4.57	2.40	4.83	1.52	4.08	0.44	0.40
4	CO_2-	−4.36	0.61	3.53	1.60	2.66	0.00	−0.10
5	$+NH_3$	−4.19	0.55	2.78	1.49	1.97	0.60	0.86
6	$PR\text{-}N(CH_3)_3+$	−4.15	3.33	6.88	1.52	5.49	−0.01	0.06
7	CH_2NH_3+	−4.09	1.01	4.02	1.52	3.05	0.29	0.32
8	IO_2	−3.46	6.35	4.25	2.15	3.66	0.78	0.68
9	$C(CN)_3$	−2.33	1.86	3.99	2.87	4.12	0.96	0.97
10	$NHNO_2$	−2.27	1.07	4.50	1.35	3.66	0.57	0.91
11	$C(NO_2)_3$	−2.01	2.27	4.59	2.55	3.72	0.82	0.72
12	$SO_2(NH_2)$	−1.82	1.23	4.02	2.04	3.05	0.60	0.53
13	$C(CN)=C(CN)_2$	−1.77	2.58	6.46	1.61	5.17	0.98	0.77
14	$CH_2C=O(NH_2)$	−1.68	1.44	4.58	1.52	4.37	0.07	0.06
15	$N(COCH_3)_2$	−1.68	2.48	4.45	1.35	4.33	0.33	0.35
16	SO_2CH_3	−1.63	1.35	4.11	2.03	3.17	0.72	0.60
17	$P(O)(OH)_2$	−1.59	1.26	4.22	2.12	2.88	0.42	0.36
18	$S=O(CH_3)$	−1.58	1.37	4.11	1.40	3.17	0.49	0.52
19	$N(SO_2CH_3)_2$	−1.51	3.12	4.83	1.36	3.72	0.49	0.47
20	$C=O(NH_2)$	−1.49	0.98	4.06	1.50	3.07	0.36	0.28
21	$CH(CN)_2$	−1.45	1.43	3.99	1.85	4.12	0.52	0.53
22	$CH_2NHCOCH_3$	−1.43	1.96	5.67	1.52	4.75	−0.05	0.05
23	$NHC=S(NH_2)$	−1.40	2.22	5.06	1.35	4.18	0.16	0.22
24	NH(OH)	−1.34	0.72	3.87	1.35	2.63	−0.34	−0.04
25	$CH=NNHCONHNH_2$	−1.32	2.42	7.57	1.60	4.55	0.16	0.22
26	$NHC=O(NH_2)$	−1.30	1.37	5.06	1.35	3.61	−0.24	−0.03
27	$C=O(NHCH_3)$	−1.27	1.46	5.00	1.54	3.16	0.36	0.35
28	2-Aziridinyl	−1.23	1.19	4.14	1.55	3.24	−0.10	−0.06
29	NH_2	−1.23	0.54	2.78	1.35	1.97	−0.66	−0.16
30	$NHSO_2CH_3$	−1.18	1.82	4.70	1.35	4.13	0.03	0.20
31	$P(O)(OCH_3)_2$	−1.18	2.19	5.04	2.42	3.25	0.53	0.42
32	$C(CH_3)(CN)_2$	−1.14	1.90	4.11	2.81	4.12	0.57	0.60
33	$N(CH_3)SO_2CH_3$	−1.11	2.34	4.83	1.35	3.72	0.24	0.21
34	SO_2Et	−1.10	1.81	4.92	2.03	3.49	0.77	0.66
35	CH_2NH_2	−1.04	0.91	4.02	1.52	3.05	−0.11	−0.03
36	1-Tetrazolyl	−1.04	1.83	5.28	1.71	3.12	0.50	0.52
37	CH_2OH	−1.03	0.72	3.97	1.52	2.70	0.00	0.00
38	$N(CH_3)COCH_3$	−1.02	1.96	4.77	1.35	3.71	0.26	0.31
39	NHCHO	−0.98	1.03	4.22	1.35	3.61	0.00	0.19
40	$NHC(=O)CH_3$	−0.97	1.49	5.09	1.35	3.61	0.00	0.21
41	$C(CH_3)(NO_2)_2$	−0.88	2.17	4.59	2.55	3.72	0.61	0.54
42	$NHNH_2$	−0.88	0.84	3.47	1.35	2.97	−0.55	−0.02
43	OSO_2CH_3	−0.88	1.70	4.66	1.35	4.10	0.36	0.39
44	$SO_2N(CH_3)_2$	−0.78	2.19	4.83	2.03	4.08	0.65	0.51
45	$NHC=S(NHC_2H_5)$	−0.71	3.17	7.22	1.45	4.38	0.07	0.30
46	$SO_2(CHF_2)$	−0.68	1.31	4.11	2.03	3.70	0.86	0.75
47	OH	−0.67	0.29	2.74	1.35	1.93	−0.37	0.12
48	CHO	−0.65	0.69	3.53	1.60	2.36	0.42	0.35
49	$CH_2CHOHCH_3$	−0.64	1.64	4.92	1.52	3.78	−0.17	−0.12
50	$CS(NH_2)$	−0.64	1.81	4.10	1.64	3.18	0.30	0.25
51	$OC=O(CH_3)$	−0.64	1.25	4.74	1.35	3.67	0.31	0.39
52	$SOCHF_2$	−0.63	1.33	4.70	1.40	3.70	0.58	0.54
53	4-Pyrimidinyl	−0.61	2.18	5.29	1.71	3.11	0.63	0.30
54	2-Pyrimidinyl	−0.61	2.18	6.28	1.71	3.11	0.53	0.23

Table 1.4 (Continued)

No.	Substituent	Pi	MR	L	B1	B5	S-P	S-M
55	$P(CF_3)_2$	−0.59	1.99	4.96	1.40	3.86	0.69	0.60
56	CH_2CN	−0.57	1.01	3.99	1.52	4.12	0.18	0.16
57	CN	−0.57	0.63	4.23	1.60	1.60	0.66	0.56
58	$COCH_3$	−0.55	1.12	4.06	1.60	3.13	0.50	0.38
59	$CH_2P{=}O(OEt)_2$	−0.54	3.58	7.10	1.52	5.73	0.06	0.12
60	$P{=}O(OEt)_2$	−0.52	3.12	6.26	2.52	5.58	0.60	0.55
61	NHCOOMe	−0.52	1.57	5.84	1.45	3.99	−0.17	−0.02
62	$NHC{=}O(NHC_2H_5)$	−0.50	2.32	7.29	1.45	3.98	−0.26	0.04
63	$NHC{=}O(CH_2Cl)$	−0.50	1.98	6.26	1.55	4.26	−0.03	0.17
64	$NHCH_3$	−0.47	1.03	3.53	1.35	3.08	−0.70	−0.21
65	$N(CH_3)COCF_3$	−0.46	1.95	5.20	1.56	3.96	0.39	0.41
66	$C{=}S(NHCH_3)$	−0.46	2.23	5.00	1.88	3.18	0.34	0.30
67	$NHC{=}S(CH_3)$	−0.42	2.34	5.09	1.45	4.38	0.12	0.24
68	$C(Et)(NO_2)_2$	−0.35	3.66	4.92	2.55	3.72	0.64	0.56
69	CO_2H	−0.32	0.69	3.91	1.60	2.66	0.45	0.37
70	$C(OH)(CH_3)_2$	−0.32	1.64	4.11	2.40	3.17	0.60	0.47
71	$EtCO_2H$	−0.29	1.65	5.97	1.52	3.31	−0.07	−0.03
72	NO_2	−0.28	0.74	3.44	1.70	2.44	0.78	0.71
73	$CH{=}NNHCSNH_2$	−0.27	2.96	7.16	1.60	5.41	0.40	0.45
74	NHCN	−0.26	1.01	3.90	1.35	4.05	0.06	0.21
75	$CH_2C(OH)(CH_3)_2$	−0.24	2.11	4.92	1.52	4.19	−0.17	−0.16
76	$CH{=}CHCHO$	−0.23	1.69	5.76	1.60	3.46	0.13	0.24
77	$NHCH_2CO_2Et$	−0.21	2.69	7.91	1.35	5.77	−0.68	−0.10
78	CH_2OCH_3	−0.21	1.21	4.78	1.52	3.40	0.01	0.08
79	$NHC{=}OCH(CH_3)_2$	−0.18	2.43	5.53	1.35	4.09	−0.10	0.11
80	$CH_2OC{=}O(CH_3)$	−0.17	1.65	5.46	1.52	4.46	0.05	0.04
81	$CH_2N(CH_3)_2$	−0.15	1.87	4.83	1.52	4.08	0.01	0.00
82	CH_2SCN	−0.14	1.81	6.63	1.52	3.41	0.14	0.12
83	1-Aziridinyl	−0.12	1.35	4.14	1.35	3.24	−0.22	−0.07
84	NO	−0.12	0.52	3.44	1.70	2.44	0.91	0.62
85	ONO_2	−0.12	0.85	4.46	1.35	3.62	0.70	0.55
86	$S{=}O(C_6H_5)$	−0.07	3.34	4.62	1.40	6.02	0.44	0.50
87	$CH_2SO_2C_6H_5$	−0.06	3.79	8.33	1.52	3.78	0.16	0.15
88	OCH_3	−0.02	0.79	3.98	1.35	3.07	−0.27	0.12
89	$C{=}O(OCH_3)$	−0.01	1.29	4.73	1.64	3.36	0.45	0.36
90	H	0.00	0.10	2.06	1.00	1.00	0.00	0.00
91	$C{=}O(CF_3)$	0.02	1.12	4.65	1.70	3.67	0.80	0.63
92	$CH{=}C(CN)_2$	0.05	1.97	6.46	1.60	5.17	0.84	0.66
93	$SO_2(F)$	0.05	0.87	3.33	2.01	2.70	0.91	0.80
94	COEt	0.06	1.58	4.87	1.63	3.45	0.48	0.38
95	$C(CF_3)_3$	0.07	2.08	4.11	3.13	3.64	0.55	0.55
96	NH—Et	0.08	1.50	4.83	1.35	3.42	−0.61	−0.24
97	$NHC{=}O(CF_3)$	0.08	1.43	5.62	1.79	3.61	0.12	0.30
98	$SC{=}O(CH_3)$	0.10	1.84	5.11	1.70	4.01	0.44	0.39
99	CF_3	0.10	0.50	3.30	1.99	2.61	0.54	0.43
100	OCH_2F	0.10	0.72	4.57	1.35	3.07	0.02	0.20
101	$CH{=}CHNO_2$(TR)	0.11	1.64	4.29	1.60	4.78	0.26	0.32
102	CH_2F	0.13	0.54	3.30	1.52	2.61	0.11	0.12
103	F	0.14	0.09	2.65	1.35	1.35	0.06	0.34
104	$C(OMe)_3$	0.14	2.48	4.78	2.56	4.29	−0.04	−0.03
105	$SECF_3$	0.15	1.63	4.50	1.85	4.09	0.45	0.44
106	$NHC{=}O(OEt)$	0.17	2.12	7.25	1.35	3.92	−0.15	0.11
107	CH_2Cl	0.17	1.05	3.89	1.52	3.46	0.12	0.11
108	$N(CH_3)_2$	0.18	1.56	3.53	1.35	3.08	−0.83	−0.16

Table 1.4 (Continued)

No.	Substituent	Pi	MR	L	B1	B5	S-P	S-M
109	CHF_2	0.21	0.52	3.30	1.71	2.61	0.32	0.29
110	$CCCF_3$	0.22	1.41	5.90	1.99	2.61	0.51	0.41
111	$SO_2C_6H_5$	0.27	3.32	5.86	2.03	6.02	0.68	0.62
112	$COCH(CH_3)_2$	0.29	1.98	4.84	1.99	4.08	0.47	0.38
113	OCHF,	0.31	0.79	3.98	1.35	3.61	0.18	0.31
114	$CH_2SO_2CF_3$	0.33	1.75	5.35	1.52	4.07	0.31	0.29
115	$C(NO_2)(CH_3)_2$	0.33	2.06	4.59	2.58	3.72	0.20	0.18
116	$P(O)(OPR)_2$	0.35	4.05	7.07	2.52	6.90	0.50	0.38
117	$CH_2S{=}O(CF_3)$	0.37	1.90	5.35	1.52	4.07	0.24	0.25
118	OCH_2CH_3	0.38	1.25	4.80	1.35	3.36	−0.24	0.10
119	SH	0.39	0.92	3.47	1.70	2.33	0.15	0.25
120	$N{=}NCF_3$	0.40	1.39	5.45	1.70	3.48	0.68	0.56
121	CCH	0.40	0.96	4.66	1.60	1.60	0.23	0.21
122	$N{=}CCl_2$	0.41	1.84	5.65	1.70	4.54	0.13	0.21
123	SCCH	0.41	1.62	4.08	1.70	4.85	0.19	0.26
124	SCN	0.41	1.34	4.08	1.70	4.45	0.52	0.51
125	$P(CH_3)_2$	0.44	2.12	3.88	2.00	3.32	0.06	0.03
126	$NHSO_2C_6H_5$	0.45	3.79	8.24	1.35	3.72	0.01	0.16
127	$SO_2NHC_6H_5$	0.45	3.78	8.24	2.03	4.50	0.65	0.56
128	CH_2CF_3	0.45	0.97	4.70	1.52	3.70	0.09	0.12
129	NNN	0.46	1.02	4.62	1.50	4.18	0.08	0.37
130	NNN	0.46	1.02	4.62	1.50	4.18	0.08	0.37
131	4-Pyridyl	0.46	2.30	5.92	1.71	3.11	0.44	0.27
132	$N{=}NN(CH_3)_2$	0.46	2.09	5.68	1.77	3.90	0.44	0.27
133	$C{=}O(NHC_6H_5)$	0.49	3.54	8.24	1.63	4.85	−0.03	−0.05
134	2-Pyridyl	0.50	2.30	6.28	1.71	3.11	0.41	0.23
135	$OCH_2CH{=}CH_2$	0.51	1.61	6.22	1.35	4.42	0.17	0.33
136	$C{=}O(OEt)$	0.51	1.75	5.95	1.64	4.41	−0.25	0.09
137	$S{=}O(CF_3)$	0.53	1.31	4.70	1.40	3.70	0.45	0.37
138	$CHOHC_6H_5$	0.54	3.15	4.62	1.73	6.02	0.69	0.63
139	OCH_2Cl	0.54	1.20	5.44	1.35	3.13	−0.03	0.00
140	$SO_2(CF_3)$	0.55	1.29	4.70	2.03	3.70	0.08	0.25
141	CH_3	0.56	0.57	2.87	1.52	2.04	0.96	0.83
142	SCH_3	0.61	1.38	4.30	1.70	3.26	−0.17	−0.07
143	$SC{=}O(CF_3)$	0.66	1.82	5.55	1.70	4.51	0.00	0.15
144	$COC(CH_3)_3$	0.69	2.44	4.87	1.87	4.42	0.46	0.48
145	$CH{=}NC_6H_5$	0.69	3.30	8.50	1.70	4.07	0.32	0.27
146	$P{=}O(C_6H_5)_2$	0.70	5.93	5.40	2.68	6.19	0.42	0.35
147	Cl	0.71	0.60	3.52	1.80	1.80	.530	.380
148	$N{=}CHC_6H_5$	0.72	3.30	8.40	1.70	4.65	0.23	0.37
149	$SeCH_3$	0.74	1.70	4.52	1.85	3.63	−0.55	−0.08
150	SCH_2F	0.74	1.34	4.89	1.70	3.41	0.00	0.10
151	$OCH{=}CH_2$	0.75	1.14	4.98	1.35	3.65	0.20	0.23
152	CH_2Br	0.79	1.34	4.09	1.52	3.75	−0.09	0.21
153	$CCCH_3$	0.81	1.41	5.47	1.60	2.04	0.14	0.12
154	CH=CH2	0.82	1.10	4.29	1.60	3.09	0.03	0.21
155	Br	0.86	0.89	3.82	1.95	1.95	−0.16	−0.08
156	$NHSO_2CF_3$	0.93	1.75	5.26	1.35	4.00	0.23	0.39
157	$OSO_2C_6H_5$	0.93	3.67	8.20	1.35	3.64	0.39	0.44
158	1-Pyrryl	0.95	1.95	5.44	1.71	3.12	0.33	0.36
159	$N(CH_3)SO_2CF_3$	1.00	2.28	5.26	1.54	4.00	0.37	0.47
160	$SCHF_2$	1.02	1.38	4.30	1.70	3.94	0.44	0.46
161	CH_2CH_3	1.02	1.03	4.11	1.52	3.17	0.37	0.33
162	OCF_3	1.04	0.79	4.57	1.35	3.61	−0.15	−0.07

Table *1.4 (Continued)*

No.	Substituent	Pi	MR	L	B1	B5	S-P	S-M
163	$OCH_2CH_2CH_3$	1.05	1.71	6.05	1.35	4.42	0.35	0.38
164	$C{=}O(C_6H_5)$	1.05	3.03	4.57	1.92	5.98	−0.25	0.10
165	$NHCO_2C_4H_9$	1.07	3.05	9.50	1.45	5.05	0.43	0.34
166	S—Et	1.07	1.84	5.16	1.70	3.97	−0.05	0.06
167	$N(CF_3)_2$	1.08	1.43	4.01	1.52	3.58	0.03	0.18
168	$CHCl_2$	1.09	1.53	3.89	1.88	3.46	0.53	0.40
169	$CH_2CH{=}CH_2$	1.10	1.45	5.11	1.52	3.78	0.32	0.31
170	CH_2I	1.10	1.86	4.36	1.52	4.15	−0.14	−0.11
171	NH—Bu	1.10	2.43	6.88	1.35	4.87	0.11	0.10
172	$CClF_2$	1.11	1.07	3.89	1.99	3.46	−0.51	−0.34
173	I	1.12	1.39	4.23	2.15	2.15	0.46	0.42
174	Cyclopropyl	1.14	1.35	4.14	1.55	3.24	0.18	0.35
175	$C(CH_3){=}CH_2$	1.14	1.56	4.29	1.73	3.11	−0.21	−0.07
176	NCS	1.15	1.72	4.29	1.50	4.24	0.05	0.09
177	$SCH_2CH{=}CH_2$	1.15	2.26	6.42	1.70	5.02	0.38	0.48
178	$N(Et)_2$	1.18	2.49	4.83	1.35	4.39	0.12	0.19
179	OSO_2CF_3	1.23	1.45	5.23	1.35	3.24	−0.72	−0.23
180	SF_5	1.23	0.99	4.65	2.47	2.92	0.53	0.56
181	$OCHCl_2$	1.26	1.69	3.98	1.35	4.41	0.68	0.61
182	CF_2CF_3	1.26	0.92	4.11	1.99	3.64	0.26	0.38
183	$C(OH)(CF_3)_2$	1.28	1.52	4.11	2.61	3.64	0.52	0.47
184	$SCH{=}CH_2$	1.29	1.77	5.33	1.70	4.23	0.30	0.29
185	NHC_6H_5	1.37	3.00	4.53	1.35	5.95	0.20	0.26
186	$SCH(CH_3)_2$	1.41	2.41	5.16	1.70	4.41	−0.56	−0.02
187	SCF_3	1.44	1.38	4.89	1.70	3.94	0.07	0.23
188	$OC{=}O(C_6H_5)$	1.46	3.23	8.15	1.64	4.40	0.50	0.40
189	$COOC_6H_5$	1.46	3.02	8.13	1.94	3.50	0.13	0.21
190	Cyclobutyl	1.51	1.79	4.77	1.77	3.82	0.44	0.37
191	O—Bu	1.52	2.17	6.86	1.35	4.79	−0.14	−0.05
192	$CH(CH_3)_2$	1.53	1.50	4.11	1.90	3.17	−0.32	0.10
193	$CHBr_2$	1.53	1.68	4.09	1.92	3.75	−0.15	−0.04
194	Pr	1.55	1.50	4.92	1.52	3.49	0.32	0.31
195	$C(F)(CF_3)_2$	1.56	1.34	4.11	2.45	3.64	−0.13	−0.06
196	$C_6H_4(NO_2)$-p	1.64	3.17	7.66	1.71	3.11	0.05	0.09
197	$CH_2OC_6H_5$	1.66	3.22	8.19	1.52	3.53	0.13	0.10
198	$N{=}NC_6H_5$	1.69	3.13	8.43	1.70	4.31	0.07	0.06
199	$SO_2CF_2CF_3$	1.73	1.97	5.35	2.03	4.07	0.39	0.32
200	$CF_2CF_2CF_2CF_3$	1.74	1.77	6.76	1.99	5.05	1.08	0.92
201	1-Cyclopentenyl	1.77	2.21	5.24	1.91	3.08	0.52	0.47
202	OCF_2CHF_2	1.79	1.08	5.23	1.35	3.94	−0.05	−0.06
203	$C_6H_4(OCH_3)$-p	1.82	3.17	7.71	1.80	3.11	0.25	0.34
204	CH_2SCF_3	1.83	1.76	5.82	1.52	4.10	−0.08	0.05
205	C_6H_5	1.96	2.54	6.28	1.71	3.11	0.04	0.01
206	$C(CH_3)_3$	1.98	1.96	4.11	2.60	3.17	−0.01	0.06
207	CCl_3	1.99	2.01	3.89	2.64	3.46	−0.20	−0.10
208	$CH_2Si(CH_3)_3$	2.00	2.96	5.39	1.52	4.75	0.46	0.40
209	$CH_2C_6H_5$	2.01	3.00	4.62	1.52	6.02	−0.21	−0.16
210	$CH(CH_3)(Et)$	2.04	1.96	4.92	1.90	3.49	−0.09	−0.08
211	C_6H_4F-p	2.04	2.53	6.87	1.71	3.11	−0.12	−0.08
212	OC_5H_{11}	2.05	2.63	8.11	1.35	5.81	0.06	0.12
213	$N(C_3H_7)_2$	2.08	3.24	6.07	1.35	5.50	−0.34	0.10
214	OC_6H_5	2.08	2.77	4.51	1.35	5.89	−0.93	−0.26
215	$C_6H_4N(CH_3)_2$-p	2.10	3.99	7.75	1.79	3.11	−0.03	0.25
216	Bu	2.13	1.96	6.17	1.52	4.54	−0.56	−0.06

Table 1.4 (Continued)

No.	Substituent	Pi	MR	L	B1	B5	S-P	S-M
217	Cyclopentyl	2.14	2.20	4.90	1.90	4.09	0.28	0.35
218	CHI_3	2.15	3.15	4.36	1.95	4.15	−0.14	−0.05
219	SC_6H_5	2.32	3.43	4.57	1.70	6.42	0.26	0.26
220	1-Cyclohexenyl	2.33	2.67	6.16	2.23	3.30	0.07	0.23
221	$OCCl_3$	2.36	2.18	5.44	1.35	4.41	−0.08	−0.10
222	$C(Et)(CH_3)_2$	2.37	2.42	4.92	2.60	3.49	0.35	0.43
223	$CH_2C(CH_3)_3$	2.37	2.42	4.89	1.52	4.18	−0.18	−0.06
224	$SC_6H_4NO_2$-p	2.39	4.11	4.92	1.70	7.86	−0.17	−0.05
225	SCF_2CHF_2	2.43	1.84	5.60	1.70	4.55	0.24	0.32
226	C_6H_4Cl-p	2.61	3.04	7.74	1.80	3.11	−0.07	−0.04
227	C_6F_5	2.62	2.40	6.87	1.71	3.67	0.12	0.15
228	C_5H_{11}	2.63	2.42	6.97	1.52	4.94	0.27	0.26
229	CCC_6H_5	2.65	3.32	8.88	1.71	3.11	−0.15	−0.08
230	CBr_3	2.65	2.88	4.09	2.86	3.75	0.16	0.14
231	EtC_6H_5	2.66	3.47	8.33	1.52	3.58	0.29	0.28
232	$C_6H_4(CH_3)$-p	2.69	3.00	7.09	1.84	3.11	−0.12	−0.07
233	C_6H_4I-p	3.02	3.91	8.45	2.15	3.11	−0.03	0.06
234	C_6H_4I-m	3.02	3.91	6.72	1.84	5.15	0.12	0.15
235	1-Adamantyl	3.37	4.03	6.17	3.16	3.49	−0.15	−0.05
236	$C(Et)_3$	3.42	3.36	4.92	2.94	4.18	0.10	0.14
237	$CH(C_6H_5)_2$	3.52	5.43	5.15	2.01	6.02	0.06	0.13
238	$N(C_6H_5)_2$	3.61	5.50	5.77	1.35	5.95	0.01	0.08
239	Heptyl	3.69	3.36	9.03	1.52	6.39	−0.13	−0.12
240	$C(SCF_3)_3$	4.17	4.40	5.82	3.32	5.00	−0.20	−0.07
241	C_6Cl_5	4.96	4.95	7.74	1.81	4.48	−0.05	−0.03

methods are based on molecular fragments, atomic contributions, or computer-identified fragments (1, 106, 107, 144–147). Whole-molecule approaches use molecular properties or spatial properties to predict log P values (148–150). They run on different platforms (e.g., Mac, PC, Unix, VAX, etc.) and use different calculation procedures. An extensive, recent review by Mannhold and van de Waterbeemd addresses the advantages and limitations of the various approaches (143). Statistical parameters yield some insight as to the effectiveness of such programs.

Recent attempts to compute log P calculations have resulted in the development of solvatochromic parameters (151, 152). This approach was proposed by Kamlet et al. and focused on molecular properties. In its simplest form it can be expressed as follows:

$$\text{Log } P_{\text{oct}} = aV + b\pi^* + c\beta_{\text{H}} + d\alpha_{\text{H}} + e \quad (1.49)$$

V is a solute volume term; π^* represents the solute polarizability; β_{H} and a, are measures of hydrogen bond acceptor strength and hydrogen bond donor strength, respectively; and e is the intercept. An extension of this model has been formulated by Abraham and used by researchers to refine molecular descriptors and characterize hydrophobicity scales (153–156).

3.3 Steric Parameters

The quantitation of steric effects is complex at best and challenging in all other situations, particularly at the molecular level. An added level of confusion comes into play when attempts are made to delineate size and shape. Nevertheless, sterics are of overwhelming importance in ligand-receptor interactions as well as in transport phenomena in cellular systems. The first steric parameter to be quantified and used in QSAR studies was Taft's E_{S} constant (157). E_{S} is defined as

$$E_{\text{S}} = \log(k_{\text{X}}/k_{\text{H}})_{\text{A}} \quad (1.50)$$

where k_{X} and k_{H} represent the rates of acid hydrolysis of esters, XCH_2COOR and CH_3COOR,

respectively. To correct for hyperconjugation in the a-hydrogens of the acetate moiety, Hancock devised a correction on E_S such that

$$E_S^C = E_S + 0.306(n - 3) \qquad (1.51)$$

In Equation 1.51, n represents the number of a-hydrogens and 0.306 is a constant derived from molecular orbital calculations (158). Unfortunately, the limited availability of E_S and E_S^C values for a great number of substituents precludes their usage in QSAR studies. Charton demonstrated a strong correlation between E_S and van der Waals radii, which led to his development of the upsilon parameter υ_X (159).

$$\upsilon_X = r_X - r_H = r_X - 1.20 \qquad (1.52)$$

where r_X and r_H are the minimum van der Waals radii of the substituent and hydrogen, respectively. Extension of this approach from symmetrical substituents to nonsymmetrical substituents must be handled with caution.

One of the most widely used steric parameters is molar refraction (MR), which has been aptly described as a "chameleon" parameter by Tute (160). Although it is generally considered to be a crude measure of overall bulk, it does incorporate a polarizability component that may describe cohesion and is related to London dispersion forces as follows: $MR = 4\pi N\alpha/3$, where N is Avogadro's number and a is the polarizability of the molecule. It contains no information on shape. MR is also defined by the Lorentz-Lorenz equation:

$$MR = [(n^2 - 1)/(n^2 + 2)] \times (MW/density) \qquad (1.53)$$

MR is generally scaled by 0.1 and used in biological QSAR, where intermolecular effects are of primary importance. The refractive index of the molecule is represented by n. With alkyl substituents, there is a high degree of collinearity with hydrophobicity; hence, care must be taken in the QSAR analysis of such derivatives. The MR descriptor does not distinguish shape; thus the MR value for amyl ($—CH_2CH_2CH_2CH_2CH_3$) is the same as that for [$—C(Et)(CH_3)_2$]: 2.42. The coefficients with MR terms challenge interpretation, although extensive experience with this parameter suggests that a negative coefficient implies steric hindrance at that site and a positive coefficient attests to either dipolar interactions in that vicinity or anchoring of a ligand in an opportune position for interaction (161).

The failure of the MR descriptor to adequately address three-dimensional shape issues led to Verloop's development of STERIMOL parameters (162), which define the steric constraints of *a* given substituent along several fixed axes. Five parameters were deemed necessary to define shape: L, B1, B2, B3, and B4. L represents the length of a substituent along the axis of a bond between the parent molecule and the substituent; B1 to B4 represent four different width parameters. However, the high degree of collinearity between B1, B2, and B3 and the large number of training set members needed to establish the statistical validity of this group of parameters led to their demise in QSAR studies. Verloop subsequently established the adequacy of just three parameters for QSAR analysis: a slightly modified length L, a minimum width B1, and a maximum width B5 that is orthogonal to L (163). The use of these insightful parameters have done much to enhance correlations with biological activities. Recent analysis in our laboratory has established that in many cases, B1 alone is superior to Taft's E_S and a combination of B1 and B5 can adequately replace E_S (164).

Molecular weight (MW) terms have also been used as descriptors, particularly in cellular systems, or in distribution/transport studies where diffusion is the mode of operation. According to the Einstein-Sutherland equation, molecular weight affects the diffusion rate. The Log MW term has been used extensively in some studies (159–161)and an example of such usage is given below. In correlating permeability (Perm) of noneledrolytes through chara cells, Lien et al. obtained the following QSAR (168):

$$\begin{aligned}\text{Log Perm} &= 0.889 \log P^* - 1.544 \log \text{MW} \\ &\quad - 0.144 H_b + 4.653 \\ n &= 30, \quad r^2 = 0.899, \\ s &= 0.322, \quad F = 77.39\end{aligned} \tag{1.54}$$

In QSAR 54, Log P^* represents the olive oil/water partition coefficient, MW is the molecular weight of the solute and defines its size, and H_b is a crude approximation of the total number of hydrogen bonds for each molecule. The molecular weight descriptor has also been an omnipresent variable in QSAR studies pertaining to cross-resistance of various drugs in multidrug-resistant cell lines (169). $\sqrt[3]{\text{MW}}$ was used because it most closely approximates the size (radii) of the drugs involved in the study and their interactions with GP-170. See QSAR 1.55.

$$\begin{aligned}\text{Log CR} &= 0.70 \sqrt[3]{\text{MW}} \\ &\quad - 1.01 \log(\beta \cdot 10^{\sqrt[3]{\text{MW}}} + 1) \\ &\quad - 0.10 \log P + 0.381 \\ &\quad - 3.08 \\ n &= 40, \quad r^2 = 0.794, \quad s = 0.344 \\ \log \beta &= -6.851 \quad \text{optimum } \sqrt[3]{\text{MW}} = 7.21\end{aligned} \tag{1.55}$$

3.4 Other Variables and Variable Selection

Indicator variables (I) are often used to highlight a structural feature present in some of the molecules in a data set that confers unusual activity or lack of it to these particular members. Their use could be beneficial in cases where the data set is heterogeneous and includes large numbers of members with unusual features that may or may not impact a biological response. QSAR for the inhibition of trypsin by X-benzamidines used indicator variables to denote the presence of unusual features such as positional isomers and vinyl/carbonyl-containing substituents (170). A recent study on the inhibition of lipoxygenase catalyzed production of leukotriene B4 and 5-hydroxyeicosatetraenoic from arachidonic acid in guinea pig leukocytes by X-vinyl catechols led to the development of the following QSAR (171):

$$\begin{aligned}\text{Log } 1/C &= 0.49(\pm 0.11)\log P \\ &\quad - 0.75(\pm 0.22)\log(\beta \cdot 10^{\log P} + 1) \\ &\quad - 0.62(\pm 0.18)\text{D2} \\ &\quad - 1.13(\pm 0.20)\text{D3} + 5.50(\pm 0.33) \\ n &= 51, \quad r^2 = 0.801, \quad s = 0.269, \\ \text{Log } P_o &= 4.61(\pm 0.49) \quad \text{Log } \beta = -4.33\end{aligned} \tag{1.56}$$

The indicator variables are D2 and D3; for simple X-catechols, D2 = 1 and for X-naphthalene diols, D3 = 1. The negative coefficients with both terms (D2 and D3) underscore the detrimental effects of these structural features in these inhibitors. Thus, discontinuities in the structural features of the molecules of this data set are accounted for by the use of indicator variables. An indicator variable may be visualized graphically as a constant that adjusts two parallel lines so that they are superimposable. The use of indicator variables in QSAR analysis is also described in the following example. An analysis of a comprehensive set of nitroaromatic and heteroaromatic compounds that induced mutagenesis in TA98 cells was conducted by Debnath et al., and QSAR 1.57 was formulated (172).

$$\begin{aligned}\text{Log TA98} &= 0.65(\pm 0.16)\log P \\ &\quad - 2.90(\pm 0.59)\log(\beta \cdot 10^{\log P} + 1) \\ &\quad - 1.38(\pm 0.25) E_{\text{LUMO}} \\ &\quad + 1.88(\pm 0.39) I_1 - 2.89(\pm 0.81) I_a \\ &\quad - 4.15(\pm 0.58) \\ n &= 188, \quad r^2 = 0.810, \quad s = 0.886, \\ \text{Log } P_o &= 4.93(\pm 0.35) \quad \text{Log } \beta = -5.48\end{aligned} \tag{1.57}$$

TA98 represents the number of revertants per nanomole of nitro compound. E_{LUMO} is the energy of the lowest unoccupied molecular or-

bital and I_a is an indicator variable that signifies the presence of an acenthrylene ring in the mutagens. **I,** is also an indicator variable that pertains to the number of fused rings in the data set. It acquires a value of 1 for all congeners containing three or more fused rings and a value of zero for those containing one or two fused rings (**e.g.**, naphthalene, benzene). Thus, the greater the number of fused rings, the greater the mutagenicity of the nitro congeners. The E_{LUMO} term indicates that the lower the energy of the LUMO, the more potent the mutagen. In this QSAR the combination of indicator variables affords a mixed blessing. One variable helps to enhance activity, whereas the other leads to a decrease in mutagenicity of the acenthrylene congeners. In both these QSAR, Kubinyi's bilinear model is used (21). See Section 4.2 for a description of this approach.

3.5 Molecular Structure Descriptors

These are truly structural descriptors because they are based only on the two-dimensional representation of a chemical structure. The most widely known descriptors are those that were originally proposed by Randic (173) and extensively developed by Kier and Hall (27). The strength of this approach is that the required information is embedded in the hydrogen-suppressed framework and thus no experimental measurements are needed to define molecular connectivity indices. For each bond the C_k term is calculated. The summation of these terms then leads to the derivation of X, the molecular connectivity index for the molecule.

$$C_k = (\delta_i \delta_j)^{-0.5} \quad \text{where} \quad \delta = \sigma - h \qquad (1.58)$$

S is the count of formally bonded carbons and ***h*** is the number of bonds to hydrogen atoms.

$$^1\mathrm{X} = \sum C_k = \sum (\delta_i \delta_j)_k^{-0.5} \qquad (1.59)$$

^{1}**X** is the first bond order because it considers only individual bonds. Higher molecular connectivity indices encode more complex attributes of molecular structure by considering longer paths. Thus, ^{2}**X** and ^{3}**X** account for all two-bond paths and three-bond paths, respectively, in a molecule. To correct for differences in valence, Kier and Hall proposed a valence delta (δ^{v}) term to calculate valence connectivity indices (175).

Molecular connectivity indices have been shown to be closely related to many physicochemical parameters such as boiling points, molar refraction, polarizability, and partition coefficients (174, 176). Ten years ago, the E-State index was developed to define an atom- or group-centered numerical code to represent molecular structure (28). The E-State was established as a composite index encoding both electronic and steric properties of atoms in molecules. It reflects an atom's electronegativity, the electronegativity of proximal and distal atoms, and topological state. Extensions of this method include the HE-State, atom-type E-State, and the polarity index ***Q***. Log P showed a strong correlation with the ***Q*** index of a small set ($n = 21$) of miscellaneous compounds (28). Various models using electrotopological indices have been developed to delineate a variety of biological responses (177–179). Some criticism has been leveled at this approach (180, 181). Chance correlations are always a problem when dealing with such a wide array of descriptors. The physicochemical interpretation of the meaning of these descriptors is not transparent, although attempts have been made to address this issue (27).

4 QUANTITATIVE MODELS

4.1 Linear Models

The correlation of biological activity with physicochemical properties is often termed an ***extrathermodynamic relationship.*** Because it follows in the line of Hammett and Taft equations that correlate thermodynamic and related parameters, it is appropriately labeled. The Hammett equation represents relationships between the logarithms of rate or equilibrium constants and substituent constants. The linearity of many of these relationships led to their designation as linear free energy relationships. The Hansch approach represents an extension of the Harnmett equation from physical organic systems to a biological milieu. It should be noted that the simplicity

of the approach belies the tremendous complexity of the intermolecular interactions at play in the overall biological response.

Biological systems are a complex mix of heterogeneous phases. Drug molecules usually traverse many of these phases to get from the site of administration to the eventual site of action. Along this random-walk process, they perturb many other cellular components such as organelles, lipids, proteins, and so forth. These interactions are complex and vastly different from organic reactions in test tubes, even though the eventual interaction with a receptor may be chemical or physicochemical in nature. Thus, depending on the biological system involved—isolated receptor, cell, or whole animal—one expects the response to be multifactorial and complex. The overall process, particularly in *vitro* or in *vivo*, studies a mix of equilibrium and rate processes, a situation that defies easy separation and delineation.

Meyer and Overton were the first to attempt to get a grasp on biological responses by noting the relationship between oil/water partition coefficients and their narcotic activity. Ferguson recognized that equitoxic concentrations of small organic molecules was markedly influenced by their phase distribution between the biophase and exobiophase. This concept was generalized in the form of Equation 1.60 and extended by Fujita to Equation 1.61 (182,183).

$$C = kA^m \tag{1.60}$$

$$\text{Log } 1/C = \text{m Log}(1/A) + \text{constant} \tag{1.61}$$

C represents the equipotent concentration, k and m are constants for a particular system, and A is a physicochemical constant representative of phase distribution equilibria such as aqueous solubility, oil/water partition coefficient, and vapor pressure. In examining a large and diverse number of biological systems, Hansch and coworkers defined a relationship (Equation 1.62) that expressed biological activity as a function of physicochemical parameters (e.g., partition coefficients of organic molecules) (19).

$$\text{Log } 1/C = a \log P + b \tag{1.62}$$

Model systems have been devised to elucidate the mode of interactions of chemicals with biological entities. Examples of linear models pertaining to nonspecific toxicity are described. The effects of a series of alcohols (ROH) have been routinely studied in many model and biological systems. See QSAR 1.63–1.67.

4.1.1 Penetration of ROH into Phosphatidylcholine Monolayers (184)

$$\text{Log } 1/C = 0.87(\pm 0.01)\log P + 0.66(\pm 0.01) \tag{1.63}$$
$$n = 4, \quad r^2 = 0.998, \quad s = 0.002$$

4.1.2 Changes in EPR Signal of Labeled Ghost Membranes by ROH (185)

$$\text{Log } 1/C = 0.93(\pm 0.09)\log P - 0.41(\pm 0.16) \tag{1.64}$$
$$n = 6, \quad r^2 = 0.996, \quad s = 0.092$$

4.1.3 Induction of Narcosis in Rabbits by ROH (184)

$$\text{Log } 1/C = 0.72(\pm 0.16)\log P + 1.35(\pm 0.12) \tag{1.65}$$
$$n = 11, \quad r^2 = 0.924, \quad s = 0.142$$

4.1.4 Inhibition of Bacterial Luminescence by ROH (185)

$$\text{Log } 1/C = 1.10(\pm 0.07)\log P + 0.16(\pm 0.12) \tag{1.66}$$
$$n = 8, \quad r^2 = 0.996, \quad s = 0.103$$

4.1.5 Inhibition of Growth of Tetrahymena pyriformis by ROH (76, 186)

$$\text{Log } 1/C = 0.82(\pm 0.04)\text{Clog } P + 0.89(\pm 0.10) \tag{1.67}$$
$$n = 34, \quad r^2 = 0.982, \quad s = 0.173$$

In all cases, there is a strong dependency on

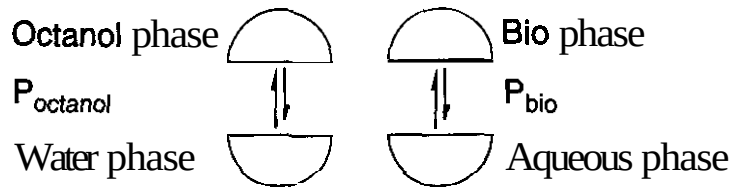


Figure 1.1. Log $P_{octanol}$ mirrors Log P_{bio}.

log P_{oct} because all these processes involve transport of alcohols through membranes. The low intercepts speak to the nonspecific nature of the alcohol-mediated toxic interaction. An equilibrium-pseudoequilibrium modeled by log P can be defined as shown in Fig. 1.1.

The Hammett-type relationship for this conceptual idea of distribution is

$$\text{Log } P_{bio} = \text{a} \cdot \log P_{octanol} + \text{b} \qquad (1.68)$$

This postulate assumes that steric, hydrophobic, electronic, and hydrogen bonding factors that affect partitioning in the biophase are handled by the octanol/water system. Given that the biological response (log $1/C$) is proportional to log P_{bio}, then it follows that

$$\text{Log } 1/C = \text{a} \cdot \log P_{octanol} + \text{constant} \qquad (1.69)$$

Hansch and coworkers have amply demonstrated that Equation 1.69 applies not only to systems at or near phase distribution equilibrium but also to systems removed from equilibrium (184, 185).

4.2 Nonlinear Models

Extensive studies on development of linear models led Hansch and coworkers to note that a breakdown in the linear relationship occurred when a greater range in hydrophobicity was assessed with particular emphasis placed on test molecules at extreme ends of the hydrophobicity range. Thus, Hansch et al. suggested that the compounds could be involved in a "random-walk" process: low hydrophobic molecules had a tendency to remain in the first aqueous compartment, whereas highly hydrophobic analogs sequestered in the first lipoidal phase that they encountered. This led to the formulation of a parabolic equation, relating biological activity and hydrophobicity (187).

$$\text{Log } 1/C = -a(\log P)^2 + \text{b} \cdot \log \text{P} + \text{constant} \qquad (1.70)$$

In the random-walk process, the compounds partition in and out of various compartments and interact with myriad biological components in the process. To deal with this conundrum, Hansch proposed a general, comprehensive equation for QSAR 1.71 (188).

$$\text{Log } 1/C = -a(\log P)^2 + \text{b} \cdot \log \text{P} + \rho\sigma + \delta E_S + \text{constant} \qquad (1.71)$$

The optimum value of log P for a given system is log P_O and it is highly influenced by the number of hydrophobic barriers a drug encounters in its walk to its site of action. Hansch and Clayton formulated the following parabolic model to elucidate the narcotic action of alcohols on tadpoles (189).

4.2.1 Narcotic Action of ROH on Tadpoles

$$\text{Log}1/C = 1.38(\pm 0.34)\log P - 0.08(\pm 0.07)(\text{Log}P)^2 + 0.52(\pm 0.34) \qquad (1.72)$$

$$n = 10, \quad r^2 = 0.990, \quad s = 0.210,$$

$$\text{Log } P_O = 8.69(5.78 - 43.43)$$

This is an example of nonspecific toxicity where the last step probably involves partitioning into a hydrophobic membrane. Log P_O represents the optimal hydrophobicity (as defined by log P) that elicits a maximal biological response.

Despite the success of the parabolic equation, there are a number of worrisome limitations. This approach forces the data into a symmetrical parabola, with the result that there are usually deviations between the experimental and parabola-calculated data. Second, the ascending slope is curved and inconsistent with the observed linear data. Thus, the slope of a linear model cannot be compared to the curved slope of the parabola. In 1973 Franke devised a sophisticated, empirical

model consisting of a linear ascending part and a parabolic part (190). See Equations 1.73 and 1.74.

$$\text{Log } 1/C = a \cdot \log P + \text{c} \quad (\text{if } \log \text{P} < \log P_{\text{x}}) \tag{1.73}$$

$$\text{Log } 1/C = -a(\log P)^2 + \text{b} \cdot \log \text{P} + \text{c} \quad (\text{if } \log \text{P} > \log P_{\text{x}}) \tag{1.74}$$

The binding of drugs to proteins is linearly dependent on hydrophobicity up to a limited value, log P_{x}, after which steric hindrance causes the linear dependency to alter to a nonlinear one. The major limitation of this approach involves the inclusion of highly hydrophobic congeners that tend to cause systematic deviations between experimental and predicted values.

Another cutoff model, which deals with nonlinearity in biological systems, is one defined by McFarland (191). It attempts to elucidate the dependency of drug transport on hydrophobicity in multicompartment models. McFarland addressed the probability of drug molecules traversing several aqueous lipid barriers from the first aqueous compartment to a distant, final aqueous compartment. The probability $P_{0,n}$ of a drug molecule to access the final compartment n of a biological system was used to define the drug concentration in this compartment.

$$\text{Log } C_{\text{R}} = a \cdot \log P - 2a \cdot \log(P + 1) + \text{constant} \tag{1.75}$$

The ascending and descending slopes are equal (=1) and linear. However, a major drawback of this model is that it forces the activity curves to maximize at log P = 0. These studies were extended by Kubinyi, who developed the elegant and powerful bilinear model, which is superior to the parabolic model and is extensively used in QSAR studies (192).

$$\text{Log } 1/C = a \cdot \log P - b \cdot \log(\beta \cdot P + 1) + \text{constant} \tag{1.76}$$

where β is the ratio of the volumes of the organic phase and the aqueous phase. An important feature of this model lies in the symmetry of the curves. For aqueous phases of this model system, symmetrical curves with linear ascending and descending sides (like a teepee) and a limited parabolic section around the hydrophobicity optimum are generated. Unsymmetrical curves arise for the lipid phases. It is highly compatible with the linear model and allows for quick comparisons of the ascending slopes. It can also be used with other parameters such as MR and σ, where it appears to pinpoint a change in mechanism similar to the breaks in linearity of the Hammett equation. The following example of the bilinear model reveals the symmetrical nature of the curve.

4.2.2 Induction of Ataxia in Rats by ROH

$$\begin{aligned}\text{Log } 1/C = {} & 0.77(\pm 0.10)\log P \\ & - 1.53(\pm 0.12)\log(\beta \cdot P + 1) \\ & + 1.68(\pm 0.12) \\ & n = 35, \quad r^2 = 0.887, \\ & s = 0.165, \quad \log P_{\text{O}} = 2.0\end{aligned} \tag{1.77}$$

The bilinear model has been used to model biological interactions in isolated receptor systems and in adsorption, metabolism, elimination, and toxicity studies, although it has a few limitations. These include the need for at least 15 data points (because of the presence of the additional disposable parameter β and data points beyond optimum Log P. If the range in values for the dependent variable is limited, unreasonable slopes are obtained.

4.3 Free-Wilson Approach

The Free-Wilson approach is truly a structure-activity-based methodology because it incorporates the contributions made by various structural fragments to the overall biological activity (22, 193, 194). It is represented by Equation 1.78.

$$\text{BA}_i = \sum_j a_j \text{X}_{ij} + \mu \tag{1.78}$$

Indicator variables are used to denote the presence or absence of a particular structure feature.

Like classical QSAR, this *de novo* approach assumes that substituent effects are additive and constant. BA is the biological activity; X_j is the jth substituent, which carries a value 1 if present, 0 if absent. The term a_j represents the contribution of the jth substituent to biological activity and p is the overall average activity. The summation of all activity contributions at each position must equal zero. The series of linear equations that are formulated are solved by linear regression analysis. It is necessary for each substituent to appear more than once at a position in different combinations with substituents at other positions.

There are certain advantages to the Free-Wilson method that have been addressed (193–195). Any type of quantitative biological data can be subject to such analysis. There is no need for any physicochemical constants. The molecules of a series may be structurally dissected in any way and multiple sites of substitution are necessary and easily accommodated (196). Limitations include the large number of molecules with varying substituent combinations that are needed for this analysis and the inability of the system to handle nonlinearity of the dependency of activity on substituent properties. Intramolecular interactions between the substituent are not handled very well, although special treatments can be used to accommodate proximal effects. Extrapolation outside of the substituents used in the study is not feasible. Another problem inherent with this approach is that usually a large number of variables is required to describe a smaller number of compounds, which creates a statistical faux pas. Fujita and Ban modified this approach in two important ways (23). They expressed the biological activity on a logarithmic scale, to bring it into line with the extrathermodynamic approach, as seen in the following equation:

$$\text{Log } X_C = \sum a_i X_i + \mu \qquad (1.79)$$

This allowed the derived substituent constants to be compared with other free energy-related parameters. The overall average intercept u took on a new look, as it were, akin to an intercept in other QSAR analyses.

Recent analyses of a Free-Wilson type have included the in *vitro* inhibitory activity of a series of heterocyclic compounds against K. pneumonia (197). Other applications of the Free-Wilson approach have included studies on the antimycobacterial activity of 4-alkylthiobenzanilides, the antibacterial activity of fluoronapthyridines, and the benzodiazepine receptor-binding ability of some non-benzodiapzepine compounds such as 3-X-imidazo-[1,2-*b*]pyridazines, 2-phenylimidazo[1,2-α]pyridines, 2-(alkoxycarbony)imidazo[2,1-*p*]benzothiazoles, and 2-arylquinolones (198–200).

4.4 Other QSAR Approaches

The similarity in approaches of Hansch analysis and Free-Wilson analysis allows them to be used within the same framework. This is based on their theoretical consistency and the numerical equivalencies of activity contributions. This development has been called the mixed approach and can be represented by the following equation:

$$\text{Log } 1/C = \sum a, + \sum c_j \varnothing_j + \text{constant} \qquad (1.80)$$

The term a_i denotes the contribution for each ith substituent, whereas $\varnothing_j$ is any physicochemical property of a substituent X_j. For a thorough review of the relationship between Hansch and Free-Wilson analyses, see the excellent reviews by Kubinyi (58, 195). A recent study of the P-glycoprotein inhibitory activity of 48 propafenone-type modulators of multidrug resistance, using a combined Hansch/Free-Wilson approach was deemed to have higher predictive ability than that of a stand-alone Free-Wilson analysis (201). Molar refractivity, which has a high collinearity with molecular weight, was a significant determinant of modulating ability. It is of interest to note that molecular weight has been shown to be an omnipresent parameter in cross-resistance profiles in multidrug-resistance phenomena (167).

5 APPLICATIONS OF QSAR

Over the last 40 years, the glut in scientific information has resulted in the development of thousands of equations pertaining to struc-

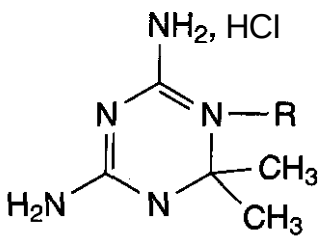


Figure 12. 4,6-Diamino-1,2-dihydro-2,2-dimethyl-1*R*-*s*-triazines.

ture-activity relationships in biological systems. In its original definition, the Hansch equation was defined to model drug-receptor interactions involving electronic, steric, and hydrophobic contributions. Nonlinear relationships helped refine this approach in cellular systems and organisms where pharmacokinetic constraints had to be considered and tackled. They have also found increased utility in addressing the complex QSAR of some receptor–ligand interactions. In many cases the Kubinyi bilinear model has provided a sophisticated approach to delineation of steric effects in such interactions. Examples of ligand–receptor interactions will be drawn from receptors such as the much-studied dihydrofolate reductases (DHFR), α-chymotrypsin and 5α-reductase (202–204).

5.1 Isolated Receptor Interactions

The critical role of DHFR in protein, purine, and pyrimidine synthesis; the availability of crystal structures of binary and ternary complexes of the enzyme; and the advent of molecular graphics combined to make DHFR an attractive target for well-designed heterocyclic ligands generally incorporating a 2,4-diamino-1,-3-diazapharmacophore (205). The earliest study focused on the inhibition of DHFR by 4,6-diamino-1,2-dihydro-2,2-dimethyl-1*R*-*s*-triazines, the structure of which is shown in Fig. 1.2 (202).

5.1.1 Inhibition of Crude Pigeon Liver DHFR by Triazines (202)

$$\begin{aligned}\text{Log } 1/\text{IC}_{50} = {} & 2.21(\pm 1.00)\pi \\ & - 0.28(\pm 0.17)\pi^2 \\ & + 0.84(\pm 0.76)\text{D} \\ & + 2.58(\pm 1.30)\end{aligned} \tag{1.81}$$

$$n = 15, \quad r^2 = 0.861,$$

$$s = 0.553, \quad \pi_0 = 4(3.6 - 6.0)$$

In all equations, n is the number of data points, r^2 is the square of the correlation coefficient, s represents the standard deviation, and the figures in parentheses are for construction of the 95% confidence intervals. π represents the hydrophobicity of the substituent R and π_0 is the optimum hydrophobic contribution of the R substituent. D is an indicator variable that acquires a value of 1.0 when a phenyl ring is present on the nitrogen and a value of zero for all other R. This is an example of a Hansch-Fujita-Ban analysis, where the indicator variable D establishes the contribution and thus the importance of a phenyl ring in DHFR inhibition. This equation has some limitations. Improper choice of N-substituents led to a high degree of collinearity between size and hydrophobicity and in terms of electronic contributions, spanned space was limited and thus inadequate. A subsequent study on the binding of these compounds to DHFR isolated from chicken liver was more revealing.

5.1.2 Inhibition of Chicken Liver DHFR by 3-X-Triazines (207)

$$\begin{aligned}\text{Log } 1/K_i = {} & 1.01(\pm 0.14)\pi' \\ & - 1.16(\pm 0.19)\log(\beta \cdot 10^{\pi'} + 1) \\ & + 0.86(\pm 0.57)\sigma + 6.33(\pm 0.14) \\ & n = 59, \quad r^2 = 0.821, \quad s = 0.906, \\ & \pi'_0 = 1.89(\pm 0.36) \quad \log \beta = -1.08\end{aligned} \tag{1.82}$$

In this example, the R group on the 2-nitrogen was restricted to an (3-X-phenyl) aromatic ring (205). Accurate K_i values were obtained from highly purified DHFR isolated from chicken liver. In most cases, π' represented the hydrophobicity of the substituent except in certain instances where X = $-$OR or $-CH_2ZC_6H_4$-Y. It was ascertained that alkoxy substituents were not making direct hydro-

phobic contact with the enzyme, given that their inhibitory activities were essentially constant from the methoxy to the nonyloxy substituent. In the bridged substituents where Z = O, NH, S, Se, the Y substituent again did not contact the enzyme surface. Variation in Y led to the same, constant biological activity. The coefficient with a' suggests that the substituent is engulfed in a hydrophobic pocket that has an optimal π'_O of 2. This value is consistent with that seen in the crude pigeon liver DHFR corrected for the presence of the phenyl group (4.0 – 2.0 = 2). The 0.86 p value (coefficient with σ) suggests that there could be a dipolar interaction between the electron deficient phenyl ring and a region of positively charged electrostatic potential in the enzyme, perhaps an arginine, lysine, or histidine residue. Hathaway et al. developed a QSAR for the inhibition of human DHFR by 3-X-triazines and obtained Equation 1.83 (208).

5.1.3 Inhibition of Human DHFR by 3-X-Triazines (208)

$$\begin{aligned}\text{Log } 1/K_i &= 1.07(\pm 0.23)\pi' \\ &\quad - 1.10(\pm 0.26)\log(\beta \cdot 10^{\pi'} + 1) \\ &\quad + 0.50(\pm 0.19)I + 0.82(\pm 0.66)\sigma \\ &\quad + 6.07(\pm 0.21) \end{aligned} \tag{1.83}$$

$$n = 60, \quad r^2 = 0.792, \quad s = 0.308,$$

$$\pi'_O = 2.0(\pm 0.87) \quad \log \beta = -0.577$$

The enhanced activity of the "bridged" substituents was corrected by the indicator variable I. Note that triazines bearing the bridge moieties $-CH_2NHC_6H_4Y$, $-CH_2OC_6H_4Y$, and $-CH_2SC_6H_4Y$ had unusually high enzyme binding activity. Note that the $-CH_2NHC_6H_5$ bridge is present in the endogenous substrate, folic acid. The bilinear dependency on hydrophobicity of the substituents parallels that seen in the case of chicken liver DHFR. A similar QSAR was obtained for DHFR isolated from L1210 murine leukemia cells (209).

5.1.4 Inhibition of L1210 DHFR by 3-X-Triazines (209)

$$\begin{aligned}\text{Log } 1/K_i &= 0.98(\pm 0.14)\pi' \\ &\quad - 1.14(\pm 0.20)\log(\beta \cdot 10^{\pi'} + 1) \\ &\quad + 0.79(\pm 0.57)\sigma + 6.12(\pm 0.14) \end{aligned} \tag{1.84}$$

$$n = 58, \quad r^2 = 0.810, \quad s = 0.264,$$

$$\pi'_O = 1.76(\pm 0.28) \quad \log \beta = -0.979$$

The consistency in these models versus prokaryotic DHFR is established by the coefficient with the hydrophobic term, the optimum a' value, and the rho value. These numerical coefficients can be contrasted sharply with those obtained from fungal and protozoal DHFR. Inhibition constants were determined for 3-X-triazines versus *Pneumocystis carinii* DHFR (210).

5.1.5 Inhibition of *P. carinii* DHFR by 3-X-Triazines (210)

$$\begin{aligned}\text{Log } 1/K_i &= 0.73(\pm 0.12)\pi' \\ &\quad - 1.36(\pm 0.35)\log(\beta \cdot 10^{\pi'} + 1) \\ &\quad - 0.78(\pm 0.42)I_{OR} \\ &\quad + 0.28(\pm 0.21)MR_Y \\ &\quad + 6.48(\pm 0.23) \end{aligned} \tag{1.85}$$

$$n = 43, \quad r^2 = 0.840, \quad s = 0.435,$$

$$\pi'_O = 3.99(\pm 0.68) \quad \log \beta = -3.925$$

In Equation 1.85, $I_{,}$ is an indicator variable that assumes a value of 1 when an alkoxy substituent is present and 0 for all other substituents. It is of interest to note that the Y substituent on the second phenyl ring now contributes to activity. The MR_Y term suggests that it most probably accesses a polar region of the active site of the enzyme. The positive coefficient with MR_Y suggests that an increase in bulk and/or polarizability enhances binding. The descending slope of the

bilinear equation is much steeper (1.36 – 0.73 = 0.63) than that seen with the mammalian and avian enzymes.

A similar model is obtained vs. the bifunctional protozoal DHFR from Leishmania major, which is coupled to thymidylate synthase (211).

5.1.6 Inhibition of L. *major* DHFR by 3-X-Triazines (211)

$$\begin{aligned}\text{Log } 1/K_i &= 0.65(\pm 0.08)\pi' \\ &\quad - 1.22(\pm 0.29)\log(\beta \cdot 10^{\pi'} + 1) \\ &\quad - 1.12(\pm 0.29)I_{OR} \\ &\quad + 0.58(\pm 0.16)MR_Y \\ &\quad + 5.05(\pm 0.16) \\ n &= 41, \quad r^2 = 0.931, \quad s = 0.298, \\ \pi'_O &= 4.54 \quad \log \beta = -4.491\end{aligned} \tag{1.86}$$

QSAR analysis on a limited set of 3-X-triazines assayed by Chio and Queener versus *Toxoplasmosis* gondii led to the formulation of Equation 1.87 (202, 212).

5.1.7 Inhibition of T. *gondii* DHFR by 3-X-Triazines

$$\begin{aligned}\text{Log } 1/IC_{50} &= 0.39(\pm 0.20)\pi' \\ &\quad - 0.43(\pm 0.19)MR_Y + 6.65(\pm 0.30) \\ n &= 17, \quad r^2 = 0.810, \quad s = 0.289\end{aligned} \tag{1.87}$$

A quick comparison of QSAR 1.82–1.84 reveals the strong similarity between the avian and mammalian models. In fact because of its increased stability, chicken liver DHFR has often been used as a surrogate for human DHFR in enzyme-inhibition studies. The intercepts, coefficients with π'_3 and optimum π'_O for avian (6.33, 1.01, 1.9), human (6.07, 1.07, 2.0), and mouse leukemia (6.12, 0.98, 1.76) can be compared to the corresponding values for P. carinii (6.48, 0.73, 3.99) and Leishmania major (5.05, 0.65, 4.54). QSAR 1.81 and 1.87 are not included in the comparison because crude pigeon enzyme was used in the former and the testing for QSAR 1.87 was conducted under different assay conditions; K_i values were not determined. A noteworthy difference between these models is the wide disparity in π_O values. The binding site of the protozoal and fungal species comprises an extensive hydrophobic surface unlike the abbreviated pockets in the mammalian and avian enzymes. The positive coefficients with the MR_Y terms suggests that added bulk on the bridged phenyl ring enhances inhibitory potency. The study versus T. gondii DHFR (QSAR 1.87) included a number of mostly small, polar substituents (NH_2, NO, $CONMe_2$) on the bridged phenyl and their activities were considerably lower than the unsubstituted analog. Comparative QSAR can be useful, particularly if the biological data are consistent (tested under the same assay conditions, excellent purity of enzymes, substrates, inhibitors, buffers), and the choice of substituents is appropriate.

One of the major problems that arises with some QSAR studies is extrapolation from beyond spanned space. Predictive ability is sound when one has probed an adequate range in electronic, hydrophobic, and steric space. At the onset of the study, the training set should address these concerns. Lack of adequate attention to such issues can result in QSAR models that are misleading. When examined on its own, such a model may appear to withstand statistical rigor and apparent transparency but, on being subjected to lateral validation, loopholes emerge. A brief study to illustrate this phenomenon is outlined below.

Four different QSAR were derived for the inhibition of DHFR from rat liver, human leukemia, mouse L1210, and bovine liver by 2,4-diamino, 5-Y, 6-Z-quinazolines (Fig. 1.3) (202, 213–215). A comparison of their QSAR presents an interesting study on the importance of spanned space in delineating enzyme-receptor interactions.

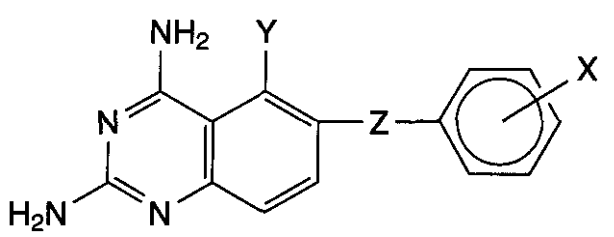


Figure 1.3. 2,4-Diamino, 5-Y, 6-Z-quinazolines.

5.1.8 Inhibition of Rat Liver DHFR by 2,4-Diamino, 5-Y, 6-Z-quinazolines (213)

$$\begin{aligned}\text{Log } 1/\text{IC}_{50} &= 0.78(\pm0.12)\pi_5 \\ &+ 0.81(\pm0.12)\text{MR}_6 \\ &- 0.06(\pm0.02)\text{MR}_6^2 \\ &- 0.73(\pm0.49)I_1 - 2.15(\pm0.38)I_2 \\ &- 0.54(\pm0.21)I_3 - 1.40(\pm0.41)I_4 \\ &+ 0.78(\pm0.37)I_6 \\ &- 0.20(\pm0.12)\text{MR}_6 \cdot I \\ &+ 4.92(\pm0.23)\end{aligned} \tag{1.88}$$

$$n = 101, \quad r^2 = 0.924, \quad s = 0.441,$$

$$\text{MR}_{6,O} = 6.4(\pm0.8)$$

5.1.9 Inhibition of Human Liver DHFR by 2,4-Diamino, 5-Y, 6-Z-quinazolines (214)

$$\begin{aligned}\text{Log } 1/K_i &= -2.87(\pm0.16)I_1 \\ &+ 0.29(\pm0.14)I_2 \\ &- 0.38(\pm0.11)\text{MR}_6 \\ &- 0.29(\pm0.06)\pi_R \\ &- 0.19(\pm0.07)\text{MR}_R + 10.12(\pm0.45)\end{aligned} \tag{1.89}$$

$$n = 47, \quad r^2 = 0.914, \quad s = 0.420$$

5.1.10 Inhibition of Murine L1210 DHFR by 2,4-Diamino, 5-Y, 6-Z-quinazolines (214)

$$\begin{aligned}\text{Log } 1/\text{IC}_{50} &= 0.49(\pm0.11)I_2 \\ &- 1.23(\pm0.25)I_3 \\ &- 0.30(\pm0.07)\text{MR}_6 \\ &- 0.12(\pm0.04)\pi_R + 9.36(\pm0.27)\end{aligned} \tag{1.90}$$

$$n = 24, \quad r^2 = 0.817, \quad s = 0.235$$

5.1.11 Inhibition of Bovine Liver DHFR by 2,4-Diamino, 5-Y, 6-Z-quinazolines (215)

$$\begin{aligned}\text{Log } 1/\text{IC}_{50} &= 0.70(\pm0.24)\text{MR}_6 \\ &+ 4.72(\pm0.59)\end{aligned} \tag{1.91}$$

$$n = 11, \quad r^2 = 0.823, \quad s = 0.420$$

These QSAR vary in size and the number of variables used to define inhibitory activity. Selassie and Klein have described a more thorough comparative analysis of these QSAR (202). *A* brief focus on the MR_6 term reveals that its coefficients vary remarkably in all four sets. QSAR ***1.88*** is a parabola with an optimum of ***6.4.*** Because it is parabolic in nature, the coefficient of the ascending slope cannot be compared with the linear slopes in QSAR ***1.89–1.91.*** Figure ***1.4*** illustrates the problems with QSAR ***1.89–1.91,*** which failed to test analogs across the available space.

Figure ***1.4*** reveals that QSAR ***1.89*** and ***1.90*** were sampled in the suboptimal MR_6 range; thus, the negative dependency on MR_6. On the other hand, QSAR ***1.91*** was focused on the ascending portion of the curve and thus only molecules in the ***0.1–3.4*** range were tested. Thus, with a limited set of compounds, one gets a misleading picture of the biological interactions. ▪

Enzymatic reactions in nonaqueous solvents have generated a great deal of interest, fueled in part by the commercial application of enzymes as catalysts in specialty synthesis. The increasing demand for enantiopure pharmaceuticals has accelerated the study of enzymatic reactions in organic solvents containing

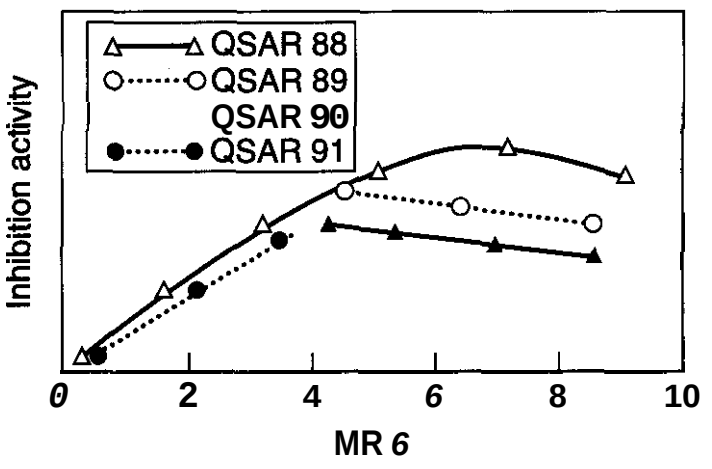


Figure 1.4. Gaps in spanned space of MR6 for 2,4-diamino-quinazolines.

little or no water (216). To investigate the substrate specificity of a-chymotrypsin in pentanol, a series of X-phenyl esters of N-benzoyl-L-alanine (Fig. 1.5) were synthesized and their binding constants were evaluated in buffer and in pentanol (203). The following QSAR 1.92 and 1.93 were derived in phosphate buffer and pentanol.

5.1.12 Binding of X-Phenyl, *N*-Benzoyl-L-alaninates to α-Chymotrypsin in Phosphate Buffer, pH 7.4 (203)

$$\begin{aligned}\text{Log } 1/K_M &= 0.28(\pm 0.11)\pi + 0.51(\pm 0.24)\sigma^- \\ &\quad + 0.38(\pm 0.23)\text{MR} + 3.70(\pm 0.24)\end{aligned} \tag{1.92}$$

$$n = 16, \quad r^2 = 0.834, \quad s = 0.198$$

5.1.13 Binding of X-Phenyl, *N*-Benzoyl-L-alaninates to α-Chymotrypsin in Pentanol (203)

$$\begin{aligned}\text{Log } 1/K_M &= 0.25(\pm 0.09)\pi \\ &\quad + 0.24(\pm 0.18)\sigma^- \\ &\quad + 4.10(\pm 0.09)\end{aligned} \tag{1.93}$$

$$n = 17, \quad r^2 = 0.762, \quad s = 0.156$$

Outliers in QSAR 1.92 included the 4-t-butyl and 4-OH analogs, whereas the 4-CONH, analog was an outlier in QSAR 1.93. These results were recently reanalyzed by Kim (217,218) with respect to the role of enthalpic and entropic contributions to ligand binding with a-chymotrypsin. Use of the Fujiwara hydrophobic enthalpy parameter r, and the hydrophobic entropy parameter π_S led to the development of QSAR 1.94 and 1.95 (219).

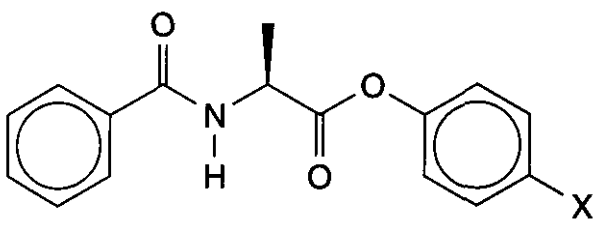


Figure 1.5. X-Phenyl, *N*-benzoyl-L-alaninates.

5.1.14 Binding of X-Phenyl, *N*-Benzoyl-L-alaninates in Aqueous Phosphate Buffer (218)

$$\begin{aligned}\text{Log } 1/K_M &= 0.38(\pm 0.11)\pi_H + 0.19(\pm 0.07)\pi_S \\ &\quad + 0.53(\pm 0.11)\sigma^- \\ &\quad + 0.26(\pm 0.10)\text{MR} + 3.77(\pm 0.11)\end{aligned} \tag{1.94}$$

$$n = 15, \quad r^2 = 0.806, \quad s = 0.200$$

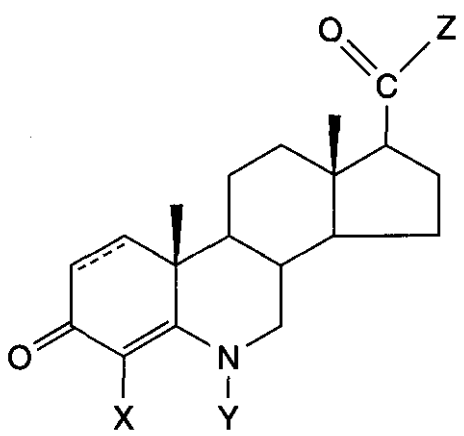
5.1.15 Binding of X-Phenyl, *N*-Benzoyl-L-alaninates in Pentanol (218)

$$\begin{aligned}\text{Log } 1/K_M &= 0.21(\pm 0.08)\pi_H + 0.31(\pm 0.05)\pi_S \\ &\quad + 0.20(\pm 0.08)\sigma^- + 4.16(\pm 0.04)\end{aligned} \tag{1.95}$$

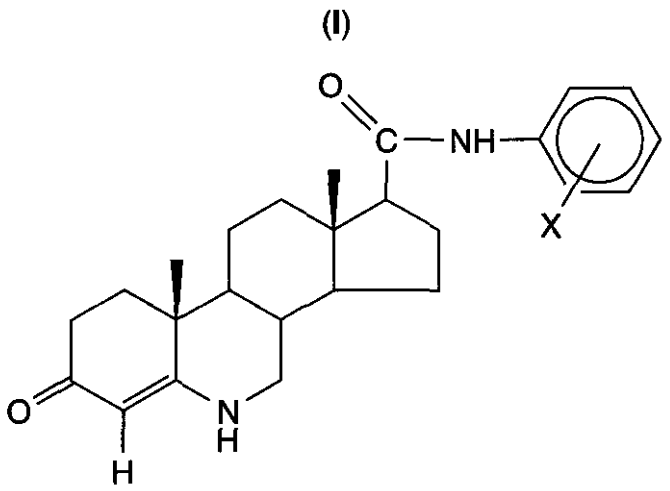
$$n = 15, \quad r^2 = 0.787, \quad s = 0.160$$

The disappearance of the MR term in QSAR 1.93 and 1.95 is significant. The MR term usually relates to nonspecific, dispersive interactions in polar space. Thus, its presence in QSAR 1.92 and 1.94 suggests that substrates bearing polarizable substituents may displace the ordered-category II water molecules. In pentanol, the substrate may be faced with the task of displacing pentanol, not water, from the enzyme and thus the MR term is no longer of consequence. QSAR 1.94 also indicates that the enthalpy term π_H plays a more critical role in binding than the entropy term π_S. Note that these roles are reversed in QSAR 1.95, suggesting that binding in pentanol is largely an entropic-driven process. Similar results were obtained by Compadre et al. in a study on the hydrolysis of X-phenyl-*N*-benzoyl-glycinates by cathepsin B in aqueous buffer and acetonitrile (220). Kim's analysis provides an excellent example of a study that focuses on mechanistic interpretation and clearly demonstrates that a thermodynamic approach in QSAR can provide pertinent information about the energetics of the ligand binding process.

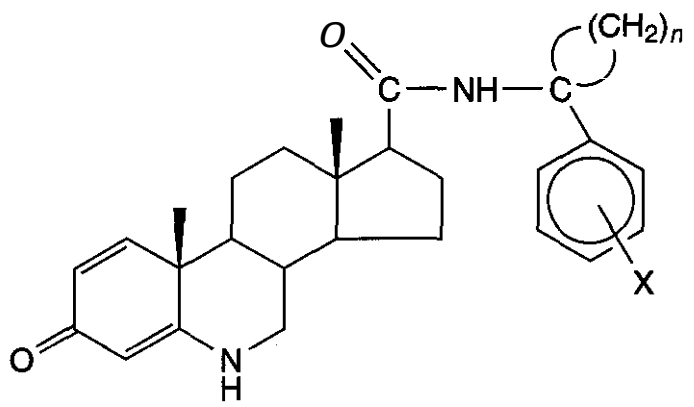
5α-Reductase, a critical enzyme in male sexual development, mediates the reduction of testosterone to dihydrotestosterone (DHT). Elevated levels of DHT in certain disease states such as benign prostatic hypertrophy and prostatic cancer drives the need for effective inhibitors of **5α-reductase. A** recent QSAR study on inhibition of human **5α-reductase,** type 1 by various steroid classes was carried out by **Kurup** et al. (204,221,222). **A** few of the models will be examined to demonstrate the importance and power of lateral validation. The three classes of steroidal inhibitors are depicted in Fig. 1.6.



(I)



(II)



(III)

Figure 1.6. Steroidal inhibitors of 5α-reductase.

5.1.16 Inhibition of 5-α-Reductase by 4-X, N-Y-6-azaandrost-17-CO-Z-4-ene-3-ones, I

Log $1/K_i$

$n = 21$, $r^2 = 0.829$, $s = 0.406$

outliers: X = Y = H, Z = $NHCMe_3$;

X = Me, Y = H, Z = CH_2CHMe_2

5.1.17 Inhibition of 5-α-Reductase by 17β-(N-(X-phenyl)carbamoyl)-6-azaandrost-4-ene-3-ones, II

$$\text{Log } 1/K_i = 0.35(\pm 0.09)\text{Clog P} + 0.26(\pm 0.11)B5_{ortho} + 5.08(\pm 0.58) \quad (1.97)$$

$n = 12$, $r^2 = 0.942$, $s = 0.154$

outlier: 2,5-$(CF_3)_2$

5.1.18 Inhibition of 5-α-Reductase by 17β-(N-(1-X-phenyl-cycloalkyl)carbamoyl)-6-azaandrost-4-ene-3-ones, III

$$\text{Log } 1/K_i = 0.32(\pm 0.17)\text{Clog P} + 6.34(\pm 1.15) \quad (1.98)$$

$n = 5$, $r^2 = 0.920$, $s = 0.090$

outlier: n = 5, X = 4-t-Bu

In all these equations, the coefficients with hydrophobicity as represented by Clog P, suggest that binding of these azaandrostene-ones occurs on the surface of the binding site where partial desolvation can occur. I is an indicator variable that pinpoints the negative effect of a double bond at C-1. **A** bulky substituent on N-6 is detrimental to activity, whereas a large substituent in the ortho position on the aromatic ring enhances activity (QSAR 1.97). The bulky ortho substituents (mostly t-Bu) may destroy coplanarity with the amide bridge by perhaps twisting of the phenyl ring and enhancing its hydrophobic contact with the

binding site on the enzyme. Note that the larger intercept in QSAR 1.98 versus QSAR 1.97 suggests that hydrophobicity is more important in this area.

5.2 Interactions at the Cellular Level

QSAR analysis of studies at the cellular level allows us to get a handle on the physicochemical parameters critical to pharmacokinetics processes, mostly transport. Cell culture systems offer an ideal way to determine the optimum hydrophobicity of a system that is more complex than an isolated receptor. Extensive QSAR have been developed on the toxicity of 3-X-triazines to many mammalian and bacterial cell lines (202, 209). A comparison of the cytotoxicities of these analogs vs. sensitive murine leukemia cells (L1210/S) and methotrexate-resistant murine leukemia cells (L1210/R) reveals some startling differences.

5.2.1 Inhibition of Growth of L1210/S by 3-X-Triazines (209)

$$\begin{aligned}\text{Log } 1/IC_{50} &= 1.13(\pm 0.18)\pi \\ &\quad - 1.20(\pm 0.21)\log(\beta \cdot 10^{\pi} + 1) \\ &\quad + 0.66(\pm 0.23)I_{R} \\ &\quad - 0.32(\pm 0.17)I_{OR} \\ &\quad + 0.94(\pm 0.37)\sigma + 6.72(\pm 0.13) \\ &\quad n = 61, \quad r^2 = 0.792, \quad s = 0.241, \\ &\quad \pi_O = 1.45(\pm 0.93) \quad \log \beta = -0.274\end{aligned} \tag{1.99}$$

5.2.2 Inhibition of Growth of L1210/R by 3-X-Triazines (209)

$$\begin{aligned}\text{Log } 1/IC_{50} &= 0.42(\pm 0.05)\pi - 0.15(\pm 0.05)\text{MR} \\ &\quad + 4.83(\pm 0.11) \\ &\quad n = 62, \quad r^2 = 0.885, \quad s = 0.220\end{aligned} \tag{1.100}$$

There is a radical difference between these two QSAR. QSAR 1.99 is very similar to the one (QSAR 1.84) obtained versus the L1210 DHFR and it can be posited that the cytotoxicity in the sensitive cell line results from the inhibition of the enzyme. The intercepts suggest that slight interference with folate metabolism significantly affects growth. A comparison of the sensitive and resistant QSAR reveals a substantial difference in the coefficients with π. The lack of many variables in QSAR 1.100 and its overall simplicity suggests that inhibition of the enzyme is not the critical step, but rather transport to the site of action in these resistant cells may be of utmost importance. This particular cell line was resistant to methotrexate by virtue of elevated levels of DHFR and also overexpression of glycoprotein, GP-170 (209). Thus, modified transport through the dysfunctional membrane would severely curtail the partitioning process, resulting in a coefficient with π that is only one-half (0.42) of what is normally seen. The negative coefficient with the MR term indicates that size plays a role, albeit a negative one, in passage through the GP-170-fortified membrane and to the site of action.

The QSAR paradigm has been shown to be particularly useful in environmental toxicology, especially in acute toxicity determinations of xenobiotics (223). There has recently been an emphasis on "transparent, mechanistically comprehensive QSAR for toxicity," a move that is welcomed by many researchers in the field (224, 225). Cronin and Schultz developed QSAR 1.101 to describe the polar, narcotic toxicity of a large set of substituted phenols. A number of phenols with ionizable or reactive groups (e.g., —COOH, —NO_2, —NO, —NH_2, or —$NHCOCH_3$) were omitted from the final analysis (226).

5.2.3 Inhibition of Growth of *Tetrahymena pyriformis* (40 h)

$$\begin{aligned}\text{Log } 1/C &= 0.67(\pm 0.02)\text{Clog } P \\ &\quad - 0.67(\pm 0.55)E_{\text{LUMO}} - 1.12 \\ &\quad n = 120, \quad r^2 = 0.893, \quad s = 0.271\end{aligned} \tag{1.101}$$

Using Hammett σ constants, Garg et al. rederived QSAR 1.102 for the same set and QSAR 1.103 and 1.104 for the diverse set of multi-, di-, and monophenols, which were se-

questered into two subsets containing electron-releasing and electron-attracting substituents, respectively (227).

5.2.4 Inhibition of Growth of T. pyriformis by Phenols (using σ) (227)

$$\begin{aligned}\text{Log } 1/C &= 0.64(\pm 0.04)\text{Clog } P \\ &+ 0.61(\pm 0.12)\sigma + 1.84(\pm 0.13) \\ n &= 119, \quad r^2 = 0.896, \quad s = 0.265\end{aligned} \tag{1.102}$$

5.2.5 Inhibition of Growth of T. *pyriformis* by Electron-Releasing Phenols (227)

$$\begin{aligned}\text{Log } 1/C &= 0.66(\pm 0.05)\text{Clog P} \\ &+ 1.63(\pm 0.15) \\ n &= 44, \quad r^2 = 0.946, \quad s = 0.182\end{aligned} \tag{1.103}$$

5.2.6 Inhibition of Growth of T. *pyriformis* by Electron-Attracting Phenols (227)

$$\begin{aligned}\text{Log } 1/C &= 0.63(\pm 0.07)\text{Clog P} \\ &+ 0.54(\pm 0.16)\sum\sigma \\ &+ 1.92(\pm 0.18) \\ n &= 100, \quad r^2 = 0.836, \quad s = 0.327\end{aligned} \tag{1.104}$$

There is excellent agreement between QSAR 1.101 and QSAR 1.104, in terms of the importance of hydrophobicity and electron demand of the substituents: the coefficients with Clog P are similar and there is a good correspondence between E_{LUMO} and σ. Nevertheless, separation of the phenols into subsets, based on their electronic attributes, indicates that different mechanisms of toxicity might be operative in this organism, a phenomenon that has been duplicated in mammalian cells (228). In a recent extension of toxicity studies on aromatics, Cronin and Schultz used a two-parameter or response-surface approach to define toxicity (229). In addition, indicator variables and group counts were included to broaden the applicability of the approach. An excellent comparison of the different modeling approaches (MLR, PLS, and Bayesian-regularized neural networks) in QSAR is also made (229).

5.2.7 Inhibition of Growth of T. pyriformis by Aromatic Compounds (229)

$$\begin{aligned}\text{Log } 1/\text{IgC}_{50} &= 0.633\log P - 0.526E_{\text{LUMO}} \\ &+ 0.721I_{2,4\text{ AP}} - 1.61I_{\text{strong acid}} \\ &+ 0.314\sum\text{H-donor} - 1.39 \\ n &= 268, \quad r^2 = 0.780, \quad s = 0.393\end{aligned} \tag{1.105}$$

The indicator variables $I_{2,4}$ ▬ and $I_{\text{strong acid}}$ suggest that 2- and 4-amino-substituted phenols enhance toxicity, whereas strong acids decrease toxicity, respectively. The H-bond donor parameter may be correcting for the added potency of amino phenols. The low r^2 may be attributed to inherent variability in biological data and to the commingling of data from four different studies. The wide variety of compounds with different toxicity mechanisms, present in this combined study, would also be a contributing factor to the low r^2. Overall, this regression-based approach shows adequate predictability and is transparent, thus aiding in mechanistic interpretation.

5.3 Interactions *In Vivo*

The paucity of QSAR studies in whole animals is understandable in terms of the costs, the heterogeneity of the biological data, and the complexity of the results. Nevertheless, in the few studies that have been done, excellent QSAR have been obtained, despite the small number of subjects in the data set (164). One particular example is insightful. The renal and nonrenal clearance rates of a series of 11 β-blockers, including bufuralol, tolamolol, propranolol, alprenolol, oxprenolol, acebutol, timolol, metoprolol, prindolol, atenolol, and nadolol were measured (230). The following QSAR were formulated using those data (164).

5.3.1 Renal Clearance of β-Adrenoreceptor Antagonists

$$\begin{aligned}\text{Log k} &= -0.42(\pm 0.12)\text{Clog P} \\ &+ 2.35(\pm 0.24) \\ n &= 10, \quad r^2 = 0.888, \quad s = 0.185\end{aligned} \tag{1.106}$$

5.3.2 Nonrenal Clearance of β-Adrenoreceptor Antagonists

$$\text{Log } k = 1.94(\pm 0.61)\text{Clog P} - 2.00(\pm 0.80)\log(\beta \cdot P + 1) + 1.29(\pm 0.30) \quad (1.107)$$

$$n = 10, \quad r^2 = 0.950, \quad s = 0.168,$$

$$\text{Clog } P_0 = 2.6 \pm 1.5 \quad \log \beta = -0.813$$

outlier: oxprenolol

It is apparent from QSAR 1.106 and 1.107, that the hydrophobic requirements of the substrates vary considerably. As expected, renal clearance is enhanced in the case of hydrophilic drugs, whereas nonrenal clearance shows a strong dependency on hydrophobicity. Note that QSAR 1.107 is stretching the limits of the bilinear model with only 10 data points! The 95% confidence intervals are also large but, nevertheless, the equations serve to emphasize the difference in clearance mechanisms that are clearly linked to hydrophobicity.

In formulating QSAR, it is useful to use a well-designed series to optimize a particular biological activity. It is also important to ensure that the ratio of compounds to parameters is 5, so that collinearity is minimized while spanned space is maximized. A normal distribution of biological data is necessary. A violation of these guidelines usually leads to statistically insignificant QSAR or models that defy predictability. One of our earliest works on the inhibition of E. *coli* DHFR by 2,4-diamino-5-X-benzyl pyrimidines led to the derivation of the following equation (231):

$$\text{Log } 1/K_i = -1.13\sigma_R + 5.54 \quad (1.108)$$

$$n = 10, \quad r^2 = 0.972, \quad s = 0.182$$

Most of the variance in these data was explained by the Hammett through-resonance constant (σ_R). It implied that electron-releasing substituents enhanced inhibitory potency. Later, expanded and extensive studies on this system revealed that inhibition of the bacterial enzyme was related to mostly steric effects and there was no dependency on electronic terms. Careful analysis of the initial data revealed that it had a limited range in hydrophobicity and steric attributes. The lack of other QSAR to validate the findings in QSAR 1.108 made it statistically significant, at that time, but mechanistically weak. Most weaknesses in QSAR formulations usually violate the compound-to-parameter ratio rule (232, 233).

6 COMPARATIVE QSAR

6.1 Database Development

There are literally dozens of databases containing information about chemical structures, synthetic methods, and reaction mechanisms. The C-QSAR database is a database for QSAR models (164,234). It was designed to organize QSAR data on physical (PHYS) organic reactions as well as chemical-biological (BIO) interactions, in numerical terms, to bring cohesion and understanding to mechanisms of chemical-biodynamics. The two databases are organized on a similar format, with the emphasis on reaction types in the PHYS database. The entries in the BIO database are sequestered into six main groups: macromolecules, enzymes, organelles, single-cell organisms, organs/tissues, and multicellular organisms (e.g., insects). The combined databases or the separate PHYS or BIO databases can be searched independently by a string search or searching using the SMILES notation. A SMILES search can be approached in three ways: one can identify every QSAR that contains a specific molecule, one can use a MERLIN search that locates all derivatives of a given structure, or one can search on single or multiple parameters. For a more thorough description of the C-QSAR database and ways to search it, see Hansch et al. (234) and Hansch et al. (164). The net result of searching the QSAR database is to "mine" for models; one could thus call it model-mining.

6.2 Database: Mining for Models

To enhance our understanding of ligand-receptor interactions and bring coherence to these relationships, there needs to be a con-

Table 1.5 Rho Values for Chemical and Biochemical Reactions

	Solvent	Radical Reagent	*n*	ρ^+ (σ^+)
		Hydrogen Abstraction from Unhindered Phenols		
1	CCl_4	$(CH_3)_3CO\cdot$	14	−1.81 (±0.77)
2	Benzene	$(CH_3)_3CO\cdot$	12	−0.82 (±0.08)
3	CCl_4	$(CH_3)_3CO\cdot$	5	−0.82 (±0.16)
		X-phenols–Enzyme Systems		
1	Horseradish peroxidase	—	13	−2.68 (±0.78)
2	Ladoperoxidase	—	11	−1.34 (±0.55)

certed effort not only to develop high-quality regressions but also to create models that resonate with those drawn from mechanistic organic chemistry. A comprehensive, integrated database C-QSAR allows us to do so; it contains over 16,000 examples drawn from all facets of chemistry and biology. An example on the toxicity of X-phenols will illustrate the usefulness of this database (164, 228, 235–238). Recently, increasing numbers of QSAR for phenols have been based on Brown's σ^+ term, an electronic term that was first designed to rationalize electronic effects of substituents on electrophilicaromatic substitution. Studies conducted at EPA gave early indications that embryologic defects of rat embryos in *vitro* could be correlated by σ^+, as seen in QSAR 1.109109 (239).

6.2.1 Incidence of Tail Defects of Embryos (235)

$$\begin{aligned}\text{Log } 1/C &= -0.58(\pm 0.21)\sigma^+ \\ &\quad + 3.51(\pm 0.14) \qquad (1.109)\\ n = 10, &\quad r^2 = 0.832, \quad s = 0.189\end{aligned}$$

Soon, this parameter was shown to correlate radical reactions in chemistry as well as chemical-biological interactions in an extensive compilation (240). Another older study by Richard et al. on the inhibition of replicative DNA synthesis in Chinese hamster ovary cells was examined and led to the development of Equation 1.110 (241). Again, there was a dependency on σ^+.

6.2.2 Inhibition of DNA Synthesis in CHO Cells by X-Phenols (236)

$$\begin{aligned}\text{Log } 1/C &= -0.74(\pm 0.34)\sigma^+ \\ &\quad - 1.02(\pm 0.41)\text{CMR} \qquad (1.110)\\ &\quad + 6.97(\pm 1.16)\\ n = 9, &\quad r^2 = 0.915, \quad s = 0.305\end{aligned}$$

These Brown ρ^+ values were in line with those obtained from chemical and biological systems (228) see Table 1.5.

Cytotoxicity studies of X-phenols versus L1210 cells in culture led to an unusual result, which was baffling but reminiscent of Hammett plots related to changes in mechanism (228).

6.2.3 Inhibition of Growth of L1210 by X-Phenols

$$\begin{aligned}&\text{Log } 1/IC_{50}\\ &= -0.83(\pm 0.18)\sigma^+ \\ &\quad + 0.74(\pm 0.28)\sigma^{+2} \qquad (1.111)\\ &\quad + 0.56(\pm 0.15)\log P\\ &\quad - 0.45(\pm 0.21)\log(\beta\cdot P + 1)\\ &\quad + 2.70(\pm 0.26)\\ &\quad n = 39, \quad r^2 = 0.913, \quad s = 0.229,\\ &\quad \text{Log } P_O = -0.18 \quad \text{Log } \beta = -2.28\\ &\quad \text{outliers: } 4\text{-}C_2H_5, 3\text{-}NH_2\end{aligned}$$

Sequestering of the data into two subsets with

varying electronic attributes ($a^- > 0$ and $\sigma^+ < 0$) led to the derivation of the following equations.

6.2.4 Inhibition of Growth of L1210 by Electron-Withdrawing Substituents ($\sigma^+ > 0$)

$$\text{Log } 1/IC_{50} = 0.62(\pm 0.16)\text{Log P} + 2.35(\pm 0.31) \quad (1.112)$$

$$n = 15, \quad r^2 = 0.845, \quad s = 0.232,$$

outlier: 3-OH

6.2.5 Inhibition of Growth of L1210 by Electron-Donating Substituents ($\sigma^+ < 0$)

$$\text{Log } 1/IC_{50} = -1.58(\pm 0.26)\sigma^+ + 0.21(\pm 0.06)\text{Log } P + 3.10(\pm 0.24) \quad (1.113)$$

$$n = 23, \quad r^2 = 0.898, \quad s = 0.191,$$

outliers: $3\text{-}NH_2$, 4-NHAc

In QSAR 1.113, 62% of the variance is accounted for by σ^+ and 28% is explained by log P. It appears that free-radical-mediated toxicity is responsible for the growth-inhibitory effects of the phenols. Homolytic bond dissociation energies related to the homolytic cleavage of the OH bond in the following reaction: ($X\text{—}C_6H_4OH + C_6H_5O\cdot \rightarrow X\text{—}C_6H_4O\cdot + C_6H_5OH$) have been used in lieu of σ^+ values. The net result is similar, as seen in QSAR 1.114 (242).

$$\text{Log } 1/IC_{50} = -0.21(\pm 0.03)\text{BDE} + 0.21(\pm 0.04)\text{Log } P + 3.11(\pm 0.17) \quad (1.114)$$

$$n = 52, \quad r^2 = 0.920, \quad s = 0.202,$$

outliers: 4-NHAc, $3\text{-}NH_2$, $3\text{-}NMe_2$

This data set contains a wide diversity of phenolic inhibitors, including a large number of ortho-substituted compounds, estrogenic phenols (β-estradiol, DES, nonyl phenol), and other antioxidants whose activities are well predicted by this model. The model suggests that cytotoxicity is an outcome of phenoxy radical formation and subsequent interaction with a relatively nonpolar receptor. The small hydrophobic coefficient suggests that DNA could be a likely target.

The appearance of the σ^+ parameter in a large number of reactions and interactions involving X-phenols indicates that the phenoxy radical can be a potent, reactive intermediate in myriad reactions. The availability of a fast, easily retrievable computerized database to corroborate this phenomenon was useful. This approach of lateral validation was crucial in establishing a QSAR model that was not only statistically significant but also mechanistically interpretable.

6.3 Progress in QSAR

The last four decades have seen major changes in the QSAR paradigm. In tandem with developments in molecular modeling and X-ray crystallography, it has impacted drug design and development in many ways. It has also spawned 3D QSAR approaches that are routinely used in computer-assisted molecular design. In terms of ligand design, it shares center stage with other approaches such as structure-based ligand design and other rational drug design approaches including docking. methods and genetic algorithms (243). Success stories in QSAR have been recently reviewed (244, 245). Bioactive compounds have emerged in agrochemistry, pesticide chemistry, and medicinal chemistry.

Bifenthrin, a pesticide, was the product of a design strategy that used cluster analysis (244) (Fig. 1.7). Guided by QSAR analysis, the chemists at Kyorin Pharmaceutical Company designed and developed Norfloxacin, a 6-fluoro quinolone, which heralded the arrival of a new class of antibacterial agents (246) (Fig. 1.7). Two azole-containing fungicides, metconazole (Fig. 1.8) and ipconazole were launched in 1994 in France and Japan, respectively (247). Lomerizine, a 4-*F*-benzhydryl-4-(2,3,4-trimethoxy benzyl) piperazine, was introduced into the market in 1999 after extensive design strategies using QSAR (248) (Fig. 1.8). Flobufen, an anti-inflammatory agent was designed by Kuchar et al. as a long-acting agent without the usual gastric toxicity

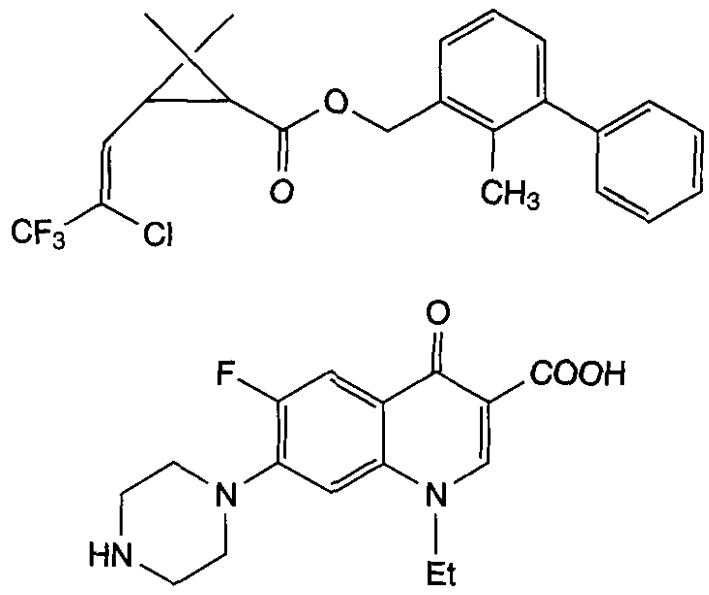


Figure *1.7. Bifenthrin and* Norfloxacin.

(249) (Fig. 1.8). It is currently in clinical trial. Other examples of the commercial utility of QSAR include the development of metamitron and bromobutide (250). In most of these examples, QSAR was used in combination with other rational drug-design strategies, which is a useful and generally fruitful approach.

In addition to these commercial successes, the QSAR paradigm has steadily evolved into a science. It is empirical in nature and it seeks to bring coherence and rigor to the QSAR models that are developed. By comparing models one is able to more fully comprehend scientific phenomena with a "global" perspective; trends in patterns of reactivity or biological activity become self-evident.

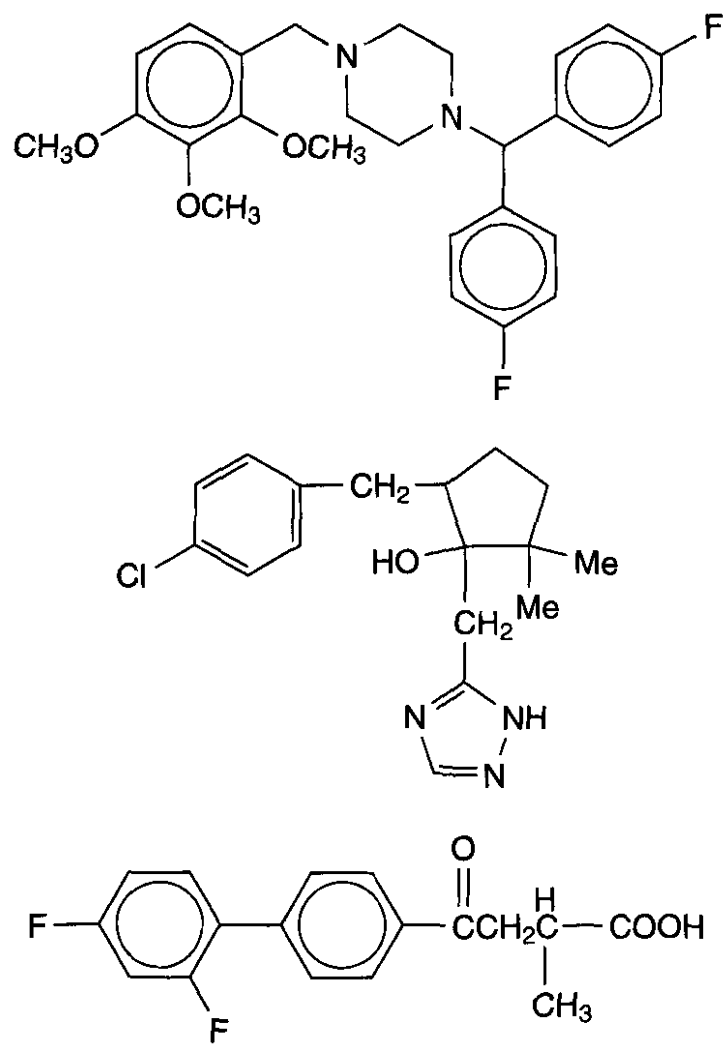


Figure *1.8.* Lomerizine, Metconazole, and ***Flobufen.***

7 SUMMARY

QSAR has done much to enhance our understanding of fundamental processes and phenomena in medicinal chemistry and drug design (251). The concept of hydrophobicity and its calculation has generated much knowledge and discussion as well as spawned a mini-industry. QSAR has refined our thinking on selectivity at the molecular and cellular level. Hydrophobic requirements vary considerably between tumor-sensitive cells and resistant ones. It has allowed us to design more selectivity into antibacterial agents that bind to dihydrofolate reductase. QSAR studies in the pharmacokinetic arena have established different hydrophobic requirements for renal/nonrenal clearance, whereas the optimum hydrophobicity for CNS penetration has been determined by Hansch et al. (252). QSAR has helped delineate allosteric effects in enzymes' such as cyclooxygenase, trypsin, and in the well-defined and complex hemoglobin system (253, 254).

QSAR has matured over the last few decades in terms of the descriptors, models, methods of analysis, and choice of substituents and compounds. Embarking on a QSAR project may be a daunting and confusing task to a novice. However, there are many excellent reviews and tomes (1, 4, 19, 58–60) on this subject that can aid in the elucidation of the paradigm. Dealing with biological systems is not a simple problem and in attempting to develop a QSAR, one must always be cognizant of the biochemistry of the system analyzed and the limitations of the approach used.

REFERENCES

1. *C. Hansch and A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, John* Wiley & *Sons, New York, 1979.*

2. D. J. Livingstone, *J. Chem. Inf. Comput. Sci.*, **40,195 (2000).**
3. C. Hansch, A. Kurup, R. Garg, and H. Gao, *Chem. Rev.*, **101,619 (2001).**
4. H. Kubinyi in M. Wolff, Ed., *Burger's Medicinal Chemistry and Drug Discovery, Volume 1: Principles and Practice*, John Wiley & Sons, New York, **1995,** p. **497.**
5. A. Crum-Brown and T. R. Fraser, *Trans. R. Soc. Edinburgh*, **25, 151 (1868).**
6. C. Richet and C. R. Seancs, *Soc. Biol. Ses. Fil.*, **9,775 (1893).**
7. H. Meyer, *Arch. Exp. Pathol. Pharmakol.*, **42, 109 (1899).**
8. E. Overton, *Studien Uber die Narkose*, Fischer, Jena, Germany, **1901.**
9. J. Ferguson, *Proc. R. Soc. London Ser. B*, **127, 387 (1939).**
10. A. Albert, S. Rubbo, R. Goldacre, M. Darcy, and J. Stove, *Br. J. Exp. Pathol.*, **26, 160 (1945).**
11. A. Albert, *Selective Toxicity: The Physicochemical Bases of Therapy*, 7th ed., Chapman and Hall, London, **1985,** p. **33.**
12. P. H. Bell and R. O. Roblin, Jr *J. Am. Chem. Soc.*, **64,2905 (1942).**
13. L. P. Hammett, *Chem. Rev.*, **17,125 (1935).**
14. L. P. Hammett, *Physical Organic Chemistry*, 2nd ed., McGraw-Hill, New York, **1970.**
15. R. W. Taft, *J. Am. Chem. Soc.*, **74,3120 (1952).**
16. C. Hansch, P. P. Maloney, T. Fujita, and R. M. Muir, *Nature*, **194, 178 (1962).**
17. R. Nelson Smith, C. Hansch, and M. M. Ames, *J. Pharm. Sci.*, **64,599 (1975).**
18. T. Fujita, J. Iwasa, and C. Hansch, *J. Am. Chem. Soc.*, **86, 5175 (1964).**
19. C. Hansch and A. Leo in S. R. Heller, Ed., *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington, DC, **1995.**
20. C. Hansch, *Acc. Chem. Res.*, **2,232 (1969).**
21. H. Kubinyi, *Arzneim.-Forsch.*, **26,1991 (1976).**
22. S. M. Free and J. W. Wilson, *J. Med. Chem.*, **7, 395 (1964).**
23. T. Fujita and T. Ban, *J. Med. Chem.*, **14, 148 (1971).**
24. G. Klopman, *J. Am. Chem. Soc.*, **106, 7315 (1984).**
25. B. W. Blake, K. Enslein, V. K. Gombar, and H. H. Borgstedt, *Mutat. Res.*, **241,261 (1990).**
26. Z. Simon, *Angew. Chem. Znt. Ed. Eng.*, **13,719** (1974).
27. L. H. Hall and L. B. Kier, *J. Pharm. Sci.*, **66, 642** (1977).
28. L. B. Kier and L. H. Hall, *Molecular Structure Description. The Electrotopological State*, Academic Press, San Diego, CA, **1999.**
29. W. Tong, D. R. Lowis, R. Perkins, Y. Chen, W. J. Welsh, D. W. Goddette, T. W. Heritage, and D. M. Sleehan, *J. Chem. Inf. Comput Sci.*, **38, 669 (1998).**
30. S. J. Cho, W. Zheng, and A. Tropsha, *Pac. Symp. Biocomput.*, **305 (1998).**
31. H. Gao and J. Bajorath, *J. Mol. Diversity*, **4, 115 (1999).**
32. H. Gao, C. Williams, P. Labute, and J. Bajorath, *J. Chem. Inf. Comput. Sci.*, **39, 164** (1999).
33. W. J. Dunn III, S. Wold, U. Edlund, S. Hellberg, and J. Gasteeger, *Quant. Struct.-Act. Relat.*, **3, 131 (1984).**
34. J. Langley, *J. Physiol.*, **1, 367 (1878).**
35. P. Ehrlich, *Klin. Jahr.*, **6, 299 (1897).**
36. J. N. Langley, *J. Physiol.*, **33,374** (1905).
37. M. Famulok, *Curr. Opin. Struct. Biol.*, **9, 324 (1999).**
38. K. Y. Wang, S. Swaminathan, and P. H. Bolton, *Biochemistry*, **33, 7617 (1994).**
39. J. W. Lown in S. Neidle and M.-J. Waring, Eds., *Molecular Aspects of Anticancer Drug-DNA Interactions*, Macmillan, Basinstoke, *UK*, **1993,** p. **322.**
40. L. Morgenstern, M. Recanatini, T. E. Klein, W. Steinmetz, C. Z. Yang, R. Langridge, and C. Hansch, *J. Biol. Chem.*, **262, 10767 (1987).**
41. R. N. Smith, C. Hansch, K. H. Kim, B. Omiya, G. Fukumura, C. D. Selassie, P. Y. C. Jow, J. M. Blaney, and R. Langridge, *Arch. Biochem. Biophys.*, **215,319** (1982).
42. C. Hansch, T. Klein, J. McClarin, R. Langridge, and N. W. Cornell, *J. Med. Chem.*, **29, 615 (1986).**
43. C. D. Selassie, Z. X. Fang, R. Li, C. Hansch, T. Klein, R. Langridge, and B. T. Kaufman, *J. Med. Chem.*, **29,621 (1986).**
44. J. M. Blaney and C. Hansch in C. A. Ramsden, Ed., *Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds, Vol. 4, Quantitative Drug Design*, Pergamon, Elmsford, NY, **1990,** p. **459.**
45. G. C. K. Roberts, *Pharmacochem. Libr.*, **6, 91** (1983).
46. A. A. Kumar, J. H. Mangum, D. T. Blankenship, and J. H. Freisheim, *J. Biol. Chem.*, **266, 8970 (1981).**

47. G. D. *Rose and R. Wolfenden, Annu. Rev. Biophys. Biomol. Struct.*, 22,381 (1993).
48. A. T. *Hagler, P. Dauber, and S. Lifson, J. Am. Chem. Soc.*, 101,5131 (1979).
49. W. *Kauzmann, Adv. Protein Chem.*, 14, 1 (1959).
50. A. *Ben-Naim, Pure Appl. Chem.*, 69, 2239 (1997).
51. W. Blokzijl *and J. B. F. N. Engberts, Angew. Chem. Znt. Ed. Engl.*, 32, 1545 (1993).
52. *N. Muller, Acc. Chem. Res.*, **23**, 23 (1990).
53. F. *Eisenhaber, Perspect. Drug Discov. Des.*, 17, 27 (1999).
54. A. R. *Fersht, J. S. Shindler, and W. C. Tsui, Biochemistry*, 19,5520 (1980).
55. P. R. *Andrews, D. J. Craik, and J. L.* Matin, *J. Med. Chem.*, 27,1648 (1984).
56. *N.* R. *Draper and H. Smith, Applied Regression Analysis, 2nd ed., John Wiley & Sons, New York*, 1981.
57. *Y. Martin in G.* Grunewald, *Ed., Quantitative Drug Design, Marcel Dekker, New York*, 1978, *p.* 167.
58. *H. Kubinyi in R. Mannhold, P.* Krogsgaard-*Larsen, and H. Timmerman, Eds., QSAR: Hansch Analysis and Related Approaches, VCH, New York*, 1993, *p.* 91.
59. *R. Franke in W. Th. Nauta and R. F. Rekker, Eds., Theoretical Drug Design Methods,* Elsevier *Science, Amsterdam/New York*, 1983, *p.* 395.
60. *C. Hansch in C. J.* Cavallito, *Ed., Structure Activity Relationships, Vol. 1, Pergamon, Oxford, UK*, 1973, *p.* 75.
61. *J.* K. *Seydel, Int. J. Quantum Chem., 20*, 131 (1981).
62. *J. G. Topliss and R. P. Edwards, J. Med. Chem.*, 22, 1238 (1979).
63. *P.* N. *Craig, J. Med. Chem.*, 14, 680 (1971).
64. J. *G. Topliss, J. Med. Chem.*, 15,1006 (1972).
65. J. *G. Topliss, J. Med. Chem.*, 20,463 (1977).
66. *T. M. Bustard, J. Med. Chem.*, 17, 777 (1974).
67. *F. Darvas,* J. *Med. Chem.*, 17, 799 (1974).
68. *P. S.* Magee *in* J. Miyamoto *and P. C. Kearney, Eds., Pesticide Chemistry: Human Welfare and Environment, Proceedings of the International Congress on Pesticide Chemistry, Vol.* 1, Pergamon, *Oxford, UK*, 1983, *p.* 251.
69. *T.* J. *Mitchell, Technometrics*, 16, 203 (1974).
70. T. *Moon, M. H. Chi, D. H. Kim,* C. *N.* Yoon, *and Y. S. Choi, Quant. Struct.-Act. Relat.*, 19, 257 (2000).
71. M. Baroni, *S. Clernenti, G.* Cruciani, *N.* Kettaneh-*Wold, and S. Wold, Quant. Struct.-Act. Relat.*, 12, 225 (1993).
72. M. *Sjostrom and* L. *Eriksson in H.* van *de Waterbeemd, Ed., Chemometric Methods in Molecular Design, VCH, Weinheim, Germany*, 1995, *p.* 63.
73. *L. Eriksson, E. Johansson, M. Muller, and S. Wold, Quant. Struct.-Act. Relat.*, 16, 383 (1997).
74. *L. Eriksson,* E. *Johansson, M. Muller, and S. Wold, J. Chemom.*, 14,599 (2000).
75. *C. Hansch* and *T. Fujita, J.* Am. *Chem.. Soc.*, 86, 1616 (1964).
76. *C-QSAR Database,* BioByte Corp., *Claremont, CA.*
77. *G.* N. *Burckhardt, W. G.* K. *Ford, and E.* Singelton, *J. Chem. Soc.*, 17 (1936).
78. L. P. *Hammett, J. Chem. Ed.*, 43,464 (1966).
79. M. Charton, *Prog. Phys. Org. Chem., 8*, 235 (1971).
80. *T. Fujita and T. Nishioka, Prog. Phys. Org. Chem.*, **12**, 49 (1976).
81. *P.* D. Bolton, *K. A. Fleming, and F. M. Hall, J. Am. Chem. Soc.*, 94,1033 (1972).
82. K. Kalfus, *J. Kroupa, M. Vecera, and O. Exner, Collect. Czech. Chem. Commun.*, 40, 3009 (1975).
83. *M.* Bergon *and J. P.* Calmon, *Tetrahedron Lett.*, 22, 937 (1981).
84. *J. Schreck,* J. *Chem. Ed.*, 48, 103 (1971).
85. *H. C. Brown and Y.* Okamoto, *J. Am. Chem. Soc.*, 80,4979 (1958).
86. *Y. Tsuno, T. Ibata,* and Y. Yukawa, *Bull. Chem. Soc. Jpn.*, 32,960,965,971 (1959).
87. J. *D. Roberts and W. T. Moreland, J. Am. Chem. Soc.*, 75,2167 (1953).
88. K. Bowden *in* C. *A.* Ramsden, *Ed., Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds, Vol. 4: Quantitative Drug Design, Pergamon,* Elmsford, NY, 1990, *p.* 212.
89. A. *Albert, Selective Toxicity: The Physicochemical Bases of Therapy, 7th ed., Chapman and Hall, London*, 1985, *p.* 379.
90. *M.* Karelson, V. *S. Lobanov, and A. R.* Katritzky, *Chem. Rev.*, 96, 1027 (1996).
91. P. S. Magee *in ACS Symposium Series 37, American Chemical Society, Washington, DC*, 1980.
92. *S.* P. *Gupta, Chem. Rev.*, 91,1109 (1991).

93. J. J. Sullivan, A. D. Jones, and K. K. Tangi, *J. Chem. Int. Comput. Sci.*, **40, 1113** (2000).

94. M. Cocchi, M. C. Menziani, F. Fanelli, P. G. Debenedetti, **J.** *Mol. Struct.*, **331, 79 (1995).**

95. M. Cocchi, M. Menziani, P. G. Debenedetti, A. Cruciani, *Chemom. Intell. Lab. Sys.*, **14, 209 (1992).**

96. **J.** H. Hildebrand, ***Proc. Natl. Acad. Sci. USA,* 76,194 (1979).**

97. G. D. Rose, A. R. Geselowitz, G. **J.** Lesser, R. H. Lee, and M. H. Zehfus, ***Science,* 229, 834 (1985).**

98. **H.** J. Schneider, *Angew. Chem. Int. Ed. Engl.*, **30,1417 (1991).**

99. J. N. Israelachvili and H. Wennerstrom, **J.** *Phys. Chem.*, **96, 520 (1992).**

100. **J. J.** H. Nusselder and **J.** B. F. N. Engberts, *Langmuir*, **7, 2089 (1991).**

101. P. **J.** Taylor in C. A. Ramsden, Ed., ***Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds, Vol. 4, Quantitative Drug Design,*** Pergamon, Elmsford, NY, **1990,** p. **241.**

102. **J.** H. Hildebrand, **J.** *Phys. Chem.*, **72, 1841** (1969).

103. H. S. Frank and M. W. Evans,*J. Chem. Phys.*, **13,507 (1945).**

104. G. Nemethy and H. A. Scheraga, **J.** *Chem. Phys.*, **36,3382 (1962).**

105. A. D. **J.** Haymet, K. A. T. Silverstin, and K. A. Dill, *Faraday Discuss.*, **103, 117 (1996).**

106. **K.** A. T. Silverstein, K. A. Dill, and A. D. **J.** Haymet, **J.** *Chem. Phys.*, **114,6303 (2001).**

107. A. **J.** Leo and C. Hansch, ***Perspect. Drug Discov. Des.*, 17, 1** (1999).

108. R. N. Smith, C. Hansch, and M. ***A.*** Ames, **J.** *Pharm. Sci.*, **64, 599 (1975).**

109. A. Leo and C. Hansch, **J.** *Org. Chem.*, **36,1539 (1971).**

110. B. C. Lippold and M. S. Adel, *Arch. Pharm.*, **305,417 (1972).**

111. S. **E.** Debolt and P. A. Kollman, **J.** *Am. Chem. Soc.*, **117, 5316 (1995).**

112. A. Leo, **J.** *Pharm. Sci.*, **76, 166** (1987).

113. A. Leo, *Methods Enzymol.*, **202,544 (1991).**

114. J. de Bruijn and **J.** Hermens, *Quant. Struct.-Act. Relat.*, **9, 11 (1990).**

115. E. Tomlinson, S. S. David, G. D. Parr, M. James, N. Farraj, **J.** F. M. Kinkel, D. Gaisser, and H. **J.** Wynn in W. **J.** Dunn **III,** J. H. Block, and R. S. Pearlman, Eds., ***Partition Coefficient, Determination and Estimation,*** Pergamon, Oxford, ***UK*, 1986,** p. **83.**

116. R. Collander, *Acta Chem. Scand.*, *5*, **774 (1951).**

117. A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, **71,525** (1971).

118. R. F. Rekker, ***The Hydrophobic Fragmented Constant. Its Derivation and Application: A Means of Characterizing Membrane Systems,*** Elsevier, Amsterdam, **1977,** p. **131.**

119. P. Seiler, *Eur.* **J.** *Med. Chem.*, **9,473 (1974).**

120. T. Fujita, T. Nishioka, and M. Nakajima, *J. Med. Chem.*, **20, 1071 (1977).**

121. D. E. Leahy, P. **J.** Taylor, and A. R. Wait, *Quant. Struct.-Act. Relat.*, **8, 17 (1989).**

122. **J.** C. Dearden, A. M. Patel, and J. M. Thubby, **J.** *Pharm. Pharmacol.*, **26** (Suppl.), **75P (1974).**

123. **W.** Draber, K. H. Buchel, and K. Dickore, ***Proc. Int. Congr. Pest. Chem., 2nd ed., 1971, 5*, 153 (1972).**

124. P. Vallat, N. El Tayar, B. Testa, I. Slacanin, A. Martson, and K. Hostettmann, **J.** *Chromatogr.*, **504,411 (1990).**

125. A. Berthod, Y. I. Han, and D. W. Armstrong, **J.** *Liq. Chromatogr.*, **11, 1441 (1988).**

126. A. Berthod, S. Carola-Broch, and M. C. Garcia-Alvarex-Cogne, *Anal. Chem.*, **71,879 (1999).**

127. K. Valko, **C.** Beran, and D. Reynolds, *Anal. Chem.*, **69,2022 (1997).**

128. ***K.*** Valko, C. M. Du, C. Bevan, D. P. Reynolds, and M. H. Abraham, *Curr. Med. Chem.*, **8, 1137 (2001).**

129. F. Lombardo, **M.** Y. Shalaeva, K. A. Tupper, F. Gao, and M. H. Abraham, **J.** *Med. Chem.*, **43, 2922 (2000).**

130. J. L. Fauchére and V. Pliska, *Eur.* **J.** *Med. Chem.*, **18,369** (1983).

131. *J.* L. Fauchére in B. Testa, Ed., ***Advances in Drug Research,*** Vol. **15,** Academic Press, London/New York, **1986,** p. **29.**

132. M. Akamatsu, Y. Yoshida. H. Nakamura, M. Asao, H. Iwamura, and T. Fujita, *Quant. Struct.-Act. Relat.*, **8, 195 (1989).**

133. **C.** Hansch, A. Leo, and D. Hoekrnan in S. R. Heller, Ed., ***Exploring QSAR: Hydrophobic, Electronic and Steric Constants,*** Vol. **2,** American Chemical Society Professional Reference Book, Washington, DC, **1995.**

134. G. G. Nys and R. F. Rekker, *Chim. Ther.*, **8, 521 (1973).**

135. A. Leo, P. Y. C. Jow, C. Silipo, and C. Hansch, **J.** *Med. Chem.*, **14,865 (1979).**

136. *D. Weininger,* **J.** *Chem. Znt. Comput. Sci.,* **28, 31 (1988).**

137. *D. Weininger, A. Weininger, and J. L.* Weininger, *J. Chem. Znt. Comput. Sci.,* **29, 97 (1989).**

138. *A. Leo in C. A.* Ramsden, *Ed., Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds, Vol. 4, Quantitative Drug Design, Pergamon, Elmsford,* NY, **1990,** *p.* **315.**

139. *A. Leo, personal communication.*

140. *A. Leo, Chem. Rev.,* **93, 1281 (1993).**

141. **A. J.** *Leo and D. Hoekman, Perspect. Drug Discov. Des.,* **18, 19 (2000).**

142. *H. van de Waterbeemd and R. Mannhold, Quant. Struct.-Act. Relat.,* **15, 410 (1996).**

143. **R.** *Mannhold and H. van de Waterbeemd, J. Cornput.-Aided Mol. Des.,* **15,337 (2001).**

144. **R. F.** *Rekker and H. M.* DeKort, *Eur. J. Med. Chem.,* **14,479 (1979).**

145. *G. Klopman, J. W. Li, S. Wang, and M.* Dimayuga, *J. Chem. Inf. Comput. Sci.,* **34, 752 (1994).**

146. *A.* **K.** *Ghose and G. M. Crippen, J. Med. Chem.,* **28,333 (1985).**

147. *T. Suzuki and Y. Kudo, J. Cornput.-Aided Mol. Des.,* **4, 155 (1990).**

148. **I.** *Moriguchi, S. Hirono, Q. Liu, I. Nakagome, and Y. Matsushita, Chem. Pharm. Bull.,* **40, 127 (1992).**

149. *G. E. Kellogg, G. J. Joshi, and D. J. Abraham, J. Med. Chem. Res.,* **1,444 (1992).**

150. J. Devillers, *D.* Domine, *C. Guillon, and W. J.* Karcher, *J. Pharm. Sci.,* **87, 1086 (1998).**

151. *M.* **J.** *Kamlet, P. W.* Carr, *R. W. Taft, and M. H. Abraham, J. Am. Chem. Soc.,* **103, 6062 (1981).**

152. *M.* **J.** *Kamlet, J. L. Abboud, M. Abraham, and R. Taft, J. Org. Chem.,* **48,2877 (1983).**

153. **J.** *A. Platts, D. Butina, M. H. Abraham, and A. Hersey, J. Chem. Inf. Comput Sci.,* **39, 835 (1999).**

154. *Y.* Ishihama *and N. Asakawa, J. Pharm. Sci.,* **88, 1305 (1999).**

155. **J.** *A. Platts, M. H. Abraham, D. Butina, and A. Hersey, J. Chem. Inf. Comput. Sci.,* **40, 71 (2000).**

156. *A.* **J.** *Leo, J. Pharm. Sci.,* **89, 1567 (2000).**

157. **R.** *W. Taft in M. S.* Newman, *Ed., Steric Effects in Organic Chemistry, John Wiley & Sons, New York,* **1956,** *p.* **556.**

158. *K. Hancock, E. A. Meyers, and B. J. Yager, J. Am. Chem. Soc.,* **83,4211 (1961).**

159. *M.* Charton *in M.* Charton *and I. Motoc, Eds., Steric Effects in Drug Design, Springer, Berlin,* **1983,** *p.* **57.**

160. *M. S. Tute in C. A.* Ramsden, *Ed., Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds, Vol. 4, Quantitative Drug Design, Pergamon,* Elmsford, NY, **1990,** *p.* **18.**

161. *C. Hansch and T. Klein, Acc. Chem. Res.,* **19, 392 (1986).**

162. *A. Verloop, W. Hoogenstraaten, and J. Tipker in E. J. Ariens, Ed., Drug Design, Vol.* VII, *Academic Press, New* York/London, **1976,** *p.* **165.**

163. *A. Verloop, The STERIMOL Approach to Drug Design, Marcel Dekker, New York,* **1987.**

164. *C. Hansch, D. Hoekman, A. Leo, D. Weininger, and C. D. Selassie, unpublished results.*

165. *V. A.* Levin, *J. Med. Chem.,* **23, 682 (1980).**

166. *E.* **J.** *Lien and P. H. Wang, J. Pharm. Sci.,* **69, 648 (1980).**

167. *C. D. Selassie, C. Hansch, and T. Khwaja, J. Med. Chem.,* **33,1914** (1990).

168. *E.* **J.** *Lien, L. L. Lien, and H. Gao in F. Sanz, J. Guiraldo, and F. Manaut, Eds., QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications,* Prous *Science,* Barcelona/Philadelphia, **1995,** *p.* **94.**

169. *C. Selassie, unpublished results.*

170. M. Recanatini, *T. Klein, C. Z. Yang, J.* McClarin, *R. Langridge, and C. Hansch, Mol. Pharmacol.,* **29, 436 (1986).**

171. *Y.* Naito, *M.* Sugiura, *Y. Yamamura, C. Fukaya, K. Yokoyama, Y.* Nakagawa, *T.* Ikeda, *M. Senda, and T. Fujita, Chem. Pharm. Bull.,* **39, 1736 (1991).**

172. *A. K. Debnath, R. L. L. de Compadre, G.* Debnath, *A. J. Shusterman, and C. Hansch, J. Med. Chem.,* **34,786 (1991).**

173. *M.* Randic, *J. Am. Chem. Soc.,* **97,6609 (1975).**

174. **L.** *B. Kier and L. H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New* York/London, **1976.**

175. *L.* **B.** *Kier and M. H. Hall, J. Pharm. Sci.,* **72, 1170 (1983).**

176. *L. H. Hall and L. B. Kier, J. Pharm. Sci.,* **64, 1978** (1975).

177. J. Gough *and L. H. Hall, J. Chem. Inf. Comput. Sci.,* **39, 356 (1999).**

178. *J. K. Boulamwini, K. Raghavan, M. Fresen, Y. Pommier, K. Kohn, and J. Weinstein, Pharm. Res.*, **13, 1892 (1995).**

179. **V.** *E. F. Heinzen, V. Cechinel, and R. A. Yunes, Farmaco*, **54, 125 (1999).**

180. *R. L. Lopez de Compadre, C. M. Compadre, R. Castillo, and W. J.* **DunnIII**, *Eur. J. Med. Chem.*, **18, 569 (1983).**

181. **H.** *Kubinyi, Quant. Struct.-Act. Relat.*, **14, 149 (1995).**

182. *P.* **A.** *J. Janssen and N. B. Eddy, J. Med. Pharm. Chem.*, **2**, 31 (1960).

183. *T. Fujita in C. A.* **Ramsden**, *Ed., Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds, Vol. 4, Quantitative Drug Design, Pergamon, Elmsford,* **NY, 1990,** *p.* **503.**

184. *C. Hansch, D. Kim, A. J. Leo, E. Novellino, C. Silipo, and A.* **Vittoria**, *CRC Crit. Rev. Toxicol.*, **19, 185 (1989).**

185. *C. Hansch and W. J. DunnIII, J. Pharm. Sci.*, **61,** 1 (1972).

186. *T. W. Schultz and M. Tichy, Bull. Environ. Contam. Toxicol.*, **51, 681 (1993).**

187. *J. T. Penniston, L. Beckett, D. L. Bentley, and C. Hansch, Mol. Pharmacol.*, **5, 333 (1969).**

188. *C. Hansch, Adv. Chem. Ser.*, **114, 20 (1972).**

189. *C. Hansch and J. M. Clayton, J. Pharm. Sci.*, **62,** 1 (1973).

190. *R. Franke and W. Schmidt, Acta Biol. Med. Germ.*, **31, 273 (1973).**

191. **J. McFarland, J.** *Med. Chem.*, **13, 1192 (1970).**

192. *H. Kubinyi and 0. H. Kehrhahn, Arzneim.-Forsch.*, **28, 598 (1978).**

193. *H. Kubinyi, Arzneim.-Forsch.*, **29, 1067 (1979).**

194. *R. Franke in W. Th. Nauta and R. F.* **Rekker**, *Eds., Theoretical Drug Design Methods, Elsevier, New York,* **1984,** *p.* **256.**

195. *H. Kubinyi in C. A.* **Ramsden**, *Ed., Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds, Vol. 4, Quantitative Drug Design, Pergamon, Elmsford,* **NY, 1990,** *p.* **539.**

196. *C. John* **Blankley** *in J. G. Topliss, Ed., Quantitative Structure Activity Relationships of Drugs, Academic Press, New York,* **1983,** *p.* **5.**

197. *E.* **Yalcin**, *S. E. Sener, I.* **Owren**, *and O. Temiz in E. Sanz, J. Giraldo, and F. Manaut, Eds., QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications,* **Prous** *Science,* **Barcelona/Philadelphia, 1995,** *p.* **147.**

198. *J. Kunes, J. Jachym, P. Tirasko, Z. Odlerova, and K.* **Waisser**, *Collect. Czech. Chem. Commun.*, **62, 1503 (1997).**

199. *Y. Terada and K. Naya, Pharmazie*, **55, 133 (2000).**

200. *S. P. Gupta and A.* **Paleti**, *Bioorg. Med. Chem.*, **6, 2213 (1998).**

201. *C. Tmej, P. Chiba, M. Huber, E. Richter, M. Hitzler, K. J. Schaper, and G. Ecker, Arch. Pharm.*, **331, 233 (1998).**

202. *C. Selassie and T. E.* **Klein** *in J. Devillers, Ed., Comparative QSAR, Taylor & Francis, Washington, DC,* **1998,** *p.* **235.**

203. *C. D. Selassie, W. X. Gan, M. Fung, and R. Shortle in F. Sanz, J. Giraldo, and F. Manaut, Eds., QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications,* **Prous** *Science,* **Barcelona/Philadelphia,** *1995, p.* **128.**

204. **A. Kurup, R.** *Garg, and C. Hansch, Chem. Rev.*, **100, 909 (2000).**

205. *J. M. Blaney, C. Hansch, C. Silipo, and A.* **Vittorio**, *Chem. Rev.*, **84, 333 (2000).**

206. *C. Hansch, Ann. N. Y. Acad. Sci.*, **186, 235 (1971).**

207. *C. Hansch, B. A. Hathaway, Z. R. Guo, C. D. Selassie, S. W. Dietrich, J. M. Blaney, R.* **Langridge**, *K. W. Volz, and B. T. Kaufman, J. Med. Chem.*, **27, 129 (1984).**

208. *B.* **A.** *Hathaway, Z.* **R.** *Guo, C. Hansch, T. J. Delcamp, S. S. Susten, and J. H. Freisheim, J. Med. Chem.*, **27, 144 (1984).**

209. *C. D. Selassie, C. D. Strong, C. Hansch, T.* **Delcamp**, *J. H. Freisheim, and T. A. Khwaja, Cancer Res.*, **46, 744 (1986).**

210. *C.* **K. Marlowe**, *C. D. Selassie, and D. V. Santi, J. Med. Chem.*, **38, 967 (1995).**

211. *R.* **G.** *Booth, C. D. Selassie, C. Hansch, and D. V. Santi, J. Med. Chem.*, **30, 1218 (1987).**

212. **L.** *C. Chio and S. F. Queener, Antimicrob. Agents Chemother.*, **37, 1916** (1993).

213. *J. Y.* **Fukunaga**, *C. Hansch, and E. E. Stellar, J. Med. Chem.*, **19, 605 (1976).**

214. *B. K. Chen, C. Horvath, and J. R. Bertino, J. Med. Chem.*, **22, 483 (1979).**

215. *N.* **V.** *Harris, C. Smith, and K.* **Bowden**, *Eur. J. Med. Chem.*, **27, 7 (1992).**

216. *A. M. Klibanov, Nature*, **409, 241 (2001).**

217. *K. H. Kim, J. Comput.-Aided Mol. Des.*, **15, 367** (2001).

218. **K.** *H. Kim, Bioorg. Med. Chem.*, **9, 1951 (2001).**

219. *K. Nakamura, K.* **Hayashi**, *I.* **Ueda**, *and H.* **Fujiwara**, *Chem. Pharm. Bull.*, **43, 369 (1995).**

220. C. M. Compadre, R. J. Sanchez, C. Bhuraneswaran, R. L. Compadre, D. Plunkett, and S. G. Novick in C. G. Wermuth, Ed., ***Trends in QSAR and Molecular Modelling,*** Escom, Strasbourg, France, **1993,** p. **112.**

221. S. V. Frye, C. D. Haffner, P. R. Maloney, R. A. Mook, Jr., G. F. Dorsey, R. N. Hiner, C. M. Cribbs, T. N. Wheeler, J. A. Ray, R. C. Andrew~K. W. Batchelor, H. N. Branson, J. D. Stuart, S. L. Schwiker, J. Van Arnold, S. Croom, D. M. Bickett, M. L. Moss, G. Tian, R. J. Unwalla, F. W. Lee, T. K. Tippin, M. K. James, M. K. Grizzle, J. E. Long, and S. V. Schuster, *J. Med. Chem.*, **37, 2352(1994).**

222. S. V. Frye, C. D. Haffner, P. R. Maloney, R. N. Hiner, G. F. Dorsey, R. A. Roe, R. J. Unwalla, K. W. Batchelor, H. N. Branson, J. D. Stuart, S. L. Schwiker, J. Van Arnold, D. M. Bickett, M. L. Moss, G. Tian, F. W. Lee, T. K. Tippin, M. K. James, M. K. Grizzle, J. E. Long, and D. K. Croom, *J. Med. Chem.*, **38, 2621**(1995).

223. M. T. D. Cronin and J. C. Dearden, *Quant. Struct.-Act. Relat.*, **14, 518** (1995).

224. M. T. D. Cronin, B. W. Gregory, and T. W. Schultz, ***Chem. Res. Toxicol.*, 11, 902** (1998).

225. T. W. Schultz, ***Chem. Res. Toxicol.*, 12, 1262** (1999).

226. M. T. D. Cronin and T. W. Schultz, *Chemosphere*, **32, 1453 (1996).**

227. R. Garg, A. Kurup, and C. Hansch, ***Crit. Rev. Toxicol.*, 31, 223 (2001).**

228. C. D. Selassie, T. V. DeSoyza, M. Rosario, H. Gao, and C. Hansch, *Chem.-Biol. Interact.*, **113, 175(1998).**

229. M. T. D. Cronin and T. W. Schultz, ***Chem. Res. Toxicol.*, 14, 1284** (2001).

230. P. H. Hinderling, O. Schmidlin, and J. K. Seydel, J. ***Pharmacokinet. Biopharm.*, 12, 263 (1984).**

231. C. Selassie and T. E. Klein in H. Kubinyi, Ed., ***3D QSAR in Drug Design. Theory, Methods and Applications,*** Escom Science, Leiden, The Netherlands, **1993,** p. **257.**

232. O. Geban, H. Ertepinar, M. Yurtsever, S. Ozden, and F. Gumus, ***Eur. J. Med. Chem.*, 34, 753 (1999).**

233. S. Daunes, C. D'Silva, H. Kendrick, V. Yardley, and S. L. Croft, ***J. Med. Chem.*, 44, 2976(2001).**

234. C. Hansch, H. Gao, and D. Hoekman in J. Devillers, Ed., ***Comparative QSAR,*** Taylor & Francis, Washington, DC, **1998,** p. **285.**

235. C. Hansch, B. R. Telzer, and L. Zhang, ***Crit. Rev. Toxicol.*, 25,** 67 (1995).

236. R. Garg, S. Kapur, and C. Hansch, ***Med. Res. Rev.*, 21, 73** (2000).

237. L. Zhang, H. Gao, C. Hansch, and C. Selassie, *J. Chem. Soc. Perkin Trans.* **2, 2553**(1998).

238. C. Hansch, S. McKarns, C. J. Smith, and D. J. Doolittle, *Chem.-Biol. Interact.*, **127, 61** (2000).

239. L. A. Oglesby, M. T. Ebon-McCoy, T. R. Logsdon, F. Copeland, P. E. Beyer, and R. J. Kavlock, *Teratology*, **45, 11 (1992).**

240. C. Hansch and H. Gao, ***Chem. Rev.*, 97, 2995** (1997).

241. A. M. Richard, J. K. Hongslo, P. F. Boone, and J. A. Holme, ***Chem. Res. Toxicol.*, 4, 151(1991).**

242. C. D. Selassie, A. J. Shusterman, S. Kapur, R. P. Verma, L. Zhang, and C. Hansch, *J. Chem. Soc. Perkin Trans.* **2, 2729(1999).**

243. D. Boyd in A. L. Parrill and M. Rami-Reddy, Eds., ***Rational Drug Design, ACS Symposium Series 719,*** American Chemical Society, Washington, DC, **1999,** p. **346.**

244. E. Plummer in C. Hansch and T. Fujita, Eds., ***Classical and Three-Dimensional QSAR in Agrochemistry, ACS Symposium Series*** 606, American Chemical Society, Washington, DC, **1995,** p. **241.**

245. T. Fujita, *Quant. Struct.-Act. Relat.*, **16, 107** (1997).

246. H. Koga, A. Itoh, S. Murayarna, S. Suzue, and T. Irikura, ***J. Med. Chem.*, 23, 1358 (1980).**

247. H. Chuman, A. Ito, T. Shaishoji, and S. Kumazawa in C. Hansch and T. Fujita, Eds., ***Classical and Three-Dimensional QSAR in Agrochemistry, ACS Symposium Series*** 606, American Chemical Society, Washington, DC, **1995,** p. **171.**

248. J. Ohtaka and G. Tsukamoto, ***Chem. Pharm. Bull.*, 35, 4117 (1987).**

249. M. Kuchar, E. Maturova, B. Brunova, J. Grimova, H. Tomkova, and K. J. Holubek, ***Collect. Czech. Chem. Commun.*, 53, 1862** (1988).

250. T. Fujita in G. Jolles and K. R. H. Wooldridge, Eds., ***Drug Design: Fact or Fantasy,*** Academic Press, London, **1984,** p. **19.**

251. J. G. Topliss, ***Perspect. Drug Discov. Des.*, 1, 253 (1993).**

252. C. Hansch, J. P. Bjorkroth, and A. Leo, *J. Pharm. Sci.*, **76, 663 (1987).**

253. C. Hansch, R. Garg, and A. Kurup, ***Bioorg. Med. Chem.*, 9, 283** (2001).

254. R. Garg, A. Kurup, S. B. Mekapati, and C. Hansch, ***Bioorg. Med. Chem.,*** in press (2002).