Quantitative Structure–Activity Relationships and the Unnamed Science

CORWIN HANSCH

Department of Chemistry, Pomona College, Claremont, California 91711

Received June 11, 1992

Introduction

In myriad ways science, through its various compartments—chemistry, biology, medicine, toxicology, environmental science, and computational science—is engaged in studying the interaction of organic compounds with all forms of life or life's component parts (DNA, enzymes, organelles, cells, membranes, etc.). This is a more massive effort than we often realize. The pharmaceutical companies in the United States spend approximately 11 billion dollars per year on research and development. To this must be added that spent for pesticide research and environmental toxicology and the NIH budget of >9 billion dollars. Thus, the expenditure in the United States must be more than 20 billion dollars, and that of the world is probably approaching 40 billion dollars. Moreover, the rate of expenditure for analysis of the chemical ↔ life interaction (the unnamed science) is increasing rapidly as we realize the many benefits to our health which are accruing from such study.

What this new science should be called is not clear, but its great and strongly growing importance cannot be denied. That is, we are trying to learn how to estimate the intended and unintended toxic effects of drugs, pesticides, and natural and industrial chemicals on the various forms and parts of life. Chemical ↔ life is a two-way interaction. While chemicals affect life processes in many ways, living organisms have their own facilities for attacking xenobiotics.

Central to understanding the chemical ↔ life interaction is the subject of structure–activity relationships. The objective of this Account is to illustrate how we can bring together two areas of science which have seemed far apart: physical organic chemistry and the study of the chemical ↔ life interactions. The cornerstone for our thinking is the Hammett equation: $\log k = \rho\sigma +$ constant. In this expression, σ values of three types are obtained from[1b,c] (1) the substituent effects on the ionization of benzoic acids, (2) the ionization of phenols (σ^-), and (3) the solvolysis of $XC_6H_4CClMe_2$ (σ^+). k is a rate or equilibrium constant while ρ is a measure of the sensitivity of the reaction to substituent changes. Although we do not understand exactly why the Hammett equation works, it is one of the most successful concepts in elucidating reaction mechanisms.

Taft extended the Hammett equation by means of the steric parameter E_s to account for effects of substituents near the reaction center,[1b] and we have shown how the use of hydrophobic constants provided a means for extending Hammett's thinking to biological problems:[1a] $\log k = f$ (steric, electronic, hydrophobic). That is, one seeks to find model parameters from physical organic systems to account for three major property changes in a set of congeneric chemicals.

In the end, the Hammett equation became established only because it gives generally meaningful and consistent answers to a variety of problems, as we have tried to illustrate with a few examples of quantitative structure–activity relationships (QSAR) (eqs 1–12). The problems we face in establishing bio QSAR are vastly more difficult, but we believe the road to success will be similar. Although bio QSAR started as an extension of Hammett's basic thinking,[1a] it is now in a state of confusion because of the rapid development of a multitude of methods.[1d]

We believe that the only way to bring order to the field is by laterally correlating each new QSAR with as many other QSAR as possible.

Over the past 20 years, we have made a start on this problem by collecting about 6000 sets of data with attendant QSAR about evenly divided between physical organic and biomedicinal chemistry and toxicology. All of the parameters and structures for about 90 000 compounds have been interfaced with an effective search program which can carry out tasks such as the following: Find all QSAR on nucleophilic substitution reactions, and list them in order of increasing values of ρ. Find all QSAR on bacteria and on physical organic reactions which contain a term in E_s with a coefficient in the range 0.5–1.5, and list in order of increasing value of the coefficient. Find all QSAR on individual subjects such as microsomes, chloroplasts, oxidoreductases, plants, insects, humans, or electrophilic substitution or additions, nucleophilic substitutions or additions, free radicals reactions, etc.

Lateral Validation in Physical Organic Chemistry

Our thesis is that a QSAR standing alone means very little. Ingenious researchers may have thrown the book of parameters at the problem and have found a statistically significant equation. The questions are, is

Corwin Hansch received his undergraduate education in chemistry at the University of Illinois and his Ph.D. in Organic Chemistry from New York University in 1944. After working with the Du Pont Company, first on the Manhattan project and then in Wilmington, DE, he joined the Pomona College faculty in 1946. The present review is the outgrowth of the study of the chemical–biological interactions of plant growth regulators commenced at that time with Professor R. M. Muir, then at Pomona, now at the University of Iowa.

(1) (a) Hansch, C. *Acc. Chem. Res.* **1969**, *2*, 232. (b) Hansch, C.; Leo, A. *Correlation Analysis in Chemistry and Biology*; Wiley-Interscience: 1979. (c) Hansch, C.; Leo, A.; Taft, R. W. *Chem. Rev.* **1991**, *91*, 165. (d) *Comprehensive Medicinal Chemistry*; Ramsden, C. A., Ed.; Pergamon Press: New York, 1990; Vol. 4.

0001-4842/93/0126-0147$04.00/0

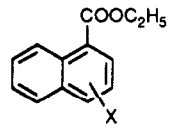
this a chance correlation, and if not, what does it mean? Statistics alone do not answer these questions: only by organizing our knowledge through lateral correlations can we attain a useful view of structure–activity relationships. A search of our data bank turns up the following examples of ester hydrolysis which can be compared with one on nitriles.

alkaline hydrolysis of $XC_6H_4COOOCMe_3$ in 50% ethanol at 20 °C[2a]

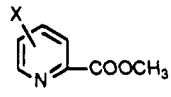
$$\log k = 2.18\sigma + 0.62 \qquad n = 5, r = 0.997, s = 0.041 \quad (1)$$

alkaline hydrolysis of $XC_6H_4COOC_2H_5$ in 87% ethanol at 30 °C[2b]

$$\log k = 2.51\sigma - 1.28 \qquad n = 18, r = 0.993, s = 0.105 \quad (2)$$

alkaline hydrolysis of  in 85% ethanol at 50 °C[2c]

$$\log k = 2.13\sigma - 2.58 \qquad n = 9, r = 0.998, s = 0.048 \quad (3)$$

alkaline hydrolysis of  in 85% methanol at 25 °C[2d]

$$\log k = 2.03\sigma - 2.09 \qquad n = 10, r = 0.997, s = 0.083 \quad (4)$$

alkaline hydrolysis of XC_6H_4CN in 60% ethanol at 82 °C[2e]

$$\log k = 2.13\sigma - 1.0 \qquad n = 5, r = 0.981, s = 0.184 \quad (5)$$

Structures and reaction conditions as well as substituents studied vary considerably, yet ρ is in the range 2.25 ± 0.25. This insight is of help beyond classical physical organic chemistry. It provides a reference point in the study of the enzymatic hydrolysis of esters. For example, in the hydrolysis of esters $XC_6H_4OCOCH_2NHCOC_6H_5$ in buffer, ρ = 1.91, but ρ for enzymatic hydrolysis by thiol hydrolases in the same buffer it is only about 0.6.[3] Polarization of the ester bond by the enzyme is the main factor, and polar effects by substituents contribute very little.

In other instances, enzymic hydrolysis of esters parallels that in nonenzymic systems.

deacylation of XC_6H_4COO-chymotrypsin at pH 7.07, 25 °C[2g]

$$\log k = 1.73\sigma - 2.07 \qquad n = 7, r = 0.971, s = 0.177 \quad (6)$$

deacylation of XC_6H_4COO-chymotrypsin at pH 8.5[2h]

$$\log k = 1.73\sigma - 3.48 \qquad n = 11, r = 0.960, s = 0.275 \quad (7)$$

In these two examples of the hydrolysis (of what can be considered esters of benzoic acids) based on different sets of compounds, we find good agreement between ρ values. The value of ρ is below that found in QSAR 1–5. In each of these examples the 4-NO_2 derivatives were poorly fit and omitted, and the 4-F congener for QSAR 7 was also omitted. These equations can be compared with eq 8 also for reactions in water.

alkaline hydrolysis of $XC_6H_4COOCH_3$ in aqueous solution[2i]

$$\log k = 1.66\sigma + 1.92 \qquad n = 14, r = 0.999, s = 0.022 \quad (8)$$

In making the search which yielded QSAR 1–5 we first isolated all data sets (304) on hydrolysis containing a term in σ (but not σ^- or σ^+). Then narrowing the list to those with ρ in the range 2–2.5 yielded 74 hits. Of these 12 were for the hydrolysis of aryl-COOR esters and one for nitriles. The above are representative examples. Searching for examples with $\rho > 2.5$ turns up, from the present database, only three examples of $XC_6H_4COOC_2H_5$. For the one run in 95% DMSO, ρ = 2.99; in 85% DMSO, ρ = 2.85, and in 65% DMSO, ρ = 2.38.[2j] This raises the question, does the non-H-bond-donating DMSO lack the power to polarize the esters that alcohols have so that the role of the substituent becomes more pronounced? The hydrolysis of 4-$XC_6H_4COOCH_3$ in water (eq 8), which has a ρ of only 1.66, supports this view. In 95% and 85% DMSO, the water would appear to be tied up by interaction with the strong H-bond-accepting DMSO, but at 65% DMSO enough water is present so that ρ falls in the same class as for eqs 1–5. It has long been known that ρ for the acid hydrolysis of esters is usually near 0. In the case of the thiol hydrolases mentioned above k_{cat} is essentially constant so that the rate of hydrolysis is controlled by K_m. Formation of the complex is rather like attaining the transition state. The small value of ρ (0.6) resembles that for acid hydrolysis of esters where the proton so polarizes the carbonyl group that there is little role for electronic effects of substituents.

Searching in the range of ρ = 1–2, we find two other aryl carboxylates. For the alkaline hydrolysis of $XC_6H_4COOC_6H_5$ in acetone–40% H_2O at 25 °C, ρ = 1.92,[2k] and for the hydrolysis of $XC_6H_4COOCH_3$ in H_2O–33% dioxane at 25 °C, ρ = 1.94.

The value of ρ is not the end of a structure–activity analysis. It is illuminating information about the sensitivity of the reaction center to electronic effects of substituents. It does suggest similarities or differences in mechanism which can then be explored by other means.

The additive effect of substituents can be illustrated by the following elimination reactions.

(2) (a) Antonovskii, V. L.; Frolova, Z. S.; Romantsova, O. N. *Zh. Org. Khim.* **1969**, 5, 42EE. (b) Kindler, K. *Chem. Ber.* **1936**, *69*, 2792. (c) Fisher, A.; Mitchell, W. J.; Ogilvie, G. S.; Packer, J.; Packer, J. E.; Vaughan, J. *J. Chem. Soc.* **1958**, 1426. (d) Campbell, A. D.; Chooi, S. Y.; Deady, L. W.; Shanks, R. A. *J. Chem. Soc. B* **1970**, 1065. (e) Cohen, L. A.; Jones, W. M. *J. Am. Chem. Soc.* **1962**, *84*, 1625. (f) Fuchs, R. *J. Org. Chem.* **1963**, *28*, 3209. (g) Caplow, M.; Jencks, W. P. *Biochemistry* **1962**, *1*, 883. (h) Amshey, J. W.; Jindal, S. P.; Bender, M. L. *Arch. Biochem. Biophys.* **1975**, *169*, 1. (i) Smith, J. H.; Menger, F. M. *J. Org. Chem.* **1969**, *34*, 77. (j) Hojo, M.; Utaka, M.; Yoshida, Z. *Tetrahedron Lett.* **1966**, 25. (k) Humffray, A. A.; Ryan, J. J. *J. Chem. Soc. B* **1967**, 468.

(3) Hansch, C.; Klein, T. E. *Acc. Chem. Res.* **1986**, *19*, 392.

(4) (a) Banger, J.; Cockerill, A. F.; Davies, G. L. O. *J. Chem. Soc. B* **1971**, 498. (b) De Puy, C. H.; Storm, D. L.; Frey, J. T.; Naylor, C. G. *J. Org. Chem.* **1970**, *35*, 2746. (c) Willi, A. V. *Helv. Chim. Acta* **1966**, *49*, 1725. (d) Hauser, C. R.; LeMaistre, J. W.; Rainsford, A. E. *J. Am. Chem. Soc.* **1935**, *57*, 1056.

$$XC_6H_4CH_2CH_2OSO_2C_6H_4Y \xrightarrow[\text{tert-butyl alcohol, 40 °C}]{\text{tert-butoxide}} XC_6H_4CH{=}CH_2 + YC_6H_4SO_3^- \quad [4a]$$

$$\log k = 2.34\sigma_x + 1.08\sigma_y - 2.22 \qquad n = 24,\ r = 0.996,\ s = 0.061 \quad (9)$$

$$XC_6H_4CH(Me)CH_2OSO_2C_6H_4\text{-}4'\text{-}Me \xrightarrow[\text{tert-butyl alcohol, 50 °C}]{\text{tert-butoxide}} XC_6H_4C(Me){=}CH_2 + CH_3C_6H_4SO_3^- \quad [4b]$$

$$\log k = 2.19\sigma_x - 3.63 \qquad n = 5,\ r = 0.998,\ s = 0.042 \quad (10)$$

$$(C_6H_5)_2CHCH_2OSO_2C_6H_4Y \xrightarrow[\text{methyl Cellosolve}]{CH_3O^-,\ 50\ °C} (C_6H_5)_2C{=}CH_2 + YC_6H_4SO_3^- \quad [4c]$$

$$\log k = 1.09\sigma_y - 2.44 \qquad n = 4,\ r = 0.999,\ s = 0.029 \quad (11)$$

$$XC_6H_4CH{=}NCl \xrightarrow[\text{93\% ethanol, 0 °C}]{CH_3O^-} XC_6H_4CN \quad [4d]$$

$$\log k = 2.37\sigma_x - 1.86 \qquad n = 9,\ r = 0.990,\ s = 0.138 \quad (12)$$

The ρ value in eq 11 is similar to the corresponding value in eq 9, and eq 10 also corresponds to eq 9, although there are differences in the structures and the reaction conditions. In eq 12 a quite different structure is involved, yet ρ is similar to that in eqs 9 and 10. The questions are, how far can such comparisons be made, and, when they seem to fail, what are the reasons?

One of the difficulties with making comparisons of ρ as above is that its value depends on conditions such as solvent and temperature. There is such lack of uniformity in the data that one must make do with what is available.

Lateral Validations in Biological QSAR

The possibilities for misleading QSAR are enormously greater with biological QSAR, where we are trying to account for more complex structural changes in terms of three *major* factors: electronic, steric, and hydrophobic. End points are not as sharp, and we are dealing with nonhomogeneous systems. Computers allow one to easily study scores of parameters and, in some methodologies, literally thousands. Our dilemmas can be illustrated with the following example.

mutagenicity of $XC_6H_4CH_2N(Me)N{=}O$ in *Salmonella typhimurium* TA1535[5a]

$$\log 1/C = 3.55\sigma - 3.88\sigma^2 + 1.62({}^3\chi_p^v) - 5.11 \qquad n = 13,\ r = 0.873 \quad (13)$$

$$\log 1/C = 0.92\pi + 2.08\sigma - 3.26 \qquad n = 12,\ r = 0.891,\ s = 0.314 \quad (14)$$

The workers who developed eq 13 determined π values experimentally and then rejected them when they failed to correlate with log $1/C$ where C is the molar concentration of compound producing a standard number of mutations in a fixed time interval. The parameter π for hydrophobicity is defined[1b] analogously to σ: $\pi = \log P_{XC_6H_5} - \log P_{C_6H_6}$ where P is the octanol/water partition coefficient.

Table I. Coefficients (h) with π or log P for Compounds Acting in Bacterial Mutagenicity Tests

no. of compound	type of compound	test	h
188	aromatic and heteroaromatic nitro compounds[5a]	TA98	0.65
117	aromatic and heteroaromatic nitro compounds[5a]	TA100	1.10
88	aromatic and heteroaromatic amines[5a]	TA98	1.08
67	aromatic and heteroaromatic amines[5a]	TA100	0.92
15	aromatic nitro compounds[5b]	*E. coli*	1.07
21	$XC_6H_4N{=}NN(R)CH_3$[5a]	TA92	0.95
12	$XC_6H_4CH_2N(Me)N{=}O$[5a]	TA1535	0.92
21	quinolines[5c]	TA100	1.14

The parameter ${}^3\chi_p^v$ is a complex parameter from graph theory the meaning of which is not clear to us, but which must be to some extent collinear with π. It was clear from an inspection of the data that errors had been made in the determination of π. Using standard π values from the benzene system[1b] we formulated eq 14. Which equation points the way for future work? Correlation coefficients are close. Equation 13 contains three variables while eq 14 contains two, but eq 14 contains one less data point. We believe that only by extensive lateral correlation with other QSAR can decisions of this type be made.

The study of the role of the hydrophobic interaction in mutagenesis illustrates the point. The slope (h) of the log P term from eight different QSAR for a variety of compounds acting in a variety of test systems is given in Table I.

In each of these equations other terms such as σ^+, ϵ_{LUMO}, ϵ_{HOMO}, and indicator variables appear. However, after these effects are accounted for, a hydrophobic term is found, and except for one example where a special collinearity problem occurs, the slopes center around 1. These observations lead one to accept eq 14, at least until χ can be shown to tie together many lateral relationships in a meaningful way. It should be noted that not all QSAR for mutagenesis contain a hydrophobic term. For example, a set of cisplatinum amine analogues is well correlated by σ^- alone,[6] suggesting that these compounds reach their site of action without a rate-limiting passage through lipophilic barriers. Since hydrophobicity is so important in mutagenicity, it is not surprising to find a critical role for it in carcinogenicity.[7]

In our database of about 3000 biological QSAR only 15% lack a term for hydrophobicity. Hence, when such an equation is encountered, it is revealing that something special is occurring and should be accounted for. So much QSAR work has been done in the past 30 years that one has certain expectations about values of h. It normally falls in the range 0.4–1.1.

A different type of lateral correlation can be made by comparing different biological systems interacting

(5) (a) Debnath, A. K.; de Compadre, R. L. L.; Shusterman, A. J.; Hansch, C. *Environ. Mol. Mutagen.* **1992**, *19*, 53. (b) Debnath, A. K.; Hansch, C. *Environ. Mol. Mutagen.* **1992**, *20*, 140. (c) Debnath, A. K.; de Compadre, R. L. L.; Hansch, C. *Mutat. Res.* **1992**, *280*, 55.

(6) Hansch, C.; Venger, B. H.; Panthananickal, A. *J. Med. Chem.* **1980**, *23*, 459.

(7) Zhang, L.; Sannes, K.; Shusterman, A. J.; Hansch, C. *Chem.-Biol. Interact.* **1992**, *81*, 149.

with the same chemicals. The action of six aliphatic alcohols, thymol, menthol, ether, and several insecticides such as BHC and DDT yielded the following QSAR.[8a]

50% inhibition of beef brain ATPase ($Na^+ + K^+$)

$$\log 1/C = 0.77 \log P + 0.53 \qquad n = 14, r = 0.988, s = 0.237 \quad (15)$$

50% inhibition of yeast growth

$$\log 1/C = 0.92 \log P + 0.53 \qquad n = 12, r = 0.998, s = 0.101 \quad (16)$$

MIC cockroach nerve conduction

$$\log 1/C = 0.91 \log P + 0.19 \qquad n = 12, r = 0.993, s = 0.162 \quad (17)$$

The slope of eq 15 is slightly different, and its standard deviation is greater, suggesting slightly different interactions at the molecular level. The other two QSAR are very similar, suggesting that hydrophobic compounds perturb a membrane to produce the end result.[8b] Molecular shape, such as the difference between ether and DDT, seems unimportant.

There are many other instances where h with log P or π terms are near 1 (0.9 ± 0.1).[8b-d]

Turning now to examples in which electronic and steric factors are involved, we see that it is possible to relate biological QSAR to counterparts from physical organic chemistry.

I_{50} of *Escherichia coli* by $XC_6H_4N{=}C{=}S$[9a]

$$\log 1/C = 2.27\sigma + 4.31 \qquad n = 9, r = 0.963, s = 0.161 \quad (18)$$

$$XC_6H_4N{=}C{=}S + C_2H_5OH \rightarrow XC_6H_4NHC(=S)OC_2H_5$$[9b]

$$\log k = 2.16\sigma - 4.80 \qquad n = 8, r = 0.975, s = 0.181 \quad (19)$$

$$XC_6H_4N{=}C{=}S + C_6H_5NH_2 \rightarrow XC_6H_4NHC(=S)NHC_6H_5$$[9b]

$$\log k = 2.14\sigma - 3.13 \qquad n = 4, r = 0.994, s = 0.060 \quad (20)$$

The close agreement among ρ values suggests that the isothiocyanates inhibit growth by reactions with nucleophilic moieties within the cell. The lack of a hydrophobic term indicates no significant lipophilic barrier between the isothiocyanates and the critical nucleophiles.

Isothiocyanates are a most interesting class of compounds under study for the prevention of cancer. They seem to do this by stimulating phase II enzymes which assist in the elimination of toxic xenobiotics or their metabolic products from the body.[9c,d]

The phenylisothiocyanates are much too reactive to be of use, but aliphatic derivatives show promise. The different SAR of this type compound is brought out by eqs 21 and 22.

inhibition of *E. coli* growth by $RN{=}C{=}S$[9e]

$$\log 1/C = 0.42\pi + 3.52 \qquad n = 11, r = 0.937, s = 0.227 \quad (21)$$

I_{50} of *Aspergillus niger* by $XC_6H_4CH_2N{=}C{=}S$[9f]

$$\log 1/C = 0.55 \log P + 3.28 \qquad n = 13, r = 0.901, s = 0.147 \quad (22)$$

These much less reactive isothiocyanates require only a hydrophobic parameter, even in the case of eq 22, where one might expect a small electronic effect of X.

The lower reactivity of the aliphatic isothiocyanates and the dependence of the QSAR on the hydrophobic parameters suggest that a different kind of reaction is involved, with their toxicities correlated by eqs 21 and 22. While reaction with a nucleophile seems likely, whether or not eqs 21 and 22 model the reaction involved in inducing the phase two enzymes is not clear. It is well-known that the hydrophobic properties of chemicals are highly important in inducing cytochrome P-450 enzymes.[9g]

Evidence shows that sulfonamides inhibit the enzyme carbonic anhydrase by their ionized forms interacting with a positively charged Zn of the enzyme. The following two QSAR support this.

sulfonamides $XC_6H_4SO_2NH_2$ inhibiting carbonic anhydrase[10a]

$$\log 1/K_i = 0.95\sigma + 0.54\pi - 0.35B_{5,3} + 6.29 \qquad n = 31, r = 0.968, s = 0.168 \quad (23)$$

change in ionization of sulfonamides with electronic effect of substituents[10b]

$$\Delta \log K = 0.86\sigma + 0.08 \qquad n = 16, r = 0.962, s = 0.146 \quad (24)$$

In the case of eq 23 it is necessary to take into account hydrophobic and steric factors ($B_{5,3}$ is the sterimol constant for meta substituents)[1d] before one can obtain a good correlation, after which ρ is in reasonable agreement with eq 24. Two other QSAR for this reaction have been derived;[10c] in one, $\rho = 1.55$, and in the other, $\rho = 0.80$. In the example of $\rho = 1.55$, the reaction was enzyme binding, not inhibition; however, we do not believe that this is the reason for the higher

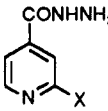
(8) (a) Uchida, M.; Kurihara, N.; Fujita, T.; Nakajima, M. *Pestic. Biochem. Physiol.* **1974**, *4*, 260. (b) Hansch, C.; Kim, D.; Leo, A. J.; Novellino, E.; Silipo, C.; Vittoria, A. *Crit. Rev. Toxicol.* **1989**, *19*, 185. (c) Kakkis, E.; Palmire, V. C.; Strong, C. D.; Bertsch, W.; Hansch, C.; Schirmer, U. *J. Agric. Food Chem.* **1984**, *32*, 133. (d) Pratesi, P.; Caliendo, G.; Silipo, C.; Vittoria, A. *Quant. Struct.-Act. Relat.* **1992**, *11*, 1.

(9) (a) Valchová, D.; Drobnica, L. *Collect. Czech. Chem. Commun.* **1966**, *31*, 997. (b) Rao, C. N. R.; Venkataraghavan, R. *Tetrahedron* **1963**, *18*, 531. (c) Talalay, P.; De Long, M. J.; Prochaska, H. J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 8261. (d) Zhang, Y.; Talalay, P.; Cho, C.-G.; Posner, G. H. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2399. (e) McKay, A. F.; Garmaise, D. L.; Gaudry, R.; Baker, H. A.; Paris, G. Y.; Kay, R. W.; Just, G. E.; Shwartz, R. *J. Am. Chem. Soc.* **1959**, *81*, 4328. (f) Drobnica, L.; Zemanova, M.; Nemec, P.; Antos, K.; Kristian, P.; Stullerova, A.; Knoppova, V.; Nemec, P. *Appl. Microbiol.* **1967**, *15*, 701. (g) Hansch, C.; Zhang, L. *Drug. Metab. Rev.*, in press.

(10) (a) Carotti, A.; Raguseo, C.; Campagna, F.; Langridge, R.; Klein, T. *Quant. Struct.-Act. Relat.* **1989**, *8*, 1. (b) Kakeya, N.; Yata, N.; Kamada, A.; Aoki, M. *Chem. Pharm. Bull.* **1969**, *17*, 2558. (c) Hansch, C.; McClarin, J.; Klein, T.; Langridge, R. *Mol. Pharmacol.* **1985**, *27*, 493.

ρ. We believe that limited variation in substituents, and hence σ, yielded a less reliable ρ.

The antibacterial action of isonicotinic hydrazides provides an interesting comparative study.[11a]

MIC of *Mycobacterium tuberculosis* by 

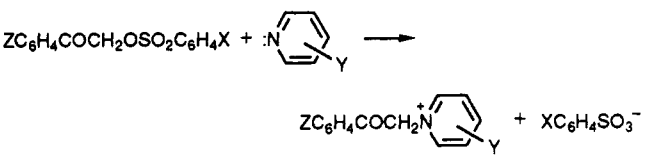
$$\log 1/C = -3.70F + 0.89E_s + 5.78$$
$$n = 17, r = 0.914, s = 0.368 \quad (25)$$

relative rates of alkylation of 2-X-pyridines by CH_3I in CH_3OH[11a]

$$\log k = -3.02F + 0.61E_s + 0.11$$
$$n = 17, r = 0.933, s = 0.275 \quad (26)$$

The agreement between F (field inductive parameter[12]) and E_s[1b] terms of the two QSAR is reasonable considering that homogeneous and heterogeneous reaction systems are being considered. The authors[11a] of this work presume that the isonicotinic hydrazides block the interaction of nicotinamide in the formation of NAD by competition with a nucleophilic substitution.

QSAR 27 represents another pyridine alkylation reaction, which also illustrates the additivity of substituent effects.[11b] Equation 27 is not strictly compa-



$$\log k = -3.24F_y + 1.30\sigma_x + 0.79\sigma_z + 2.79$$
$$n = 35, r = 0.984, s = 0.170 \quad (27)$$

rable to eq 26 in that the substituents on the pyridine ring are only in the 3- and 4-positions. The ρ of F would probably change somewhat for 2-substituents. Still, this illustrates the possibilities for comparison which are easily available from the present data bank. It also shows that three quite different points of substitution in two different molecules can be dealt with by the simple linear combination of substituent constants.

inhibition of *A. niger* spores by $BrCH_2CONHR$[13a]

$$\log 1/C = 0.82E_s + 0.63 \log P - 2.05 \log (\beta(10^{\log P}) + 1) + 3.65 \quad (28)$$
$$n = 15, r = 0.971, s = 0.235, \text{optimum } \log P = 3.40$$

A search of the present bank reveals nothing strictly comparable to the bromo amides of the above QSAR; however, the following isosteric reaction attracted our attention.

alkaline hydrolysis of CH_3COOR[13b]

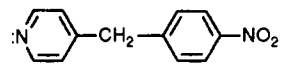
$$\log k = 0.83E_s + 1.93\sigma^* - 0.35$$
$$n = 11, r = 0.955, s = 0.206 \quad (29)$$

In QSAR 28 hydrophobic effects must be accounted for with the bilinear model[14] before the role of steric effects is seen. In this model, activity first increases linearly with a slope of 0.58 until log P of 3.9, when it begins to decrease with a slope of −1.33 (0.58 − 1.91). Only alkyl groups were present in the R of the amides, so that no electronic terms is expected or found in QSAR 28. Such variation was present in the data for eq 29, and a field inductive σ^* parameter[12] is necessary. After these factors are accounted for, the agreement between the E_s terms is good. This indicates that the toxic action may result from the reaction of a nucleophile with the amide carbonyl unit. Of course, we cannot rule out nucleophilic displacement of Br, but this seems unlikely because of its distance from R.

The difficulties of doing comparative QSAR with whole animals are greater, but not insurmountable. In a classic application of the principles of physical organic chemistry to cancer chemotherapy, Ross studied the hydrolysis of aniline mustards [$XC_6H_4N(CH_2CH_2Cl)_2$] assuming that their nucleophilic reactivity with water might parallel the activity of these drugs against cancer. We have formulated QSAR 30 from his results for hydrolysis rates.[15]

$$\log k = -1.84\sigma - 4.02 \qquad n = 11, r = 0.961, s = 0.116 \quad (30)$$

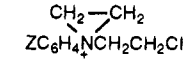
At first glance it is surprising that σ^- does not appear in eq 30 since through-resonance is involved between X and $N(CH_2CH_2Cl)_2$. The reason is that the σ and σ^- constants for the substituents considered were almost the same. Bardos et al. decided that a better model than water would be interaction with



and from their data QSAR 31 was formulated.[15]

$$\log k = -1.92\sigma^- + 1.12I - 1.77$$
$$n = 14, r = 0.972, s = 0.251 \quad (31)$$

In this expression I is an indicator variable which takes the value of 1 for the presence of an ortho substituent on the mustard and 0 for no substituent at this point. Thus, reactivity is increased by about a factor of 10, presumably by twisting the $N(CH_2CH_2Cl)_2$ moiety out of conjugation with the benzene ring, which increases electron density on N, making it a more active nucleophile. The reaction with DNA to limit tumor growth is believed to proceed via an intermediate:



Even though the reactants are rather different, the ρ values are essentially the same, which reminds us of the good correspondence between QSAR 19 and 20. Equation 32 correlates the concentration of drug which

(11) (a) Seydel, J. K.; Shaper, K.-J.; Wempe, F.; Cordes, H. P. *J. Med. Chem.* **1976**, *19*, 483. (b) Yoh, S.-D.; Lee, O.-S. *Tetrahedron Lett.* **1988**, *29*, 4431.
(12) Hansch, C.; Leo, A.; Taft, R. W. *Chem. Rev.* **1991**, *91*, 165.
(13) (a) Leonard, J. M.; Blackford, V. L. *J. Bacteriol.* **1947**, 57, 339. (b) Balakrishman, M.; Rao, G. V.; Venkatasubramanian, *J. Chem. Soc., Perkin Trans. 2* **1974**, 6.
(14) Kubinyi, H.; Kehrhahan, O. H. *Arzneim.-Forsch.* **1978**, *28*, 598.
(15) Panthananickal, A.; Hansch, C.; Leo, A.; Quinn, F. R. *J. Med. Chem.* **1978**, *21*, 16.

increases the life span of a mouse by 25% when infected with L1210 leukemia.[15]

$$\log 1/C = -0.96\sigma^- + 0.86I - 0.31\pi + 4.07$$
$$n = 19, r = 0.926, s = 0.313 \quad (32)$$

The role of σ^- is considerably reduced in eq 32, but the I term is comparable to that in eq 31. It may be that in the mouse the advantages of increasing nucleophilicity for reaction with tumor DNA are offset to some degree by spurious side reactions encountered by the more active drugs in moving from site of injection to sites of reaction.

Microsomal demethylation provides comparative QSAR from the isolated enzyme, to the enzyme in the organelle, to the reaction in man.

demethylation of $XC_6H_4N(CH_3)_2$ by an isolated cytochrome P-450 enzyme[16a]

$$\log 1/K_m = 0.46 \log P + 0.63\sigma^- + 2.62$$
$$n = 8, r = 0.928, s = 0.137 \quad (33)$$

$$\log k_{cat} = -0.68\sigma^- + 1.06 \qquad n = 8, r = 0.986, s = 0.065 \quad (34)$$

$$\log k_{cat}/K_m = 0.53 \log P + 3.47$$
$$n = 8, r = 0.937, s = 0.093 \quad (35)$$

demethylation of various drugs and $XC_6H_4N(CH_3)_2$ by microsomes[16b]

$$\log 1/K_m = 0.69 \log P + 2.91$$
$$n = 14, r = 0.920, s = 0.330 \quad (36)$$

demethylation of $C_6H_4CH(CH_3)NR_1R_2$ by a single human[16c]

$$\log k = 0.60 \text{ CLOGP} - 3.09$$
$$n = 13, r = 0.903, s = 0.262 \quad (37)$$

In the example covered by QSAR 33–35, we obtain the most detailed view of the demethylation process. Here it is seen that the electronic effect of X in forming the enzyme–substrate complex is canceled in the catalytic step, so that in the overall reaction (eq 35) only a hydrophobic effect is seen. The negative ρ of eq 34 is what is expected in an oxidation reaction. Biochemists have been reluctant to do correlations such as eqs 33 and 34 since fundamentally K_m and k_{cat} are not truly independent variables. However, little is lost in making the attempt, and often they are independent enough so that more insight can be gained about the stepwise process.[20a]

In the example of demethylation by the organelle, it was found that k_{cat} was essentially constant so that the overall reaction rate also appears to depend only on hydrophobicity. In this instance, the substrates must penetrate into the lipophilic microsome and then bind to the enzyme. The small electronic effect seen in eq 34 is masked. The standard deviation is rather large for QSAR 36, and this is due to the wide range of structures covered by the QSAR.

In QSAR 37, CLOGP indicates that calculated log P values were used.[16d] In this investigation, 13 amphetamines were fed to an individual and the degree of demethylation was assessed by analysis of the urine. Two other individuals were also so tested. From one, a similar QSAR can be derived ($h = 0.53$, $r = 0.880$, $n = 11$); from the other, a poor correlation was found, illustrating the variation in biochemistry from individual to individual.

In whole animals, or even cell culture studies, it must be remembered that a log P or π term accounts for a number of hydrophobic interactions so that it is a kind of overall average. That is, hydrophobic effects can be important in receptor binding,[3] in the random walk to the receptor[1a] and metabolic modification of the compounds.

The value of a standard parameter for the degree of hydrophobicity of chemicals can be illustrated in a general way. An early QSAR analysis found that the optimum value of log P for penetration of a set of phenylboronic acids into the brain of mice was 2.3.[18a] Redoing the results using CLOGP values and the bilinear model, rather than the parabolic model, we now find a value of 2.1. Subsequently, it was found that a variety of CNS depressants also have optimum log P near 2.[18b] Hence, to design a compound for action in the CNS one should aim for a log P near 2, other factors being equal. The flip side of this problem may be of even greater importance, namely, keeping drugs out of the CNS.

The first antihistamines and then later the first β-blockers with log P near 2 caused CNS problems. More hydrophilic drugs were developed later to avoid this problem.[18c] A more recent example concerns the cholesterol reducing agents lovastatin and pravastatin. The former causes sleeping problems while the latter does not.[18d,e] The two drugs are variations on a parent molecule, but lovastatin differs by a methyl group while pravastatin has an OH at the same position. Their respective log P values are 1.7 and –0.23 at pH 7.4 for the sodium salts. Apparently, enough of the more lipophilic compound enters the CNS to cause sleeplessness. Such observations led to the principle of minimal hydrophobicity in drug design.[18c]

The advantage of having a common definition of hydrophobicity (log P from octanol/water) can be illustrated with examples from agriculture studies as well.

(16) (a) MacDonald, T. L.; Gutheim, W. G.; Martin, R. B.; Guengerich, F. P. *Biochemistry* **1989**, *28*, 2071. (b) Martin, Y. C.; Hansch, C. *J. Med. Chem.* **1971**, *14*, 777. (c) Donike, M.; Iffland, R.; Jaenicke, L. *Arzneim.-Forsch.* **1974**, *24*, 556. (d) Leo, A. Reference 1d, p 295.

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

(18) (a) Hansch, C.; Steward, A. R.; Iwasa, J. *Mol. Pharmacol.* **1965**, *1*, 87. (b) Hansch, C.; Steward, A. R.; Anderson, S. M.; Bentley, D. *J. Med. Chem.* **1968**, *11*, 1. (c) Hansch, C.; Björkroth, J. P.; Leo, A. *J. Pharm. Sci.* **1987**, *76*, 663. (d) Vgontzas, A. N.; Kales, A.; Bixler, E. O.; Manfredi, R. L.; Tyson, K. L. *Clin. Pharmacol. Ther.* **1991**, *50*, 730. (e) Serajuddin, A. T. M.; Ranadive, S. A.; Mahoney, E. M. *J. Pharm. Sci.* **1991**, *80*, 830.

(19) (a) Briggs, G. G.; Bromilow, R. H.; Evans, A. A. *Pestic. Sci.* **1982**, *13*, 495. (b) Cross, B.; Hoffman, P. P.; Santora, G. T.; Spatz, D. M.; Templeton, A. R. *J. Agric. Food Chem.* **1983**, *31*, 260. (c) Kakkis, E.; Palmire, V. C.; Strong, C. D.; Bertsch, W.; Hansch, C.; Schirmer, U. *J. Agric. Food Chem.* **1984**, *32*, 133. (d) Hansch, C.; Kerley, R. *Chem. Ind. (London)* **1969**, 294.

(20) (a) Hansch, C.; Klein, T. E. *Methods Enzymol.* **1991**, *202*, 512. (b) Hansch, C.; Blaney, J. M. *Drug Design: Fact or Fantasy?* Jolles, G., Wooldridge, K. R. H., Eds.; Academic Press: New York, 1984; p 185. (c) Blaney, J. M.; Hansch, C.; Silipo, C.; Vittoria, A. *Chem. Rev.* **1984**, *84*, 333. (d) Hansch, C.; Klein, T. *Methods Enzymol.* **1991**, *202*, 512.

translocation of $XC_6H_4NHCONH_2$ and

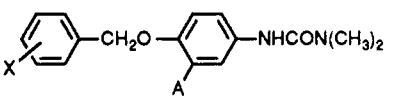
N-methylcarbamoyl oximes from roots to shoots[19a]

$$\log \text{TSCF} = 0.43 \log P - 1.01 \log(\beta(10^{\log P}) + 1) - 0.24I - 0.59 \quad (38)$$

$$n = 17,\ r = 0.917,\ s = 0.164,\ \log P_0 = 1.88$$

The transpiration stream concentration factor (TSCF) is defined as (concentration in transpiration stream)/(concentration in external solution). It was determined that the mass of chemical accumulated in the shoots for a known volume of transpired water is independent of time if the chemical is stable in the plant. The indicator variable I is assigned the value of 1 for the carbamoyl oximes and 0 for the ureas. Its small coefficient would indicate a very slight difference in the two classes of herbicides. Log P_0 is much like that seen for drugs in animals.

This study of xenobiotic movement in whole plants can be compared with the QSAR of a rather different set of compounds:[19b,c]



LD_{85} of mustard plants by the ((phenylmethoxy)phenyl)urea herbicides

$$\log 1/C = 0.48 \log P - 0.86 \log (\beta(10^{\log P}) + 1) - 5.10E_R + 0.87 \quad (39)$$

$$n = 21,\ r = 0.937,\ s = 0.335,\ \log P_0 = 2.10$$

In eq 39 the E_R term applies only to substituents A, and its negative sign indicates that A can better delocalize a free radical electron the less potent the herbicide. E_R values can be found in ref 19d. It appears that metabolic attack occurs on the CH_2 bridge unit and that it is promoted by A but not by X. Log P_0 of eq 39 compares well with that of eq 38 even though the biological end points are different. Since all E_R vaues are positive, this term has a large negative effect on the toxicity of the herbicides. This would seem to be due to metabolic loss via oxidation.

We have recently reviewed[3,20d] the most convincing method of QSAR validation, that of comparing QSAR derived from a purified enzyme with the fit of the ligands to the enzyme model constructed from the X-ray crystallographic coordinates. For example, QSAR 23 indicates that the 3- and 4-substituents in sulfonamides fall on a hydrophobic surface. From the evidence now in hand, we have found h to be near 0.5 when a substituent contacts a surface, but near 1 when it is engulfed in a hydrophobic pocket. In eq 23, h near 0.5 suggests such a surface contact and molecular graphics confirms it.[20d] For 3-substituents, in addition to the hydrophobic interaction, eq 23 calls for a steric effect. This too is confirmed by molecular graphics. Many examples of this type from enzymes whose structures have been established (i.e., dihydrofolate reductase, alcohol dehydrogenase, papain, chymotrypsin, trypsin, elastase) show that the terms in QSAR do have real meaning.[20d]

The above vignettes are only a small sampling of what is already possible with computerized sorting and comparing of QSAR, but we hope it is enough to interest others to work on providing structure to one aspect of the science of chemical ↔ biological interactions. There has been some feeling that physical organic chemistry is in a state of decline. We believe that it is on the verge of a golden age. Elucidation of the incredibly complex web of chemical ↔ life interactions of the unnamed science will not be solved by elegant computer programs and beautiful graphics; it will require decades of careful, imaginative experimental work in biochemistry, molecular biology, and pharmacology, the results of which can then be patiently fit together by the methods of physical organic chemistry.

Construction of the database was initiated with a grant from the NIH in 1970. We also thank Abbott, Burroughs Welcome, du Pont, R. J. Reynolds, and Sumitomo companies for financial support in building the database. In addition, Professors Miki Akamatsu of Kyoto University and Joop Hermens of the University of Utrecht have added numerous sets of data.