faster than τ_{max} in the transition, even though the latter is in a sense also limited by a frictional term, the resistance to fluctuations in twist.

Conclusion

The equilibrium and kinetic properties of the helix–coil transition in nucleic acids clearly reflect a number of detailed properties of the molecule, such as average dimensions and interactions in coil and helix regions and fluctuations in the local base composition. There is reason to expect similar dependences to be found for other nucleic acids such as the naturally occurring transfer and ribosomal RNAs, whose secondary structure is of considerable current interest. Studies of the details of the melting transition of these should give valuable insight into their specific structure. Nucleic acids differ from proteins in having much simpler, more predictable mechanisms of denaturation. It is for this reason that studies of unwinding can be interpreted in terms of molecular parameters for nucleic acids and that work of this kind can be expected to yield specific information more easily than is the case with proteins.

A Quantitative Approach to Biochemical Structure–Activity Relationships

CORWIN HANSCH

Department of Chemistry, Pomona College, Claremont, California 91711

Received January 9, 1969

This Account dates from 1946 when, under less affluent conditions, Robert Muir of the Pomona botany department was housed in our chemistry building. His interest in plant growth regulators and my interest in indole derivatives soon led to a joint effort to correlate chemical structure with biological activity of the indoleacetic acid-like synthetic hormones.[1] Although we at once began to obtain interesting qualitative results, attempts to formulate these in quantitative terms were frustrated by our conceptual training. As a plant physiologist, Muir was well aware of "lock and key" theory of enzyme–substrate reactions and, as a chemist, I was conditioned to explain substituent effects in the electronic terms of the Hammett equation. Neither of us having had training in pharmacological work, the arguments of Veldstra[2] that relative lipophilic character of substituents was highly important fell on deaf ears.

It was during the reflective period of a sabbatical leave in 1960 in Professor Huisgen's laboratory in Munich that it finally became apparent to me that a study of partition coefficients of the phenoxyacetic acid growth regulators might be helpful in assigning relative lipophilic character to the different members of the series. Soon after my return to Claremont, such a study with Maloney[3] showed the importance of this parameter in explaining the relative activity of the many phenoxyacetic acids which had been tested on *Avena* coleoptiles by Muir, now at the University of Iowa.

It was clear from different ways of mathematically combining Hammett σ constants and log P values (P is the octanol–water partition coefficient of the unionized molecule) that a considerable reduction in the variance in the data was possible. At this point Professor Toshio Fujita from Kyoto University joined our group and made the important suggestion that we follow an approach used by Taft[4] and linearly combine the two constants as in eq 1. In eq 1, C is the molar

$$\log (1/C) = k_1\pi + k_2\sigma + k_3 \qquad (1)$$

concentration of compound producing a standard response in a constant time interval. The constants k_1, k_2, and k_3 are obtained *via* the method of least squares. In eq 1, π is defined analogously to σ; *i.e.*, $\pi = \log P_X - \log P_H$. P_X is the octanol–water partition coefficient of a derivative and P_H that of a parent molecule, in this case phenoxyacetic acid.

Although eq 1 gave much better correlations than the simple Hammett equation ($k_1 = 0$), or what might be called a Meyer–Overton[5] equation ($k_2 = 0$), it still left a good deal to be desired. It was not until I could bring myself to postulate that log $(1/C)$ was not linearly but *parabolically* dependent on log P that a generally useful equation (eq 2) was obtained. The idea

$$\log (1/C) = -k_1(\log P)^2 + k_2(\log P) + k_3\sigma + k_4 \qquad (2)$$

behind eq 2 is that molecules which are highly hydrophilic will not penetrate lipophilic barriers readily and hence will have a low probability of reaching the biolog-

(1) (a) R. M. Muir and C. Hansch, *Ann. Rev. Plant Physiol.*, **6**, 157 (1955); (b) R. M. Muir, T. Fujita, and C. Hansch, *Plant Physiol.*, **42**, 1519 (1967).
(2) H. Veldstra, *Ann. Rev. Plant Physiol.*, **4**, 151 (1953).
(3) C. Hansch, P. P. Maloney, T. Fujita, and R. Muir, *Nature*, **194**, 178 (1962).
(4) R. W. Taft in "Steric Effects in Organic Chemistry," M. S. Newman, Ed., John Wiley & Sons, Inc., New York, N. Y., 1956, p 556.
(5) K. H. Meyer and H. Hemmi, *Biochem. Z.*, **277**, 39 (1935).

ical sites of action in the test interval. Molecules having a very high log P value will be strongly held by the first lipophilic material they encounter and thus will also be slowed down in their random walk to the sites of action.

It was of course known in pharmacological work that a linear relation between log $(1/C)$ and log P did not go on forever (consider the old saying among those practicing drug modification: ethyl, propyl, butyl, futile). However, no generally useful hypothesis for a rational mathematical treatment had been developed. From an extensive analysis, Davson and Danielli[6] concluded that a linear relation exists between the logarithm of the partition coefficient and the logarithm of the rate of penetration. Unfortunately, work in this area had been carried out using molecules having relatively low log P values (*i.e.*, ≤ 1). A more recent study[7] has uncovered a number of examples where the linear relationship obviously does not hold. Once the psychological barrier of the linear relationship between the logarithm of the rate of permeation and the logarithm of the partition coefficient was overcome, it was gratifying to find that eq 2, developed for plant growth regulators, gave as good or better correlations with many other biochemical and pharmacological systems.[8]

Although a variety of different approaches[1,2,9] to the quantitative correlation of the relation between molecular modification and the concomitant change in biological response for a set of congeners are under study, we shall be mainly concerned with work from our laboratory in what has been termed extrathermodynamic relations between structure and activity.[10] We have attempted, as a first approximation, to factor the effects of substituents on the biological response of a parent molecule into three broad categories: electronic, hydrophobic, and steric. We have been seeking suitable thermodynamically based constants which can serve as relative measures of the way in which substituents such as Cl, NO_2, CH_3, etc., affect the electronic, hydrophobic, and steric properties of a drug which in turn affects its biochemical properties. The combination of such numerical constants (eq 2) can then be empolyed in computerized regression analysis to ascertain the role of a substituent in modifying pharmacological properties of a parent drug.

The Hydrophobic Property of Substituents

To appreciate the assumptions underlying eq 2, one must go back to the work of Meyer and Overton which excited such wide interest at the turn of the century. Their work indicated that, within limits, one found a linear relation between biological activity and lipophilic character of the members of a set of related drugs. Equation 3 formulates[11] what one these days would

$$\log (1/C) = 0.94 \log P + 0.87 \qquad n = 51 \qquad r = 0.971 \qquad (3)$$

term a linear free energy relationship between the molar concentration (C) causing isonarcosis in tadpoles and the octanol–water partition coefficient. In eq 3, n is the number of compounds tested and r is the correlation coefficient. Equation 3 correlates the narcotic action of a wide variety of alcohols, esters, ketones, aromatic hydrocarbons, alkyl halides, etc., with the single physical property of the partition coefficients. Although Overton[12] published the data in 1901, only recently[11] have enough partition coefficients been measured in a single system to allow the formulation of eq 3. Previous to this work, linear plots were made between log $(1/C)$ and log P for small sets of congeners causing standard biological responses.[5] However, no attempt was made until very recently[13] to measure many partition coefficients in a standard reference system and to use these to assess the role of the hydrophobic character of a drug in causing a response.

The importance of lipophilic character in biologically active organic compounds has intensified interest[14] in what has been roughly termed hydrophobic bonding. When an organic molecule is placed in aqueous solution, a loose envelope of water molecules appears to form around it. When such a molecule is removed from this phase into a nonpolar phase, this rather orderly shell of water molecules disintegrates. The change in entropy for this process appears to be one of the most important contributing factors to "hydrophobic bonding." Dipole–dipole interaction, hydrogen bonding, and London forces are also important. In this Account, we use the octanol–water partition coefficient as an operational representation of capacity for hydrophobic bonding.

Returning to eq 3, there are a number of implications of the linear relationship between relative biological activity defined as log $(1/C)$ and hydrophobic character of a drug as represented by log P. Approximately 95% of the variance in the biological data is "explained" by the simple linear equation ($r^2 = 0.943$). Since the structure of the molecule counts so little, and since the tadpoles are swimming in a solution of the narcotic, it

(6) H. Davson and J. F. Danielli, "The Permeability of Natural Membranes, 2nd ed, Cambridge University Press, Cambridge, England, 1952.

(7) J. T. Penniston, L. Beckett, D. L. Bentley, and C. Hansch, *Mol. Pharmacol.*, **5**, in press.

(8) (a) C. Hansch, *Ann. Rept. Med. Chem.*, **1966**, 347 (1967); (b) C. Hansch, *ibid.*, **1967**, 348 (1968).

(9) (a) S. F. Free and J. Wilson, *J. Med. Chem.*, **7**, 395 (1964); (b) B. Pullman and A. Pullman, "Quantum Biochemistry," Interscience Publishers, New York, N. Y., 1963; (c) H. C. Hemker, *Biochim. Biophys. Acta*, **63**, 46 (1962); (d) D. Agin, L. Hersh, and D. Holtzman, *Proc. Natl. Acad. Sci. U. S.*, **53**, 952 (1965); (e) W. P. Purcell, *Biochim. Biophys. Acta*, **105**, 201 (1965); (f) R. Zahradník, F. Boček, and J. Kopecký, Proceedings of the 3rd International Pharmacology Meeting, 1966, Vol. 7, Pergamon Press, New York, N. Y., 1968, p 127. (g) A. Cammarata, *J. Med. Chem.*, **11**, 1111 (1968); (h) J. C. McGowan, *J. Appl. Chem.*, **4**, 41 (1954); (i) L. J. Mullins, *Chem. Rev.*, **54**, 289 (1954).

(10) J. E. Leffler and E. Grunwald, "Rates and Equilibria of Organic Reactions," John Wiley & Sons, Inc., New York, N. Y., 1963.

(11) C. Hansch in "Drug Design," Vol. I, E. J. Ariëns, Ed., Academic Press, New York, N. Y., in press.

(12) E. Overton, "Studien uber die Narkose," Fischer, Jena, Germany, 1901.

(13) (a) T. Fujita, J. Iwasa, and C. Hansch, *J. Am. Chem. Soc.*, **86**, 5175 (1964); (b) C. Hansch and S. M. Anderson, *J. Org. Chem.*, **32**, 2583 (1967); (c) C. Hansch, J. E. Quinlan, and G. L. Lawrence, *ibid.*, **33**, 347 (1968).

(14) G. Némethy, *Angew. Chem.*, **6**, 195 (1967).

has been assumed[15] that equilibrium is established between drug in exobiophase and drug in or on the sites of action. Thus the mere presence of the compound is assumed to produce narcosis. It may be that in some living systems equilibrium is essentially attained and that the chemical potential of the narcotic at the sites of action is equal to that in the applied solution. More recent second thoughts[7,15b,16] lead one to believe that approximate attainment of equilibrium is the exception rather than the rule. A stepwise formulation of our present understanding is

$$\text{drug in exobiophase} \xrightarrow[\text{step 1}]{k_1} \text{drug adjacent to site of action} \xrightarrow[\text{step 2}]{k_2} \text{drug on critical enzyme or membrane} \xrightarrow[\text{step 3}]{k_3} \text{perturbation of enzyme or membrane} \longrightarrow k_4\text{–}k_n \text{ observed biological response} \quad (4)$$

The simplest assumption would be that if eq 3 is linear with respect to log P, then each of the above processes would also be approximately linear with respect to log P. Evidence is now in hand providing some justification for such assumptions.

That the logarithm of the rate of movement of organic compounds through biological tissue is roughly linearly related to log P is exemplified in eq 5. Equation 5, from the work of Collander,[17] shows the relationship between the passive movement of a wide variety of organic molecules through *Nitella* cells. While the correlation is not as good as eq 3, this may be in considerable part due to the analytical techniques available to Collander at that time. A recent reevaluation of Collander's study[18] indicates that a slightly curvilinear relationship gives a small improvement over eq 5 and that the size of the compound, as measured by molecular weight, probably plays a small part. No doubt if molecules having much higher values of log P had been studied, the departure from linearity would have been more evident. The essentially linear nature of eq 5 and other such examples[11] would support the idea that when log P is not too high, step 1 of eq 4 is linearly dependent on log P.

	n	r	
$\log \text{Perm} = 0.96 \log P_{\text{ether}} - 3.92$	61	0.891	(5)

It is only very recently[19] that efforts have been made to apply extrathermodynamic methods to step 2. However, it is now quite evident from a number of studies that the binding of small molecules to macromolecules closely parallels partition coefficients. This is illustrated in Table I where P is from the octanol–water system. The most interesting aspect of the results in Table I is the narrow range of slopes associated with the log P term. One could generalize and say that the dependence on hydrophobic bonding is 0.6 ± 0.1 log P for systems as different as highly purified bovine serum albumin, homogenized brain, and nylon. Scholtan[20] has shown that the binding by serum protein of a great variety of drugs can be rationalized as in Table I, using the isobutyl alcohol–water reference system. The results in Table I strongly support the linear nature of step 2 of eq 4 (see eq 6–14 in Table I).

Very little quantitative data of a direct nature can be cited to support the linear nature of step 3 of eq 4. A preliminary result is embodied[19b] in eq 15, where % $\Delta\alpha_D$ is the per cent decrease in levorotation (sodium D line) caused by the binding of various organic compounds to bovine serum albumin.

	n	r	
$\% \Delta\alpha_D = 2.23 \log P - 2.70$	9	0.988	(15)

The above results on relatively simple systems help us, at least in part, to understand the few long-standing examples and the many recently uncovered examples[11] of linear relations between biological response [log $(1/C)$] and hydrophobic character (log P) of sets of drugs. A few examples are given in Table II (see eq 16–32 in Table II).

In correlations 16–19, several different examples are given of a single class of compounds, simple aliphatic alcohols, acting on different test systems. The slopes of the correlations indicate a very similar dependence of the inhibitory action of alcohols on their hydrophobic character. The intercepts (b) are surprisingly close, considering differences in the test systems as well as the difference in the definitions of biological response. We have found[11] for a larger group of sets of alcohols (nine) a mean intercept and standard deviation of 0.34 ± 0.2. The intercept is a useful parameter for comparing the intrinsic stereoelectronic activity of isolipophilic molecules. The size of the intercept will be a function of the sensitivity of the test system and the intrinsic activity of the functional part of the molecule. This can be illustrated by comparing the alcohols with correlations 20–23 for the barbiturates. The mean intercepts for alcohols and barbiturates differ by 0.8 log unit or, in other words, the barbiturates are about six times as effective in inhibiting oxidative metabolism.[21]

A similar comparison can be seen in the better defined chymotrypsin system (correlations 31 and 32). Here the phenol function is about five times more inhibitory than the isolipophilic neutral molecules such as acetonitrile, benzene, acridine, naphthalene, etc.

Using the numerical values for partition coefficients and regression equations allows one to group together information about a large number of drugs and to make easy comparisons of such sets. In this sense we are more interested in "the forest than in the individual trees." Comparison of a given individual in one set with an individual in another set might not be very close. Averaging out errors in testing as well as small differences in intrinsic activity which might be due to

(15) (a) J. Ferguson, *Proc. Roy. Soc.* (London), **B127**, 387 (1939); (b) J. Ferguson, *Chem. Ind.* (London), 818 (1964).
(16) C. Hansch and T. Fujita, *J. Am. Chem. Soc.*, **86**, 1616 (1964).
(17) R. Collander, *Physiol. Plantarum*, **7**, 420 (1954).
(18) B. V. Milborrow and D. A. Williams, *ibid.*, **21**, 902 (1968).
(19) (a) K. Kiehs, C. Hansch, and L. Moore, *Biochemistry*, **5**, 2602 (1966); (b) F. Helmer, K. Kiehs, and C. Hansch, *ibid.*, **7**, 2858 (1968); (c) C. Hansch and F. Helmer, *J. Polymer Sci., Part A-1*, **6**, 3295 (1968).
(20) W. Scholtan, *Arzneimittel-Forsch.*, **18**, 505 (1968).
(21) See Table II, footnote *b*.

Table I

Hydrophobic Binding of Small Molecules by Macromolecules[a]

$$\text{Log } K = a \log P + b$$

Type of compound	Macromolecular system	a	n	r	Correlation	Ref
Miscellaneous aromatic and aliphatic	Bovine serum albumin	0.75	42	0.960	6	19b
Miscellaneous aromatic and aliphatic	Bovine hemoglobin	0.71	17	0.950	7	19b
$RCOO^-$ [b]	Bovine serum albumin	0.59	5	0.966	8	19b
ROH	Ribonuclease	0.50	4	0.999	9	19b
Barbiturates	Homogenized liver	0.52	5	0.973	10	11
Barbiturates	Homogenized brain	0.52	5	0.973	11	11
Penicillins	Human serum	0.49	79	0.924	12	*c*
Acetanilides	Nylon	0.69	7	0.961	13	19c
Thyroxine analogs	Human albumin	0.46	8	0.950	14	11

[a] The indicated reference should be consulted for the definition of the equilibrium constants K as well as for values of the intercept. [b] In this correlation, π was used instead of log P. See text for definition. [c] A. E. Bird and A. C. Marshall, *Biochem. Pharmacol.*, **16**, 2275 (1967).

Table II

Linear Relationships between Biological Response and Hydrophobic Character

$$\text{Log } (1/C) = a \log P + b$$

Type of compound	Biological system	a	b	n	r	Correlation	Ref
ROH	Inhibition of bacterial luminescence	1.17	0.22	8	0.998	16	*b*
ROH	Toxicity to red spider	0.69	0.16	14	0.979	17	11
ROH	I_{50}, lung O_2 consumption	0.90	0.16	7	0.995	18	*b*
ROH	I_{50}, O_2 consumption cervical ganglion	1.16	0.57	4	0.994	19	11
Barbiturates	I_{50}, NADH oxidase	1.11	1.24	6	0.921	20	*b*
Barbiturates	I_{50}, brain O_2 consumption	1.04	0.96	10	0.956	21	*b*
Barbiturates	I_{50}, mitochondria O_2 consumption	1.35	0.85	8	0.990	22	11
Barbiturates	I_{50}, egg cell division	0.80	1.08	19	0.960	23	*b*
Phenols	Conversion of cytochrome $P_{450} \rightarrow P_{420}$	0.63	0.27	11	0.985	24	11
Miscellaneous	C-Mitosis of allium root	0.96	0.56	19	0.963	25	*b*
Benzimidazoles	I_{50}, Hill reaction	0.95	0.98	16	0.950	26	*c*
Ureas	Hypnosis of mice	0.55	2.42	23	0.943	27	11
Miscellaneous	Inhibition of liver succinooxidase	0.80	0.12	14	0.963	28	11
Alcohols and ketones	Denaturation of T4-phage DNA	0.68	0.01	15	0.937	29	11
Phenylacetic acids	Inhibition of *Avena* cell elongation	0.73	3.01	18	0.972	30	1b
Phenols	Inhibition of chymotrypsin[a]	0.95	−1.88	10	0.990	31	*d*
Miscellaneous neutral compounds	Inhibition of chymotrypsin[a]	1.00	−2.60	8	0.995	32	*d*

[a] In these examples an inhibition constant, K, was used instead of log $(1/C)$. [b] C. Hansch and S. M. Anderson, *J. Med. Chem.*, **10**, 745 (1967). [c] K. H. Büchel, W. Draber, A. Trebst, and E. Pistorius, *Z. Naturforsch.*, **21**, 243 (1966). [d] I. V. Berezin, A. V. Levashov, and K. Martinek, Vth Federated European Biological Society Meeting, Prague, 1968.

electronic or steric factors allows us to make a better assessment of predominate features of this process. The similar dependences of similar processes from quite different systems can be seen by comparing the slopes of correlation 23 for barbiturates inhibiting division of sea urchin egg cells and correlation 30 for phenylacetic acids inhibiting oat cell elongation. The intercepts are quite different. Further investigation is necessary to determine how much of this difference is to be attributed to the difference in sensitivity of the two systems and how much (if any) is due to the greater intrinsic activity of the phenylacetic acids. One wonders what significance is to be attached to the great similarity between correlation 26 for the inhibition of the photosynthetically related Hill reaction by benzimidazoles and the inhibition of oxidative metabolism by barbiturates (eq 20–23).

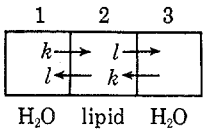
Even from the quite limited surveys of Tables I and II, it is apparent that lipophilic molecules bind to and influence all kinds of biochemical processes which are controlled by the macromolecular systems of enzymes and membranes. Little significance is to be attached to binding of a given drug to serum fractions or, say, the inhibition of an enzyme unless one can make comparisons on an isolipophilic basis. Only such comparisons can demonstrate that something more than nonspecific hydrophobic action is involved. However, establishing a linear free energy relationship such as those above gives one a reference system with which to measure and judge steric and electronic factors which contribute so much to specificity in biochemical processes.

Nonlinear Relation between Biological Response and Hydrophobic Character

The good correlations in the simple systems of Tables I and II and the many others[11] provide us with some hope and confidence in approaching more complex sys-

tems. The movement of drugs through tissue to their sites of action has long been regarded as important in explaining drug action. Several different attitudes have been assumed toward this problem by workers in the field. Some have attempted to ignore the effect of small changes in structure on this process. Others have assumed a linear relationship between movement and partition coefficient, and still others have considered the problem too complicated to yield a rational solution by the application of present knowledge. It was assumed in the formulation of eq 2 that in the most general sense there would be an optimal value for the partition coefficient of a drug for finding its site of action *via* a random walk process through macromolecular structures. The data in Table I lend support to this concept. As far as we have been able to extend our studies, there is a linear relation between log K or log $(1/C)$ and log P in the nonspecific binding of small molecules by macromolecules. Thus, highly lipophilic molecules tend to be localized by the first protein or lipid with which they come in contact.

Scheme I



The concept of an optimum log P can be justified from a simple kinetic model[7] of the type used in multicompartment analysis. In Scheme I there is a lipid barrier between two aqueous compartments. The rate constant for passive movement from aqueous to lipid phase is k, and that for the reverse process is l. Compartment 1 has a given volume V_1 and, at an initial time, a concentration of A_1 solute. Corresponding values apply to the other compartments. More compartments can be added to Scheme I so that the small molecule must cross many membranes. It can also be assumed that an irreversible first-order reaction removes drug from the final phase with rate constant m.

The general set of differential equations to be analyzed is given by eq 33. In these equations, A_i rep-

$$\frac{dA_1}{dt} = -kA_1 + lA_2$$

$$\frac{dA_{2i}}{dt} = -2lA_{2i} + k(A_{2i-1} + A_{2i+1}) \qquad (33)$$

$$\frac{dA_{2i+1}}{dt} = -2kA_{2i+1} + l(A_{2i} + A_{2i+2})$$

$$\frac{dA_n}{dt} = mA_{n-1}$$

resents the concentration in the ith phase. Since A_1/A_n does not depend on A_1^0, an arbitrary initial concentration such as 1, 10, etc., is employed. For a specific value of n, the partition coefficient ($P = k/l$) is allowed to vary over a range of values defined by the condition

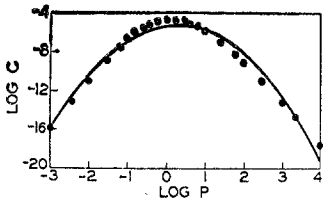


Figure 1.

that $kl = 1$. The solution for the set of equations at time t can be obtained using the Runge–Kutta approximation and a large computer. The points in Figure 1 show the concentrations of drug in the last compartment of a 20-barrier model when $t = 10$ and $m = 1$. The solid line is a parabola fit by the method of least squares to the calculated points. The good fit is justification for the equation used in correlating the structure–activity relations of Table III.

For all of the correlations (34–45) in Table III, the exponential term is quite significant. Over and above the good correlations of Table III is the interesting parameter log P_0 obtained by setting the derivative d log $(1/C)$/d log P equal to zero. The value of this constant represents the ideal lipophilic character (as defined by octanol–water) for a set of congeneric drugs. Correlations 34–36 are typical results obtained with hypnotics. For these drugs log P_0 is $\sim$2. In an extended study of hypnotics[22] a mean value of 2 was found for 16 sets of drugs comprising eight different sets of barbiturates and eight miscellaneous sets of hypnotics. Although the intercepts varied depending on the nature of the functional group (alcohol, amide, etc.), maximum activity in a given series is associated with a surprisingly constant value for log P_0. Since these drugs are relatively nonspecific, it would seem that the value of log P_0 is primarily determined by the passive movement of drug from the point of application to the sites of action. Assuming this to be so, it is of interest to compare log P_0 values from other systems. Most pertinent is the value found from correlation 38. In this work, benzeneboronic acids were injected into mice and the concentration of boron in the brain measured 15 min later. The result is that forecast by eq 33. Moreover, the log P_0 of 2.3 is close to that found in the hypnotics. The log P_0 value of correlation 37 for quite nonspecific action also supports this conclusion. Log P_0 values for correlations 39 and 40 are also near 2. In these examples one might assume that the rate-limiting process in glucuronide and hippuric acid formation in rabbits is the random walk of the drug to the sites of enzymic action.

The above discussion is not meant to imply that log P_0 will usually be near 2. Its value will depend very much on the system. Correlations 41–43 are part of a more extensive study with bacteriostatic compounds. For the Gram-positive organisms in Table III, as well

(22) C. Hansch, A. R. Steward, S. M. Anderson, and D. Bentley, *J. Med. Chem.*, **11**, 1 (1968).

Table III

Nonlinear Relationships between Biological Response and Partition Coefficients

$$\text{Log } (1/C) = k_1(\log P)^2 + k_2 \log P + k_3$$

Biological systems	Type of compound	k_1	k_2	k_3	Log P_0[a]	n	r	Correlation	Ref
Mice hypnosis	Barbiturates	−0.44	1.58	1.93	1.8	13	0.969	34	22
Mice hypnosis	Acetylenic alcohols	−0.51	2.13	0.86	2.1	8	0.944	35	22
Mice hypnosis	Diacylureas	−0.18	0.60	1.89	1.7	13	0.918	36	22
I_{50}, chick embryo hatching	Miscellaneous	−0.16	0.76	2.08	2.4	10	0.982	37	11
Localization in brain	Benzeneboronic acids	−0.53	2.47	−1.05	2.3	14	0.915	38	*b*
Glucuronide formation in rabbits	Secondary ROH[e]	−0.33	1.14	0.74	1.8	11	0.848	39	*c*
Hippuric acid formation	Benzoic acids[e]	−0.67	3.15	−1.76	2.4	8	0.916	40	*c*
Inhibition of *B. diphtheriae*	Hydrocupreines	−0.12	1.43	1.16	5.8	17	0.936	41	*d*
Inhibition of *Strep. hemolyticus*	Phenols[e]	−0.17	2.14	−3.57	6.3	33	0.956	42	*d*
Inhibition of *S. aureus*	4-Hydroxyphenyl sulfides	−0.15	1.73	−2.21	5.9	12	0.995	43	*d*
Inhibition of aphids	RSCN	−0.10	1.14	1.37	5.9	7	0.961	44	11
I_{50}, frog ventricle	Miscellaneous	−0.13	1.49	0.36	5.8	12	0.977	45	11

[a] Log P_0 is found by setting d log $(1/C)$/d log $P = 0$ in the above equations and solving for log P. [b] C. Hansch, A. R. Steward, and J. Iwasa, *Mol. Pharmacol.*, **1**, 87 (1965). [c] C. Hansch, E. J. Lien, and F. Helmer, *Arch. Biochem. Biophys.*, **128**, 319 (1968). [d] E. J. Lien, C. Hansch, and S. M. Anderson, *J. Med. Chem.*, **11**, 430 (1968). [e] In these correlations, relative biological response was used instead of log $(1/C)$.

as for a variety of other such organisms,[23] log P_0 is near 6. For a varied set of drugs and Gram-negative organisms, log P_0 near 4 was observed.[23] The results with the bacteria are from *in vitro* work with the microorganisms immersed in a solution of the drug. The random walk under these conditions is a very short one, and we find essentially a linear relation between log $(1/C)$ and log P until very lipophilic compounds are encountered. Similar conditions prevail in correlations 44 and 45. In these examples the critical reaction(s), step 3 of eq 4, may occur on the outer membranes of an organ or organelle.

It has been pointed out[11] that a variety of *neutral* drugs acting on the central nervous system have log P_0 values near 2. Equation 33 and the results in Table III, to a considerable extent, remove the mystery surrounding the concept of a blood–brain barrier. The assumption has been made that, since it is often difficult to get drugs into the brain, it was in some unique way shielded from foreign molecules injected into the blood stream. This "shielding" is brought about by the binding of very lipophilic molecules to serum protein. These molecules are then gradually stripped from the serum protein by the liver and metabolized to more water-soluble derivatives, and these are excreted in the urine. Very water-soluble molecules (low log P) cannot penetrate the vascular system composed in part of lipophilic molecules to enter the brain. The "blood–brain barrier" is seen in these terms to be no different than the "barriers" surrounding other organs or organelles of the body.

In developing the model used in the equations of Table III, a single concept (eq 33) has been employed to rationalize the empirically found parabolic dependence of log $(1/C)$ on log P. While this is justified, one recognizes that there are other factors which could cause a departure from linearity in this relationship. Brodie's work[24] on the relation between metabolism and lipophilic character inclines one to think that the value of log P_0 could be influenced by metabolism.[11] At a certain level of log P a change of mechanism could occur because increasing the value of log P gives much greater rates of metabolic destruction of the drugs. Since log P does to some extent parallel the size of the molecules, steric effects could be involved in penetration rates. This latter fact would not seem to be highly important when one considers that sets of molecules of such greatly different geometries and having quite different functions have the same log P_0 values for a standard response in similar systems. From the few quantitative correlations which have been made[11] between metabolism or elimination and chemical structure it would appear that these processes are nonspecific and highly log P dependent (correlations 39 and 40). Insofar as these processes are nonspecific, it makes little *practical* difference in deriving equations such as those of Table III whether a drug fails to reach a given reaction site in a standard time interval because it was detained on a lipophilic protein or because it fell into a log P determined metabolic trap.

Correlation of Biological Response with Electronic Parameters

Ever since the formulation of the Hammett equation, workers in the biosciences have been searching for applications of this simple equation or its modifications[10] to biochemical problems. Very few clean-cut examples have been found, and the reason for this is clearly evident from the good correlations of Tables I, II, and III showing the great importance of hydrophobic interactions. A recent survey[11] uncovered less than a half-dozen distinct examples where only electronic parameters would yield a satisfactory correlation. These are exemplified in eq 46–48. In two of these

(23) See Table III, footnote *d*.

(24) B. B. Brodie, J. R. Gillette, and B. N. La Du, *Ann. Rev. Biochem.*, **27**, 427 (1958).

I_{50} of Fly Cholinesterase Diethyl *p*-X-Phenyl Phosphates[25]

	n	r	
$\log (1/C) = 2.37\sigma^- + 4.38$	6	0.985	(46)

Nicotinelike Activity of X-$C_6H_5OCH_2CH_2N^+(CH_3)_3$[8a]

$\log (1/C) = 13.7S_0^{(N)} - 10.5$	6	0.994	(47)

Bacteriostatic Action of Sulfonamides[11]

$\log (1/C) = 1.05\sigma^- - 1.28$	17	0.966	(48)

equations, σ^- gives a better correlation than σ, revealing information about the mechanism of action. In eq 47, $S_0^{(N)}$ is the quantum mechanically calculated super delocalizability of electrons in the *ortho* position of the ring made by Fukui. The example of the sulfonamide study by Seydel is the most extensive set of biologically active molecules giving a sharp linear correlation with the simple Hammett equation. A very slight improvement in this correlation is obtained[11] by the addition of a term in π.

The Dependence of Biological Activity on Steric Effects

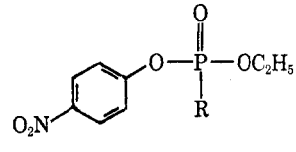
Work of the past decade[26] on the flexible character of enzymes is beginning to counteract the gross oversell the lock and key theory of enzyme–substrate interaction has had on the steric requirements of substrate–enzyme interaction. The outstanding examples of great differences in biological activity of stereoisomers, for example, conditioned many workers to think first in terms of steric factors when explaining differences in a set of congeners. No doubt these are highly important, but they must take their place along with hydrophobic and electronic influences.

There are two broad classes of steric effects to consider: intermolecular and intramolecular. Overlap between these two classifications must also be considered. Intermolecular interactions means spatial interactions between substrate and enzyme or membrane. Intramolecular interactions are interactions within the substrate molecule. It is the latter effect with which we shall be concerned; it is still too early to know what, if anything, can be done in a systematic way with the former type of problem.

A number of examples[27,28] have now been found in which Taft's E_s parameter[4] and Hancock's modification[29] (E_s^c) have proved to be of value in biochemical work. The use of regression analysis in dissecting out the steric effect from electronic and hydrophobic can be illustrated as in eq 49–51; these equations correlate[28c]

	n	r	
$\log K = -0.23\pi + 6.14$	13	0.105	(49)
$\log K = 5.68\sigma^* + 6.31$	13	0.407	(50)
$\log K = 2.58E_s^c + 7.94$	13	0.927	(51)

the inhibition constants for a set of phosphorus esters of structure I acting on cholinesterase *in vitro*. Changes in the function R were limited to normal and branched



I

alkyl groups. The least important substituent effect is that of hydrophobic bonding (eq 49). The meaningful correlation is that of eq 51 showing that variations in K are to be attributed largely to steric effects of R on the ester moiety. This high correlation with E_s^c is surprising because this parameter was formulated from hydrolysis of carboxylic esters under conditions quite different from those involved in eq 51. An equation in which all three terms (π, σ^*, and E_s^c) are combined gives only a very slight improvement.

A Multiparameter Approach to Biochemical Structure–Activity Problems

Equation 52 constitutes a useful model[16] with which one can commence to work on the immensely difficult problems involved in drug design. In eq 52, a and b are constants for a given system. Since the biological

$$\frac{d(BR)}{dt} = a \exp - \left[\frac{(\log P - \log P_0)^2}{b}\right](k_i)(C) \quad (52)$$

response (BR) is normally determined for a fixed time interval, $d(BR)/dt$ can be replaced by a constant. C is the molar concentration of applied drug producing a standard response in the fixed time interval and k_i is a rate or equilibrium constant for a limiting physical or chemical process being affected by the set of drugs. It will be recognized that the exponential term in eq 52 accounts for the probability of a drug molecule reaching the reaction centers in time so that its effect will be contained in the observed biological response. This leads to the relationship of Figure 1 based on eq 33. Keeping in mind that P_0 is a constant for a given system, eq 52 can be converted to eq 53. We assume the sub-

$$\log (1/C) = -k_1(\log P)^2 + k_2(\log P) + k_3 \log k_i + k_4 \quad (53)$$

stituent effects on the rate or equilibrium constant k_i can be treated by the linear combination of thermodynamically based but extrathermodynamically[10] applied constants; that is, we wish to factor substituent effects as in eq 54. Besides the usual problems in such factor-

(25) C. Hansch and E. W. Deutsch, *Biochim. Biophys. Acta*, **126**, 117 (1966).
(26) D. Koshland and K. E. Neet, *Ann. Rev. Biochem.*, **37**, 359 (1968).
(27) See Table III, footnote *c*.
(28) (a) C. Hansch, *J. Med. Chem.*, **11**, 920 (1968); (b) C. Hansch and E. J. Lien, *Biochem. Pharmacol.*, **17**, 709 (1968); (c) C. Hansch, *Farmaco*, **23**, 293 (1968); (d) C. Hansch, E. W. Deutsch, and R. N. Smith, *J. Am. Chem. Soc.*, **87**, 2738 (1965).
(29) C. K. Hancock and C. P. Falls, *J. Am. Chem. Soc.*, **83**, 4214 (1961).

$$\delta_X \log k_i = \delta_X \Delta G_{\text{hydrophobic}} + \delta_X \Delta G_{\text{electronic}} + \delta_X \Delta G_{\text{steric}} \quad (54)$$

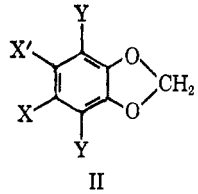
ing,[10] one must recognize that $\Delta G_{\text{hydrophobic}}$ cannot be defined in a pure sense so that electronic and steric components are excluded. The above division is arbitrary and approximate. In using π to extrathermodynamically represent $\delta_X \Delta G_{\text{hydrophobic}}$, we are using octanol–water partitioning to define hydrophobicity. Of course, partitioning will depend on electronic interactions of, say, drug and octanol or water. The size of the molecule will also influence π or log P values. Thus, by electronic and steric effects we mean highly specific effects not contained in the operational definition of hydrophobic interactions. There are a variety of free energy related constants which we can use to obtain numerical solutions for the effect of substituents (X) on k_i. Such an extrathermodynamic postulate is embodied in eq 55. While one must keep in mind that higher order approximations[21] may be necessary for

$$\log k_i = k' \log P + k''\sigma + k'''E_s + k'''' \quad (55)$$

good correlations, eq 55 illustrates how one might assess the relative importance of hydrophobic effects in terms of log P, electronic effects in terms of the various kinds of σ constants, and intramolecular steric effects in terms of E_s. Substitution of eq 55 into eq 53 gives an equation of general use for complex systems. An application of such a complex equation is contained in eq 56; this equation correlates[28a] substituent effects in

	n	r	
$\log \text{SR5} = -0.21\pi^2 + 0.71\pi + 1.46\sigma\cdot + 0.88E_s + 1.59$	16	0.943	(56)

a set of benzodioxoles (II) with the insecticide activity of 1-naphthyl-N-methylcarbamate in houseflies. SR5



II

represents the synergistic effect of a given benzodioxole when five times as much synergist as insecticide was used. The constant, $\sigma\cdot$, is a radical parameter formulated from the homoarylation studies of Hey.[28a] The E_s term in eq 56 pertains to only four derivatives in which bulky groups in the X′ position were adjacent to a NO_2 in the X position. Such groups greatly reduced the activity of the synergists, presumably by decoupling the π electron system of the NO_2 function from that of the aromatic system. A variety of σ constants were tried in eq 56; however, the best result was obtained with the radical parameter. This is taken as *prima facie* evidence for a radical mechanism of action in which the synergists react with enzymes which destroy the insecticide. Some muttering about the large number of terms used in eq 56 to treat 16 derivatives may be in order. Our experience with such equations is still quite limited; however, such an approach does provide insight unattainable from the usual intuitive inspection of the data. Knowing that substituent effects are most probably influencing radicals helps in the design of the next derivatives. The calculation of π_0 also helps design in setting a limit to the amount of hydrophobic character one should attempt to build into the synergist.

Summary

Our discussion has centered around the kind of symbols of organic chemistry which the medicinal and biochemist might use in the discussion of structure–activity studies. The structural formulas of classical organic chemistry are very inadequate for such discussions. For example, computer programs are now being used to store and print out on demand all of the formulas of a particular type of drug. The chemist is then handed an enormous list of, say, all compounds active against malaria. The effect of looking at such a list of several thousand compounds is completely bewildering because our present symbolism in organic chemistry is not well suited for discussing the dynamics of organic reactions in nonhomogeneous systems (or even homogeneous ones). The English school of chemists started in the 1930's to formulate symbols such as ±I, ±D, etc., to describe the dynamic effects of substituents on organic reactions. A huge advance was made by Hammett, who showed how the qualitative symbols should be formulated in numerical terms. The extension of his breakthrough by Taft, Brown, Charton,[30] and others[10] has provided those who work with organic reactions powerful statistical tools with which to evaluate organic reaction mechanisms. Our efforts have been directed toward finding the missing link which will allow us to bring to bear on biochemical reactions (often occurring in media of blood and guts) some of the powerful tools of physical organic chemistry. The evidence in hand is that log P or π can enable us to employ computers in a numerical analysis of biochemical struture–activity problems.

This work has been supported by National Institutes of Health Grants GM 07492 and CA 11110. We are also much indebted to the Smith Kline and French Company for financial help. None of the biological data summarized in the equations of this report was obtained in the Pomona laboratory. I regret that lack of space does not permit direct acknowledgment of the many sources from which these data were drawn. I wish to thank Professor R. Nelson Smith for many helpful discussions and Professor Donald McIntyre without whose instruction in the use of computers this work would not have been possible.

(30) M. Charton, *J. Am. Chem. Soc.*, 91, 624 (1969).