Journal of Medicinal Chemistry

Volume 24, Number 3 March 1981

Perspective

A Practitioner's Perspective of the Role of Quantitative Structure-Activity Analysis in Medicinal Chemistry

Yvonne C. Martin

Department of Medicinal Chemistry, Pharmaceutical Products Division, Abbott Laboratories, North Chicago, Illinois 60064. *Received October 31, 1980*

This article is written from the point of view of a practitioner of QSAR (quantitative structure activity relationships) methods for the past decade in an industrial setting. It will examine the utility of QSAR in a medicinal chemistry research program. In particular, attention will be paid to its special advantages compared to a more qualitative analysis of data and to its current and its inherent limitations. A number of questions will underly the discussion: Is it an advantage to statistically analyze apparent relationships? Does QSAR ever suggest new directions of analogue design? How often have its predictions been correct? What are the advantages of a statistically based design of a set of analogues? Does QSAR have provisions for helping the scientist decide that a series based on a lead has been explored enough? How does it tie into receptor mapping and conformational analysis? Are there generalizations that can be made with respect to the applicability of a QSAR from one series to that for the same biological activity with another series? What are the limitations of QSAR? What is its future? How should it be integrated into a research program? These are important questions to consider; it is not guaranteed that an answer or even a strong opinion will be presented here.

What Is QSAR

QSAR methods are characterized by two assumptions with respect to the relationship between the chemical structure and the biological potency of a compound. The first is that one can derive a *quantitative* measure from the *structure* of those global and local properties of significance to the biological *activity* of a compound. The properties are usually assumed to be either physicochemical (such as pK_a or partition coefficient) or substructural (such as the presence or absence of certain chemical features, for example, CO_2R or SH). The other assumption is that one can mathematically describe the *relationship* between the biological property one wishes to optimize and the molecular properties calculated from the structure. In other words, the two characteristics of QSAR methods are (1) a method to transform the chemical structure of a compound into a set of numerical descriptors of the properties relevant to biological activity, and (2) a method to establish the quantitative relationship between these descriptors and the biological properties of the compounds.

The most popular QSAR method is the linear free energy, extrathermodynamic, or Hansch method.[1] Its basic assumption is that the effect of substituents on the strength of interactions between a drug and its receptor and other biomolecules is an additive combination of the effects of the substituents on various types of simpler model intermolecular interactions. From physical chemistry the interactions are assumed to be electrostatic, steric (repulsion), hydrophobic, and dispersion in nature.

The substituent effects on these noncovalent interactions are assumed to be proportional to the tabulated values of Hammett σ, Taft E_s, Hansch π, and molar refractivity values of the substituents, respectively. The Hammett σ value of a substituent is the logarithm of the effect of that substituent on the acid dissociation constant of benzoic acid. The Taft E_s value is experimentally derived from the relative rates of hydrolysis of esters, but for spherically symmetric substituents it is proportional to the radius of the substituent. The Hansch π value is the effect of the substituent on the logarithm of the octanol–water partition coefficient P. (Unless explicitly stated otherwise, all references in this paper to partition coefficient and log P are references to the value for the neutral form: it frequently complicates interpretations of data if ionization and partitioning equilibria are mixed into the hybrid constant, apparent log P at some pH.[2]) Molar refractivity, derived from the refractive index, parameterizes dispersion interactions. Many other descriptors of molecules have been proposed: values of substituent constants and their experimental basis are available,[3] as are a discussion of how to apply the Hansch methodology[4] and recent advances in the methodology.[5] The data base of π, σ, and E_s constants is nowhere near complete.

The proportionality assumption stated above leads to the mathematical expression of the Hansch equation (eq 1). The log $(1/C)$ term is the relative potency of the

(1) (a) Hansch, C. *J. Med. Chem.* **1976**, *19*, 1; (b) Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616.

(2) Martin, Y. C.; Hackbarth, J. J. *J. Med. Chem.* **1976**, *19*, 1033.

(3) Hansch, C.; Leo, A. J. "Substituent Constants for Correlation Analysis in Chemistry and Biology"; Wiley: New York, 1979.

(4) Martin, Y. C. "Quantitative Drug Design. A Critical Introduction"; Marcel Dekker: New York, 1978.

(5) Martin, Y. C. *Drug Des.* **1979**, *8*, 1.

0022-2623/81/1824-0229$01.00/0

$$\log (1/C) = a + b\pi + cE_s + d\mathrm{MR} + e\sigma \quad (1)$$

analogue, in which C is the concentration required to produce some standard biological response. The values of the coefficients a, b, c, and d in eq 1 are fit by least-squares multiple regression analysis. This statistical method also provides a measure of (1) whether each coefficient is significantly different from zero; (2) the value of the R^2 statistic, which is the fraction of the variance in the log $(1/C)$ data that is explained by the equations; and (3) the value of s, which is the standard deviation of the observed log $(1/C)$ values from those calculated. For the statistical part of a QSAR analysis, one examines the possible equations of a data set to find those of interest, that is, equations that contain only statistically significant terms that make mechanistic sense and that do not overfit the data.

One of the earliest observations in QSAR was that the relationship between potency and log P is not necessarily linear.[1] The first QSAR equation was parabolic in log P or π, that is, one that contains a π and π^2 term (eq 2). The basis for this was a random-walk assumption. It leads to an optimum log P, log P_0.

$$\log (1/C) = a + b\pi + c\pi^2 + dE_s + e\mathrm{MR} + f\sigma \quad (2)$$

Almost simultaneously with the first use of the Hansch method of QSAR was the publication of the Free–Wilson method.[6] In this procedure molecules are described not in terms of physicochemical properties but rather by a set of indicator or all-or-none variables that denote the presence or absence of certain groups. The underlying assumption is that adding a particular substituent at a particular position on the molecule always results in a quantitatively similar effect on biological potency of the rest of the molecule. This assumption is a quantitation of the more traditional approach to SAR. The assumptions are put into equation form (eq 3). In this equation, i is

$$\log (1/C) = x + \sum^{mn} a_{ij}G_{ij} \quad (3)$$

the number of the position of substitution, j is the number of the substituent at that particular position, m is the total number of positions of substitution, and n is the number of substituents. The value of the a_{ij} term indicates the presence (1.0) or absence (0.0) of the substituent ij. The group contribution values G_{ij} are established by multiple regression analysis. For example, substituting a 4″-*O*-acetyl for the usual –OH in erythromycin decreases the activity vs. *S. aureus* by 0.69 (±0.04).[7]

This method can be used only for series with multiple sites of substitution and only if each substituent that occurs at any position is present at that position in at least two compounds of the series. In some cases the group contributions are correlated with the physicochemical properties of the substituents. The method is further discussed in ref 7. It has the advantage that one does not need physicochemical properties of the compounds and the disadvantage that one cannot make predictions based on such properties. It does not fit a data set if there is an optimum of some global property, since then the assumption of the independent contribution of substituents at each position would not hold.

Hybrid Hansch/Free–Wilson equations are constructed by the addition to eq 2 of indicator variables that describe the presence or absence in the molecule of certain groups or features such as hydrogen-bond acceptors. These indicator variables quantitate that effect of a substituent on biological activity that cannot be attributed to the physicochemical properties considered. In those cases in which the biological measurement is receptor binding, the coefficients of the indicator variables measure fit to a receptor uncomplicated by (because they have been statistically removed) differences in binding due to differences in physicochemical properties.

A characteristic of the QSAR methods is that they are relatively inexpensive and easy to do. The computer and personnel time and resources required are at least one and probably two orders of magnitude less than that required for quantum chemical, conformational, computer graphics oriented methods of structure–activity analyses. For this reason, QSAR methods are available to most medicinal chemists, whereas the others may not be.

Immediate Implications of the First QSAR Papers[1b]

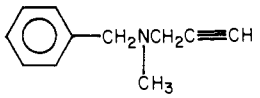
Historical Perspective: Appreciation of the Relationship between Physical and Biological Properties of Molecules in 1961. In order to discuss the accomplishments of QSAR, one must first appreciate the insight that scientists had into the relationship between physical and biological properties when the first papers on QSAR were written. First, some scientists appreciated the importance of partition coefficient to drug potency. In the U.S., Fieser[8] and Brodie[9] had published such data using measured log P values. Fieser showed a constant log P_0 from series to series and also used this observation to correctly predict the activity of new analogues. Brodie and his group showed in tables the relationship between the various processes of absorption, distribution, metabolism, and excretion and lipophilicity. Pauling had recently attributed the anesthetic potency of various compounds to their ability to structure water.[10] Finally, a 1961 statement by Burger may be quoted: "for, without proper solubility and partition, biochemical interaction is unlikely".[11] Hydrophobicity was a concept that was well established in biochemistry at that time, but its importance was not a central point of new compound design by medicinal chemists.

By 1960 scientists also appreciated the importance of other physicochemical properties to drug potency. For example, Bell and Roblin showed that there is a pK_a optimum for the antibacterial activity of sulfonamides, the glamour drugs of the day.[12] They also correctly predicted the potency of two new analogues from this relationship. Albert also discussed in detail the relationship between pK_a, redox potential, log P, or metal binding and biological activity.[13] The contribution of QSAR is, thus, *not* in the recognition that there may be a relationship between physicochemical and biological properties of molecules but rather in how to discover, quantitate, and evaluate possible relationships.[1b]

For example, in 1959 W. Martin discovered the mono-

(6) Free, S. M.; Wilson, J. W. *J. Med. Chem.* **1964**, *7*, 395.
(7) Martin, Y. C. in ref 4, p 329.
(8) Fieser, L. F.; Ettlinger, M. G.; Fawaz, G. *J. Am. Chem. Soc.* **1948**, *70*, 3228.
(9) Brodie, B. B.; Hogben, C. A. *J. Pharm. Pharmacol.* **1957**, *9*, 345.
(10) Pauling, L. *Science* **1961**, *134*, 15.
Buznikov, G. A.; Kabankin, A. S.; Kolbanov, V. M.; Landau, M. A.; Aroyan, A. A.; Ovespyan, T. R.; Teplits, N. A. *Khim, Farm. Zh.* **1976**, *10*, 23–27.
(11) Burger, A.; *Proc. Int. Pharmacol. Meet., 1st* **1963**, *7*, 35.
(12) Bell, P. H.; Roblin, Jr., R. O. *J. Am. Chem. Soc.* **1942**, *64*, 2905.
(13) Albert, A. "Selective Toxicity", 2nd ed.; Methuen: London, 1960.

amine oxidase inhibitor pargyline (I) and prepared a



I

number of analogues.[14] Under the direction of J. D. Taylor, I measured their pK_a and in vitro potency (pI_{50}). Drs. Martin and Taylor arranged for the measurement of the chloroform–buffer log P and in vivo potency of some of the analogues. The three of us examined plots of the pI_{50}, the pI_{50} corrected to the concentration of the neutral form, the pI_{50} corrected to the concentration of the ionic form, and the in vivo potency plotted vs. pK_a, partition coefficient, molecular weight, and approximate solubility. By the time the group disbanded in 1960 we were frustrated that none of the physical properties that others had found to be useful were revealing for our series.

Application of Statistics to Test Hypotheses of QSAR. One important contribution of Hansch and Fujita is the demonstration that the relationships between physicochemical properties and biological activity can be examined statistically. Although today this insight seems obvious, it was not recognized by any of the earlier workers on drug–physical property relationships. The use of statistics allows one (1) to accept or reject apparent relationships, (2) to detect outliers, and (3) to compare the relative attractiveness of alternate apparent physical property–activity relationships.

Consideration of Multiple Physical Properties. Their second important contribution is the use of multivariate statistics. The result is that one statistically tests the possibility that more than one physical property may work in concert to influence potency.

In 1966 Dr. Taylor and I returned to the pargyline QSAR using the measured physical properties and multiple regression analysis. Our only significant correlation was a positive one between pI_{50} and the date of synthesis. Thus, the demonstrated usefulness of multiple regression analysis alone was not sufficient to help others solve their QSAR problems.

Parameterization of Log P as an Additive and Constitutive Property. The third important contribution of Hansch and Fujita was the recognition that partition coefficient is an additive constitutive property, that is, that it can be calculated from the values of related compounds.[15] The ability to calculate relative log P values means that one can do retrospective QSAR studies on literature series and also predict potency of new analogues.

For the pargyline series, the availability of π values allowed us to include more analogues and to notice that our earlier partition measurements on some of the compounds were apparently in error. By 1969 we were able to derive a statistically significant equation but not one that explained enough of the variance.

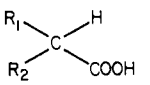
Publication of Many Examples of QSAR. The third important contribution of Hansch and Fujita is their enthusiasm, as evidenced by the number of data sets they have examined. The sheer number of such experiences reinforces the notion that frequently there is some mathematical relationship between biological potency and physicochemical or substructural features.

This large body of successful correlations encouraged us to continue to examine the pargyline QSAR. Ultimately it became clear that there is an optimum pK_a, a significant lipophilic effect, and that ortho substituents increase potency.[16] Thus, the direct contribution of Hansch and Fujita to helping us solve our problem were (1) the use of regression analysis, (2) the consideration of multiple physicochemical properties simultaneously, (3) the presentation of methods to calculate log P, and (4) their enthusiasm for the generality of QSAR.

Differences between QSAR in 1980 and 1970

Factoring π and MR by Position. Current use of the linear free energy approach typically considers the possibility that the biological potency depends not necessarily on overall log P or MR but possibly on π or MR at a specific site on the molecule. The use of such site-specific descriptors fits very well with the analysis of well-characterized receptor binding assays. Together they form a powerful tool for mapping a receptor, which will be discussed in more detail later.

A second use of factoring π and MR by position is the simple treatment of stereoisomers. This involves distinguishing substituents of the same stereochemical relationship to the pharmacophore from those with the opposite relationship. For example, the plant growth inhibition of a series of phenoxypropionic acids (II) in which



II

either R_1 or R_2 is an aryloxy function was analyzed.[17] The QSAR showed that the chiral carbon atom and its attached H and COOH determine the stereochemistry of binding to the receptor. In the R series the phenoxy group binds in a hydrophobic region, whereas in the S series it is the nonphenoxy substituent that binds in this region.

Use of Other Multivariate Methods. In the past 10 years several different mathematical methods have been applied to QSAR problems. Cluster, principal component, factor, discriminant, and nonlinear regression analysis and pattern-recognition techniques are now routinely used.[18-28] It is one measure of the productivity of the QSAR philosophy that it has generated the application of these techniques. These methods expand the field into exploring (1) an active/inactive type of biological response, (2) the relationships between or independence of physical or bi-

(14) Taylor, J. D.; Wykes, A. A.; Gladish, Y. C.; Martin, W. B. *Nature (London)* **1960**, *187*, 941.
(15) Fujita, T.; Iwasa, J.; Hansch, C. *J. Am. Chem. Soc.* **1964**, *86*, 5175.
(16) Martin, Y. C.; Martin, W. B.; Taylor, J. D. *J. Med. Chem.* **1975**, *18*, 883.
(17) Lien, E. J.; Rodrigues de Miranda, J. F.; Ariens, E. J. *Mol. Pharmacol.* **1976**, *12*, 598.
(18) Hansch, C.; Unger, S. H.; Forsythe, A. B. *J. Med. Chem.* **1973**, *16*, 1217.
(19) Dunn, W. J., III; Wold, S. *J. Med. Chem.* **1978**, *21*, 1001.
(20) Franke, R.; Dove, S.; Kuhne, R. *Eur. J. Med. Chem.* **1979**, *14*, 363.
(21) Cammarata, A.; Menon, G. K. *J. Med. Chem.* **1976**, *19*, 739.
(22) Martin, Y. C.; Holland, J. B.; Jarboe, C. H.; Plotnikoff, N.; *J. Med. Chem.* **1974**, *17*, 409.
(23) Hyde, R.; Lord, E. *Eur. J. Med. Chem.*. **1979**, *14*, 199.
(24) Kubinyi, H. *Arzneim.-Forsch.* **1976**, *26*, 1991.
(25) Martin, Y. C. "Physical Chemical Properties of Drugs"; Yalkowsky, S.; Sinkula, A.; Valvani, V., Eds.; Marcel Dekker: New York, 1980; p 49.
(26) Kirschner, G.; Kowalski, B. R. *Drug Des.* **1979**, *8*, 73.
(27) Stuper, A. J.; Brugger, W. E.; Jurs, P. C. "Computer Assisted Studies of Chemical Structure and Biological Function"; Wiley: New York, 1979.
(28) Moriguchi, I.; Komatsu, K.; Matsushi, Y. *J. Med. Chem.* **1980**, *23*, 20.

ological properties, (3) the relatedness of molecules as determined by their physicochemical or substructural properties, and (4) nonlinear nonparabolic relationships between potency and a molecular property. The unique strengths and weaknesses of these methods will not be explicitly discussed, rather they will be included only implicitly by their commonality with the extrathermodynamic method.

Replacement of the Parabola with Other Nonlinear Equations.[5] Compartment modeling studies suggest equations that have two important differences from a parabola: the slopes of log $(1/C)$ vs. log P at low log P and high log P are linear and are not identical. For example, from a probabalistic kinetic viewpoint eq 4 was derived.[24]

$$\log (1/C) = a + b \log P - c \log (P^d + 1) + \ldots \quad (4)$$

Equilibrium compartment models led to eq 5 for series of

$$\log (1/C) = X - \log [1 + cP^d + 1/(aP^b)] + \ldots \quad (5)$$

constant pK_a in which there is an aqueous, nonaqueous, and receptor compartment.[2]

Modeling exercises also reevaluated the concept of a general log P_0 or optimum log P for a particular biological response. It is now recognized that the optimum log P may be strongly dependent on the time after administration at which response is measured in in vivo assays[29] or on the amount of extraneous lipid (i.e., the method of preparation of the receptor) for in vitro assays.[2] The pK_a of the compounds and the pH of the test system will also influence the observed optimum log P[2]. Only when all of these variables are kept constant will the log P_0 values be expected to compare between series. For example, Hansch and Lien showed that for many acidic and neutral compounds the log P_0 for in vitro antibacterial activity is higher for Gram-positive organisms than it is for Gram-negative organisms.[30] However, for analogues of the basic antibiotic erythromycin the optimum log P is independent of the species of microorganism.[31] Since the mechanism of action and the ionization properties are different in the two series, it is not surprising that the competing factors that determine log P_0 are different in the two cases.

The other changes in the practice of QSAR are more subtle; they include more reliable methods to calculate log P, other descriptors for the shape and size of a substituent, and incorporation of more stringent statistical criteria for accepting an equation.[4,5,32]

Accomplishments of QSAR

Forcasting of Biological Activity. Many bona fide correct predictions have been published (Table I).[33-64] Most of the entries include several compounds predicted from an equation. It also seems likely that the practitioners of QSAR in industry have a number of other predictions in their files. The number of predictions and the variety of chemical types and biological systems give support to the contention that QSAR equations can predict the potency of untested analogues.

Table I. Correct Predictions of Biological Activity from QSAR

type of compd	biological act.	ref
propynylamines	inhibn of monoamine oxidase	16
erythromycins	antibacterial	31
fusaric acid analogues	inhibn of dopamine β-oxidase	31
dopamine amides	hydrolysis by arylamidases	31
benzylpyridinium ions	inhibn of complement	33
neutral compounds	binding to BSA	33
carboxylate esters	clot lysis	34
amides	hydrolysis by papain	34
phenanthrenes	antimalarials	34
thyroxin analogues	thyroxin	36
pyrimidones	inhibn of Hill reaction	37
1,3-benzodioxoles	synergism of carbaryl toxicity	38
acrylates and methacrylates	acute toxicity	39
pyridine-2-methanols	spasmolytic	40
methoxychlors	insecticidal	41
arylamidinoureas	antimalarial	42
rifamicins	antibacterial	43
cyclopropylamines	inhibn of monoamine oxidase	44
3-carbamoylpiperidines	cholinesterase inhibn	45
pyranamines	antiallergy	46
miscellaneous	rat LD_{50}	47
azapurin-6-ones	antiallergy	48
naphthoquinones	antimalarial	49
β-carbolines	inhibn of monoamine oxidase	50
clonidine analogues	antihypertensive	51
copper chelates	cytotoxicity	52
chlorinated hydrocarbons	bioconcn in fish	53
phenyl oxazolidines	radioprotectives	54
miscellaneous	mouse brain tumor	55
peptidyl-*p*-nitroanilides	hydrolysis by *Subtilisin carlsberg*	56
quinoxaline 1,4-dioxides	antibacterial	57
tuberins	antimycoplasma	58
thiazole β-blockers	inhibn of adenylate cyclase	58
nitrosoureas	antileukemia	59
mitomycins	antileukemia	60
ketones	inhibn of chymotrypsin	61
sulfonamides	antibacterial	62
sulfonamides	pharmacokinetics	63
hydantion derivatives	anti-CNS tumor	64

Proper Design of Series. Since many data sets fit a

(29) (a) Dearden, J. C.; Townend, M. S. *Abh. Akad. Wiss. DDR, Abt. Math., Naturwiss., Tech.* **1978**, *2N*, 387. (b) Dearden, J. C.; Townend, M. S. *J. Pharm. Pharmacol.* **1976**, *28*, 13P.
(30) Lien, E. J.; Hansch, C.; Anderson, S. M. *J. Med. Chem.* **1968**, *11*, 430.
(31) Martin, Y. C. in ref 4, p 271.
(32) Topliss, J. G.; Edwards, R. P. "Computer Assisted Drug Design"; Olson, E. C.; Christoffersen, R. E., Eds.; American Chemical Society: Washington, D.C., 1979, p 131.
(33) Hansch, C.; Yoshimoto, M.; Doll, M. H. *J. Med. Chem.* **1976**, *19*, 1089.
(34) Hansch, C. "Biological Activity and Chemical Structure"; J. A. Keverling Buisman, J. A., Ed.; Elsevier: Amsterdam, 1977; p 47.
(35) Purcell, W. P. "Biological Activity and Chemical Structure"; Keverling Buisman, J. A., Ed.; Elsevier: Amsterdam, 1977; p 293.
(36) Jorgensen, E. C.; Reid, J. A. *J. Med. Chem.* **1965**, *8*, 533.
(37) Gibbons, L. K.; Koldenboven, E. F.; Nethery, A. A.; Montgomery, R. E.; Purcell, W. P. *J. Agr. Food Chem.* **1976**, *24*, 203.
(38) Desmarchelier, J.; Krieger, R. I.; Lee, P. W.; Fukuto, T. R.; *J. Econ.. Entomol.* **1973**, *66*, 631.
(39) Bass, G. E.; Lawrence, W. H.; Purcell, W. P.; Autrian, U. *J. Dent. Res.* **1974**, *53*, 756.
(40) Cousse, H.; Mouzin, G.; d'Hinterland, L. D. *Eur. J. Med. Chem.* **1973**, *4*, 466.
(41) Goodford, P. J.; Hudson, A. T.; Sheppey, G.; Wootton, R.; Black, M.; Sutherland, G. J.; Wickham, J. C. *J. Med. Chem.* **1976**, *19*, 1239.
(42) Goodford, P. J.; Norrington, F. E.; Richards, W.; Walls, L. P. *Br. J. Pharmacol.* **1973**, *48*, 650.
(43) Quinn, F. R.; Driscoll, J. S.; Hansch, C. *J. Med. Chem.* **1975**, *18*, 332.
(44) Fuller, R. W.; Marsh, M. M.; Mills, J. *J. Med. Chem.* **1968**, *11*, 397.

QSAR equation, it has been recognized that one should plan a series to optimize the chance that the analogues will reveal any QSAR.[18,58,65-67] This planning gives one a good chance of finding that combination of properties which optimizes potency. The analogues should vary substantially in each of the properties proposed to be important in the determination of potency. The variation in each property should be independent of the variation in the other properties. For efficiency, each compound should be unique in these properties. Of course, one must also pay attention to synthetic difficulty, for without that consideration one could lose efficiency.

The various strategies and criteria of series design can be applied to test conformational, quantum chemical, or substructural hypotheses, as well as QSAR ones. The planned series will be only as relevant to the optimization of structure as are the properties used in its design.

Experience in our laboratories has shown that adding such design strategy to the intuition of experienced medicinal chemists decreases the average number of analogues required to investigate 1 physical property from 11 to 4, or a 2.75-fold increase in the information gained per compound synthesized. Even a severe critic of QSAR finds that "the Hansch approach has been of immense value in forcing the chemist to think quantitatively about the chemical properties of the compounds he makes".[68]

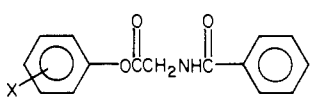
The criteria for good series design are also useful in the decision to terminate synthesis in a series that has not met the biological criteria for success. One is more willing to move on to other work if it is known that the existing analogues represent a thorough search of property space.

It is the opinion of many in the QSAR field that if the only contribution of QSAR to medicinal chemistry were to emphasize the importance of objective series design, this in itself would be a major contribution to increased efficiency.

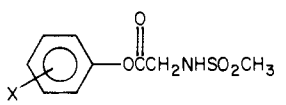
Bioisosterism. One of the concepts of medicinal chemistry is that certain pairs of substituents may be bioequivalent or bioisosteric.[1a,69,70] For many years bioisosterism was qualitative and intuitive, but the QSAR parameterization of substituents now allows one to quantitate this similarity. This concept may be useful in the initial stages of investigating the structural specificity of a lead and also before synthesis in a series is terminated. Of course, if a satisfactory QSAR is obtained the "isosteres" will be determined by that equation.

QSAR and Receptor Mapping. Drug molecules are frequently considered to consist of two parts: an essential pharmacophore that interacts directly with the receptor and accessory regions that can be structurally modified without destroying the essential drug-receptor interaction. (Such modifications may, however, change the strength of this interaction.) Hence, if substitutions are in accessory regions of the molecule or if at least all analogues bind in a comparable orientation, it may be possible to derive a QSAR of the drug–receptor interaction. This QSAR would describe in quantitative terms the forces involved in the interaction. Such studies have been done on peptidyl ester substrates of chymotrypsin[62] and the inhibition of mammalian and bacterial dihydrofolate reductase,[71-73] for example.

A specific example of QSAR receptor mapping is the study of the K_m for substrates of papain.[34] A preliminary regression analysis of two series (III and IV) and the X-ray

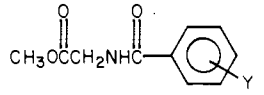


III



IV

structure of the protein suggested that X binds to a group of polar amino acids near the mouth of a hydrophobic cleft and that the amide moiety binds within this cleft. To test the possibility, analogues of type V were synthesized. The

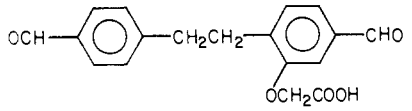


V

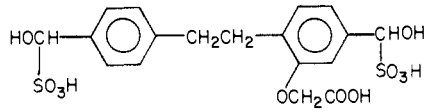
hypothesis was confirmed. This study is important because it illustrates that QSAR need not be done in isolation from X-ray studies, but rather is synergistic with them.

A second example is compounds VI and VII. They were originally designed solely on the basis of their close fit with the atoms in the diphosphoglycerate site on adult human hemoglobin.[74] However, when VI and VII were tested on hemoglobins that differ in the number and types of amino

(45) Beasley, J. G.; Purcell, W. P. *Biochim. Biophys. Acta* **1969**, 175.
(46) Cramer, R. D.; Snader, K. M.; Willis, C. R.; Chakrin, L. W.; Thomas, J.; Sutton, B. M. *J. Med. Chem.* **1979**, *22*, 714.
(47) Enslein, K. *Genesee Stat. Newsl.* **1978**, *6*, 3.
(48) Broughton, B. J.; Chaplen, P.; Knowles, P.; Lunt, E.; Marshall, S. M.; Pain, D. L.; Woolridge, K. R. H. *J. Med. Chem.* **1975**, *18*, 1117.
(49) Martin, Y. C.; Bustard, T. M.; Lynn, K. R. *J. Med. Chem.* **1973**, *16*, 1089.
(50) Martin, Y. C.; Biel, J. H. "Neuropsychopharmacology of Monoamines and Their Regulatory Enymes"; Usdin, E., Ed.; Raven Press: New York, 1974; p 37.
(51) Rouot, B.; Leclerc, G.; Wermuth, C. G.; Miesch, F *J. Pharmacol.* **1977**, *8*, 95.
(52) Coats, E. A.; Milstein, S. R.; Holbein, G.; McDonald, J.; Reed, R.; Petering, H. G. *J. Med. Chem.* **1976**, *19*, 131.
(53) Neely, W. B.; Branson, D. R.; Blau, G. E.; *Environ. Sci. Technol.* **1974**, *8*, 1113.
(54) Fernandez, J.; Grassy, G.; Terol, A.; Robbe, Y.; Chapar, P.; Granger, R.; Sentenac-Roumanou, H. *Eur. J. Med. Chem.* **1978**, *13*, 245.
(55) Chu, K. C.; Feldman, R. J.; Shapiro, M. B.; Hazard, G. F.; Geran, R. I. *J. Med. Chem.* **1975**, *18*, 539.
(56) Pozsgay, M.; Gaspar, R.; Bajusz, S.; Elodi, P. *Eur. J. Biochem.* **1979**, *95*, 115.
(57) Dirlma, J. P.; Czuba, L. J.; Dominy, B. W.; James, Richard B.; Pezzullo, R. M.; Presslitz, J. E.; Windisch, R. M. *J. Med. Chem.* **1979**, *22*, 1118.
(58) Unger, S. H. *Drug Des.* **1980**, *9*, 47.
(59) Montero, J. L.; Imbach, J. L.; *C. R. Hebd. Seances. Acad. Sci., Ser. C* **1974**, *279*, 809.
(60) Remers, W. A.; Schepman, C. S. *J. Med. Chem.* **1974**, *17*, 729.
(61) Grieco, C.; Silipo, C.; Vittoria, A.; Hansch, C. *J. Med. Chem.* **1977**, *20*, 586.
(62) Seydel, J. K. *J. Med. Chem.* **1971**, *14*, 724.
(63) Seydel, J. K.; Wempe, E. *Arzneim.-Forsch.* **1971**, *21*, 187.
(64) Peng, G. W.; Marquez, V. E.; Driscoll, J. S. *J. Med. Chem.* **1975**, *18*, 846.
(65) Craig, P. N. *J. Med. Chem.* **1971**, *14*, 680.
(66) Wootton, R.; Cranfield, R.; Sheppey, G.; Goodford, P. J. *J. Med. Chem.* **1975**, *18*, 607.
(67) Martin, Y. C.; Panas, H. N. *J. Med. Chem.* **1979**, *22*, 784.
(68) Barlow, R. B. *Trends Pharmacol. Sci.* **1979**, *1*, 109.
(69) Hansch, C.; Clayton, J. M. *Intra-Sci. Chem. Rep.* **1974**, *8*, 17.
(70) Thornber, C. W. *Q. Rev. Chem. Soc.* **1979**, *8*, 563.
(71) Stepan, A.; *Badilescu*, I. I.; Simon, Z.; *An. Univ. Timisoara, Ser. Stiinte Fiz.-Chim.* **1977**, *15*, 61.
(72) Dietrich, S. W.; Smith, R. N.; Brendler, S.; Hansch, C. *Arch. Biochem. Biophys.* **1979**, *194*, 612.
(73) Hansch, C.; Dietrich, S.; Smith, R. *Chem. Biol. Pteridines, Proc. Int. Symp., 6th, 1978*, **1979**, 425.
(74) Beddell, C. R.; Goodford, P. J.; Stammers, D. K.; Wootton, R. *Br. J. Pharmacol.* **1979**, *65*, 535.



VI



VII

acids that are proposed to interact with them, it was possible to derive a quantitative correlation between the strength of binding and the number and types of bonds possible. In this case both the receptor and the drug were varied in the QSAR.

This combination of QSAR calculations and X-ray analysis is potentially very valuable. The detailed structure of the proposed binding site will reveal which regions are hydrophobic, which are ionized, and which are open. A successful QSAR, on the other hand, may suggest at which positions of the small molecule increased hydrophobicity increases binding, how changes in the strength of potential hydrogen bonds affects binding, etc. With the use of sophisticated color computer graphics[75] one should be able to integrate the two types of information. For certain series, integration of conformational considerations and/or potential energy calculations on the docking of the drug to the receptor may be necessary. In short, such a combined approach should correct the QSAR deficiency of lack of specific attention to the three-dimensional shape of molecules. A view of the drug–receptor interaction may also help one decide which electronic parameters are important.

QSAR and Pharmacokinetics. It has become a rule of thumb that the various processes of gastrointestinal absorption, distribution, metabolism, and excretion are less sensitive to the specific molecular structure than is true interaction with a receptor. Over the years a number of generalizations have been made: among these are that absorption, penetration into the brain, and metabolism by the liver increase with increasing log P, whereas urinary excretion decreases with increasing log P.[76] What has QSAR methodology contributed to the testing and refinement of these hypotheses? If the only contribution of QSAR were to make these ideas more quantitative and predictive, this would be an important contribution.

A number of studies of the effect of structural variation on the rate of absorption of substances from the rat stomach or intestine or the human buccal cavity have been published. Much of this work showed that the rate of absorption increases with increasing solvent–water partition coefficient. As part of a prodrug project at Abbott, we reanalyzed much of this data with QSAR methodology.[77] For example, we analyzed the data for 193 compounds for which the rate of intestinal absorption was studied by perfusion in the rat.[78-88] The standard deviation from the equation was more than 10 times larger than that from a homologous series studied in buccal absorption.[25] It does not appear that the deviations are due to absorption of ions or to the use of octanol–water log P values rather than hexane–water or $CHCl_3$ log P values. A plot of the compounds that are less than 5% ionized at the pH of the measurement is shown in Figure 1. The important thing to notice is that the lack of fit is not due to the use of the wrong equation, but actually represents a true scatter. For example, the points labeled "8" refer to sulfonamides; the low points are for N-4-acetyl analogues. Although acetylation increases log P by 0.26, the absorption rate for five different analogues decreased by an average of 0.62 h^{-1}. Similar deviations were seen from the analysis of buccal absorption where aromatic acids did not fit the relationship seen with aliphatic acids nor did they form a relationship of their own.[25]

Independently of our work, others have come to the same conclusions.[89] Indeed, several groups have observed that even for a set of related molecules there is not a linear correlation between the rate of partitioning or interphase transfer and the ultimate extent or equilibrium value.[25] The logarithm of the rate of transfer is a bilinear function of log P.

Hence, the conclusion from QSAR analyses is that although there appears to be a general tendency for an increase in absorption rate with increasing log P, (1) this tendency levels off at a log P of approximately 2.0 and (2) absorption rate also depends on the structural change made in the molecule. That is, the contribution of QSAR has been to point out the limitations in the previous qualitative conclusions. A lack of fit to a QSAR equation may be as important to future plans as is a fit.

Binding to serum proteins is generally considered to be a hydrophobic interaction, and a correlation with log P is frequently seen.[90] However, recent careful thermodynamic work has demonstrated that the hydrophobic binding area for anions is limited to binding four methylene units.[91]

Another general property of compounds that has been characterized by QSAR is the relative tendency of compounds to penetrate into the brain. One of the earliest QSAR studies was of this phenomenon; at that time the conclusion was reached that there is a log P_0 of 2.0 for this process.[34] In view of modeling studies[29] and the observations that the rate of metabolism of barbiturates (which is the partial basis for this value) is increased with increasing log P,[92] one should be cautious about applying this

(75) Langridge, R. M.; Jorgensen, E. C.; Ferrin, T. E.; Pensak, M. J.; Anderson, R. *Abstr. Pap. ACS Comp.* **1979**, *177*, 31.
(76) Goldstein, A.; Aronow, L.; Kalman, S. M. "Principles of Drug Action: The Basis of Pharmacology", 2nd. ed.; Wiley: New York, 1974; p 129.
(77) Martin, Y. C.; Hackbarth, J. J., unpublished observations, 1975.
(78) Kitao, K.; Kubo, K.; Morishita, T.; Yata, N.; Kamada, A. *Chem. Pharm. Bull.* **1973**, *21*, 2417.
(79) Morishita, T.; Yamazaki, M.; Yata, N.; Kamada, A. *Chem. Pharm. Bull.* **1973**, *21*, 2309.
(80) Schanker, L. S.; Tocco, D. J.; Brodie, B. B.; Hogben, C. A. M. *J. Pharmacol. Exp. Ther.* **1958**, *123*, 81.
(81) Nogami, H.; Hanano, M.; Yamada, H. *Chem. Pharm. Bull.* **1963**, *11*, 395.
(82) Nogami, H.; Hanano, M.; Yamada, H. *Chem. Pharm. Bull.* **1968**, *16*, 389.
(83) Nogami, H.; Hanano, M.; Yamada, H. *Chem. Pharm. Bull.* **1968**, *16*, 580.
(84) Nogami, H.; Hanano, M.; Yamada, H. *Chem. Pharm. Bull.* **1968**, *16*, 586.
(85) Koizumi, T.; Arita, T.; Kakemi, K. *Chem. Pharm. Bull.* **1964**, *12*, 413.
(86) Koizumi, T.; Arita, T.; Kakemi, K. *Chem. Pharm. Bull.* **1964**, *12*, 421.
(87) Kakemi, K.; Arita, T.; Hori, R.; Konishi, R. *Chem. Pharm. Bull.* **1967**, *15*, 1534.
(88) Kakemi, K.; Arita, T.; Hori, R.; Konishi, R. *Chem. Pharm. Bull.* **1967**, *15*, 1883.
(89) Ho, N. H.; Park, J. Y.; Morozowich, W.; Higuchi, William I. *Proc. Acad. Pharm. Sci. Meet., 1977*, **1978**, p 136.
(90) Bird, A. E.; Nayler, J. *Drug Des.* **1971**, *2*, 277.
(91) Rodrigues de Miranda, J. F.; Eikelboom, T. D.; van Os, G. J.; *Mol. Pharmacol.* **1976**, *12*, 454.
(92) Yih, T. D.; van Rossum, J. M. *Biochem. Pharmacol.* **1977**, *26*, 2117.

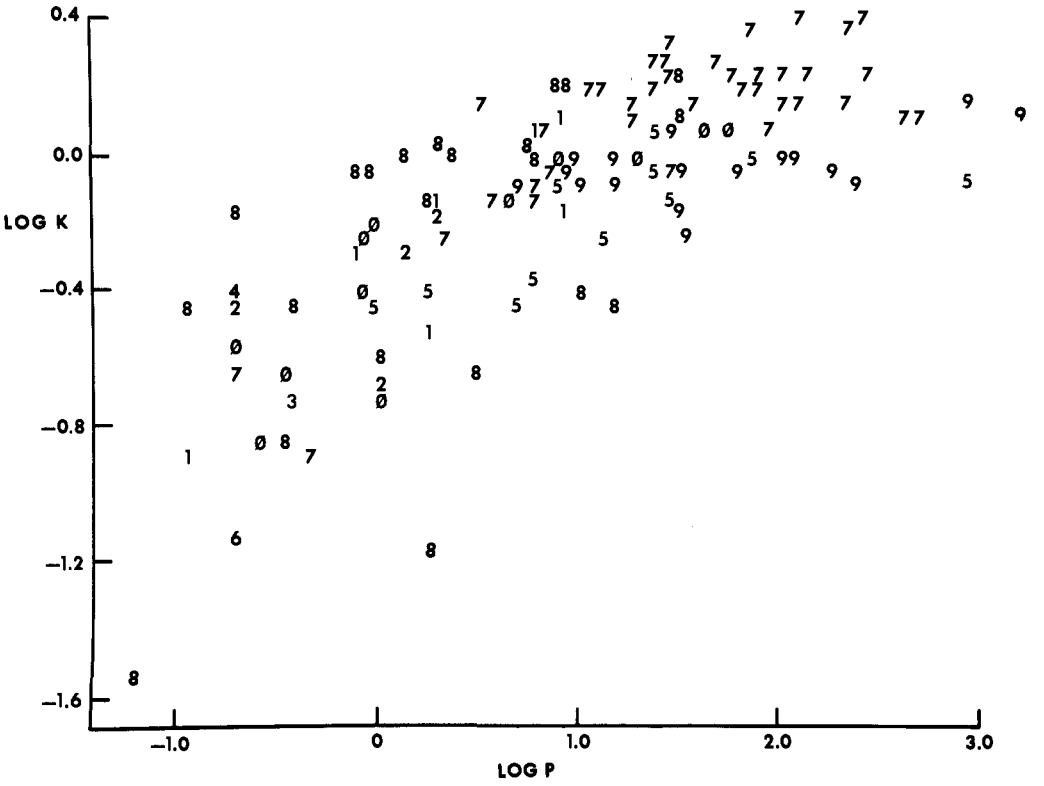


Figure 1. A plot of log K for disappearance from rat intestine vs. octanol–water log P. Only compounds which are >95% nonionized are plotted. 0 from ref 78, 1 from ref 79 measured at pH 4.0, 2 from ref 79 measured at pH 5.8, 3 from ref 79 measured at pH 6.5, 4 from ref 79 measured at pH 7.6, 5 from ref 80, 6 from ref 81, 7 from ref 82–84, 8 from ref 85 and 86, and 9 from ref 87 and 88.

value. Rather, if there is a QSAR study in which the biological measurements correspond rather exactly with those one plans to make and if the chemical series is rather close, then that study, rather than a general one, should be used.

A number of studies have been made of the relationship between rate of metabolism of compounds by various microsomal preparations and physicochemical properties.[92-97] Here the evidence is that although there may be some sort of tendency for increased rate of metabolism with increased log P, this trend is often specific as to the nature of the metabolic transformation investigated.[95] Indeed, one study produced data that did not yield to QSAR analysis.[94] Hence, in the case of metabolism by microsomes the accomplishment of QSAR has been to show that, when one examines the issue carefully, the apparent relationship between rate of metabolism and lipophilicity is not universal, but rather its occurrence depends on the series. It is just as important to know what one can predict as what one cannot predict.

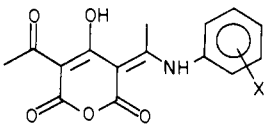
The passive reabsorption of substances from the urinary filtrate to decrease the total amount of substance excreted in the urine has also been studied by QSAR. Again the excretion as a function of log P has been shown to be not linear but linear to an asymptote; in this case the slope is negative and this asymptote is essentially reached at log P of 2 and lower. The studies here are rather sparse, so no comments can be made about the generality of the observed relationships.

More detailed reviews of the QSAR of pharmacokinetics are in preparation by others.[98,99]

Disappointments and Limitations of QSAR

Lack of Universal Success at Predicting Potency. A disappointment with QSAR comes when an equation does not correctly predict the potency of a new analogue. This can result from at least three preventable circumstances: (1) the prediction was based on a poorly designed series or an invalid or ambiguous regression equation, (2) it was based on an extrapolation outside the range of the physical properties represented by the original substituents (while it is valid to make such predictions, one should consider them to be tests of the hypothesis that the QSAR holds outside the original region), and (3) the conditions of the biological tests were different.

A more serious failure occurs when the physical properties on which a prediction was based are not those of importance to the determination of potency. For example, in the investigation of analogues of VIII the equations



VIII

indicated that increased potency could be realized by making compounds less lipophilic.[46] When this hypothesis was tested by synthesis it was confirmed, but one compound had a potency one order of magnitude higher than expected. When the possibility of hydrogen bonding from

(93) Martin, Y. C.; Hansch, C. *J. Med. Chem.* **1971**, *14*, 777.
(94) Gammans, R. E.; Sehon, R. D.; Anders, M. W.; Hanna, P. E. *Drug Metab. Dispos.* **1977**, *5*, 310.
(95) Kapetanovic, I. M.; Strong, J. M.; Mieyal, J. J. *J. Pharmacol. Exp. Ther.* **1979**, *209*, 20.
(96) Yih, T. D.; van Rossum, J. M. *J. Pharmacol. Exp. Ther.* **1977**, *203*, 184.
(97) Ebel, R. E.; Okeeffe, D. H.; Peterson, J. A. *J. Biol. Chem.* **1978**, *253*, 3888.
(98) Seydel, J. K.; Schaper, K. J.; *Int. Encycl. Pharmacol. Ther. Pharmacokin.-Theor. Method.*, in press.
(99) Austel, V.; Kutter, E., in preparation.

the various positions of the molecule was considered, a satisfactory regression equation was obtained. In spite of the poor predictions, this analysis had a happy ending in that the QSAR analysis ultimately resulted in a compound that was 1000 times more potent than any in the original data set.

Lack of Absolute Ways to Describe Molecules. From both a practical and a theoretical viewpoint, a serious limitation of QSAR lies in the description of the molecules. How does one transform a three-dimensional structural diagram into a set of numbers that describe the potential affinity for a receptor and the ability to trigger some biochemical event, as well as the affinity for the various active and passive transport systems that a drug must transverse? The basis of this difficulty is the lack of fundamental understanding of how to quantitatively describe substituent effects on noncovalent intermolecular interactions. Two problems are present: the first is what sort of interactions might occur, and the second is how to parameterize the substituent effect on the interactions.

(1) **Descriptors of the Three-Dimensional Features.** Both drugs and their receptors are three-dimensional objects with a distribution of chemical features superimposed upon a three-dimensional shape. The interaction between them involves a complementarity or fit between these two objects. Recently, the concept of fit has been expanded to include mutual conformational adaptation of drug and receptor upon interaction. Although these three-dimensional considerations may be implicit in the use of E_s and factoring π and MR by position, eq 1–5 include no terms (because no descriptors have been devised) that specifically describe the variation in conformation, conformational flexibility, or three-dimensional aspects of the analogues. Thus, a successful QSAR will provide only indirect information about the three-dimensional aspects of the drug–biomolecule interaction. Hence, the major sucessful receptor mapping QSAR's have dealt with series in which the common feature is a relatively rigid molecule[72,73] or with series of enzyme substrates where one at least knows some of the atoms that interact with the receptor.[62]

In certain cases no QSAR equation will be found because no conformational descriptors of the molecules were used. On the positive side, the failure of eq 2–4 to fit a set of data on flexible molecules may thus suggest that it will be worthwhile to do a careful analysis of conformational possibilities within that series.[100,101] If there are several rotational degrees of freedom, the QSAR analysis may eliminate the necessity to consider certain ones and, accordingly, simplify the calculations and save time and money. On the other hand, if the QSAR analysis fails on relatively rigid molecules this indicates that the next study should be pharmacophore pattern matching.[102,103]

(2) **Hydrophobicity.** The simplest property to parameterize would appear to be the substituent effect on hydrophobicity. In spite of the work of many investigators, the calculation of log P is still characterized by a large set of empirical rules, interaction terms, and special adjustments.[3,104] For example, consider the series that contains substituted benzenes and α-, β-, and γ-substituted pyridines. The log P values of pyridines do not differ by a constant amount from the analogous benzenes; rather, the change in log P also depends on the position, the Hammett σ value of the substituent, and whether the added substituent contains hydrogen-bonding groups.[105] What descriptor of hydrophobicity does one use in a regression equation? Should one use π values from benzene, indicator variables for hydrogen-bonding groups, and σ values in the regression equation? Since the coefficients of the indicator variables and σ in the log P equation will differ for each position of substitution on the pyridine, they will each require a separate set of indicator variables and σ constants in the QSAR equation. This results in a large number of variables. On the other hand, one may choose to calculate log P and do the QSAR analysis on that variable. One difficulty here is that the log P equation may not be available.

A final complication is that if the interaction critical to potency is a hydrophobic interaction between one position of substitution and a hydrophobic pocket on the receptor, how does one parameterize the relative tendency of the same substituent on the various parent compounds to participate in such an interaction? Do the electronic and hydrogen-bonding terms discussed in the previous paragraph reflect an effect of the ring nitrogn on the hydrophobicity of the substituent or vice versa (see, for example ref 20)? The fact that calculating an octanol–water log P is so uncertain makes one realize the problem of trying to model a protein–drug hydrophobic interaction with its finite size requirements and perhaps nonuniform hydrophobicity within the binding site.

(3) **Electronic Effects.** An even more serious problem arises with the numerical description of electronic effects on substituents. Substituents can change both the degree of ionization of the molecule and the charge distribution. The former may affect the amount of interacting species available to the receptor, while the latter may affect the strength of the drug–receptor interaction. The difficulty is to sort out these possibly competing effects so that the correct interpretation of the data is made. This problem has been addressed but by no means solved.[25] A second difficulty with electronic properties of molecules is that the details of the chemistry of the drug–target molecule interaction are usually not known and, hence, the type of parameter to use to model this interaction is not defined. In principle, quantum chemical calculations should provide some help; however, they remain time consuming and expensive and are of little help if one does not know the atoms involved or the chemical nature of the interaction. This problem of parameterizing electronic effects was apparent in the QSAR analysis of chloramphenicols in which a poor predictability of an equation was found[106] and that of antitumor acridines in which close series had to be handled by separate regression equations.[107]

Problems with the Precise Measurement of the Biological Property of Interest. Since QSAR methodology depends critically on the quantitation of the biological response of interest, any deficiencies in this respect also reflect on the QSAR. Obviously, one is building a straw castle if receptor mapping QSAR is done on compounds that bind in different ways to the receptor or that bind to different receptors. One would hope that the statistics would reveal a problem in such cases, but it is always a danger. A second complication occurs if the

(100) Potenzone, R.; Weintraub, H. R.; Hopfinger, A. J. *Comput. Chem.* **1977**, *13*, 187.
(101) Weintraub, H. J. in ref 33, p 353.
(102) Gund, P. *Annu. Rep. Med. Chem.* **1979**, *14*, 299.
(103) Marshall, G. R.; Barry, C.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. in ref 33, p 205.
(104) Rekker, R. F.; de Kort, H. M. *Eur. J. Med. Chem.* **1979**, *14*, 479.
(105) Fujita, T., private communication.
(106) Hansch, C.; Nakamoto, K.; Gorin, M.; Denisevich, P.; Garrett, E. R.; Heman-Ackah, S. M.; Won, C. H. *J. Med. Chem.* **1973**, *16*, 917.
(107) Denny, W. A.; Atwell, G. J.; Cain, B. F. *J. Med. Chem.* **1979**, *22*, 1453.

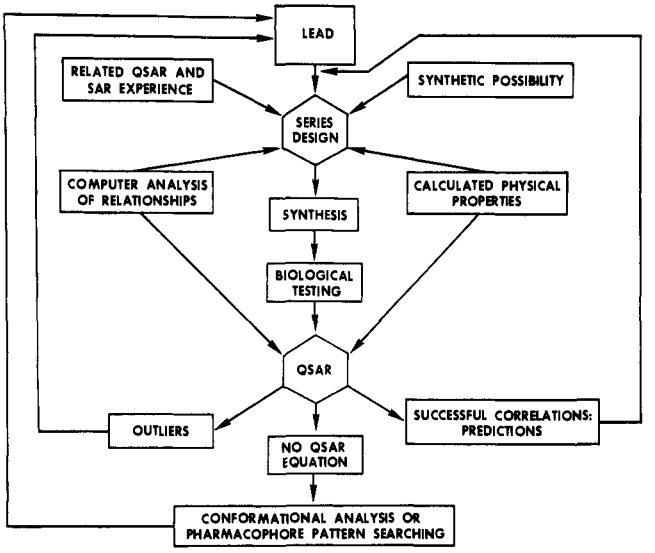


Figure 2. Flow chart of the place of QSAR within a medicinal chemistry program.

biological response measured is a complex result of several processes. For example, many drug–receptor interactions involve drug binding and then conformational changes in the drug and receptor to activate some response. The substituent effect could be on one or both of these processes: if they are not measured independently, then any QSAR would be ambiguous on this point. The problem is that one may not know that there are several processes going on and, hence, one may interpret a QSAR in a completely incorrect fashion.

Role of QSAR in a Medicinal Chemistry Research Program

The previous discussion has alluded to several uses of the QSAR philosophy. Some of these ideas are summarized in Figure 2, a flow chart of decisions in a medicinal chemistry research program. QSAR contributes to the design of the first series to follow up a lead by experiences from similar structural types and/or biological activities and by computer analysis of the relationships between physical properties of proposed compounds. After the first series has been synthesized and tested, QSAR may produce correlations to be tested by further synthesis, outliers that suggest a whole new lead structure, or no significant relationship, in which case more elaborate conformational or receptor–mapping studies may be indicated. In addition, two QSAR circumstances may lead to the decision to terminate synthesis: one case would be that in which a carefully planned series was designed and tested but found to be lacking in the required biological properties so no QSAR was possible; a second would be that in which a series can be fit by a regression equation that suggests that the best compound has already been synthesized. In the latter case some isosteres may also be planned. QSAR is most effective when it is an integral part of the planning and evaluation of the synthetic medicinal chemistry program.[108]

Future Practice of QSAR

As with any tool, it is to be expected that QSAR will evolve in the future. One trend is already apparent, that is, the incorporation of current QSAR methodology into a larger complex of computer programs and data bases. For example, consider Figure 3. When one starts a QSAR study, in the future it might be possible to enter the structure graphically and have the computer look up or calculate the various physical properties of interest. This could be as simple as looking up a σ constant or as complex as a complete conformational analysis or an ab initio molecular orbital calculation. Some such programs are currently available, and others are being written for the Chemical Information System network. Since the Pomona College Medicinal Chemistry Project also has a master file of all data it has correlated, searching this file for the QSAR of compounds or biological activity similar to the one of current interest would also be possible. Both in prodrug work and in straight analogue modification it would be helpful to have an estimate of the type of metabolic transformations the compounds might undergo: this information is also being developed in programs. In the series design aspect of a program, as well as when one wants to test predictions, one would like to know which compounds are relatively easy and which one's are harder or perhaps impossible to synthesize. For this purpose one would possibly tap into one of the programs and data bases currently under development. The appropriate mathematical programs to do the regression, cluster, principal component, and factor analyses and pattern-recognition calculations would form the final module. The master program would query the user and from the answers select a strategy of providing the most relevant information.

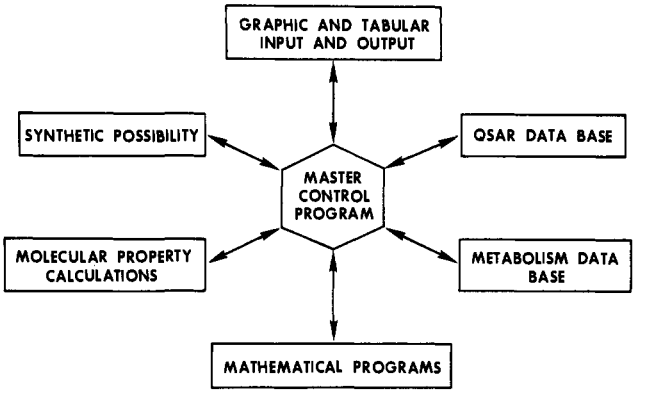


Figure 3. Relationship between the various QSAR modules and other computerized aids for medicinal chemistry.

Another trend that is likely to continue is the integration of QSAR with methods which explicitly consider the three-dimensional aspects of molecules.

A final area of development is the integration of multiple biological activities into one QSAR analysis, that is, to devise means to optimize the profile of a set of analogues rather than the activities one by one. The profile may be as simple as simultaneous optimization of potency and time course or it might involve a multiple optimization of several biological activities.[109]

Note Added in Proof: Since the manuscript was prepared several additional examples of correct predictions have been found: (1) embryotoxicity of alkoxybenzylalkylamines for *A. lixula*,[110] (2) embryotoxicity of alkoxybenzylalkylamines for other sea urchins,[110] and (3) sedative activity of spirobarbituric acids.[111]

(108) Kutter, E. "Arzneimittelentwicklung Grundlagen Strategien Perspektiven"; Kutter, E., Ed.; Georg Thieme: Stuttgart, 1978.

(109) Mager, P. P. *Drug Des.* **1980**, *9*, 187–236.

(110) Buznikov, G. A.; Kabankin, A. S.; Kolbanov, V. M.; Landau, M. A.; Aroyan, A. A.; Ovespyan, T. R.; Teplits, N. A. *Khim. Farm. Zh.* **1976**, *10*, 23–27.

(111) Voronin, V. G.; Polevaya, O. Y.; Mukhanova, V. G.; Landau, M. A.; Kolbanov, V. M.; Privolneva, T. P.; Chugnov, V. V.; Lavretskaya, E. F. *Khim. Farm. Zh.* **1976**, *10*, 43–46.