

Relationships between the Biological and Physicochemical Properties of Series of Compounds

Richard M. Hyde

Wellcome Research Laboratories, Beckenham, Kent, England. Received July 8, 1974

An equilibrium model is described for relationships between the biological activity and the physicochemical properties of compounds in a congeneric series. An equation derived from the model is of the form, $\log C = \log (a + 10^{-\alpha}) + b$, where C is the total molar concentration of drug to produce a standard response, α is a parameter of the form $\log K$ where K is an equilibrium constant, and a and b are coefficients. This equation provides a plausible fit to some nonlinear observations which have been studied.

When quantitative biological data are available for a series of compounds, relationships between the physicochemical properties and observed biological activities of the molecules have frequently been sought. This approach is known as the method of physicochemical-activity relationships (the PAR method¹). It is common practice to express the biological activity as the logarithm of the molar concentration of the compound which is needed to produce a standard effect. The physicochemical properties are selected both for their relevance to the biological system and for more practical criteria. The partition parameter² π and Hammett's constant³ σ are the most commonly used physicochemical parameters. Bulk parameters have also been widely applied, as have the components of σ which were derived by Swain and Lupton⁴ and called \mathfrak{F} and \mathfrak{R} . All these parameters can be predicted for benzenoid compounds, but measured physicochemical properties have also been used by some workers.

Current practice¹ entails the use of regression analysis to find an equation which relates biological activity to a linear or parabolic function of one or more physicochemical parameters. A typical PAR analysis would involve the testing for statistical significance of relationships such as

$$\log ED_{50} = a\sigma + b\pi + c\pi^2 + d \quad (1)$$

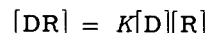
where a , b , c , and d are coefficients determined by the method of least squares.

It has been suggested that each linear term in such an equation may represent a process-limiting step to which the corresponding physicochemical parameter is relevant, while a specific justification for the use of square terms has been based on a diffusion argument.⁵ However, the general justification for seeking a relationship between any biological activity and these physicochemical parameters is by no means obvious.

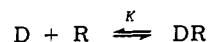
A concise review of earlier theoretical work in this field was published by Higuchi and Davis⁶ who extended the approach by considering the influence of structural changes in a drug molecule upon its distributive tendencies to compartments in the body other than the biological

receptor site itself. Their approach led to another function based on the equilibrium treatment of the bio-system. The present approach is also based on an equilibrium model and leads to a simple equation which can be used to interpret both linear relationships and some nonlinear results which were previously interpreted by the use of square terms. Two forms of the model, both of which lead to the same conclusion, are described.

Model 1. According to occupancy theory⁷ the interaction of drug, D, and receptor, R, can be expressed as



where DR represents the complex between drug and receptor, and K is the equilibrium constant governing its formation. The biological response is assumed to be a function of the concentration of DR, which is determined by the mass action equation



If the total molar concentration of drug is $C = [D] + [DR]$ and the total concentration of receptors is $R' = [R] + [DR]$ which is assumed constant for the system, it follows that

$$[DR] = K(C - [DR])(R' - [DR])$$

$$[DR] = KC(R' - [DR]) - K[DR](R' - [DR])$$

$$K(R' - [DR])C = [DR](1 + K(R' - [DR]))$$

$$\log C = \log [DR] + \log (1 + K(R' - [DR])) - \log K - \log (R' - [DR])$$

The above hypothetical equilibrium constant, K , may be written as $K = k \times 10^\alpha$ where k is a constant for the series and α is a parameter which varies from compound to compound. The parameter α must be in the form $\log K'$ where K' is an equilibrium constant relating to a model physicochemical system. This constant K' may be measured or predicted. For example, α may be set equal to $\log P$ where P is a measured partition coefficient. If predicted physicochemical parameters are to be used, care

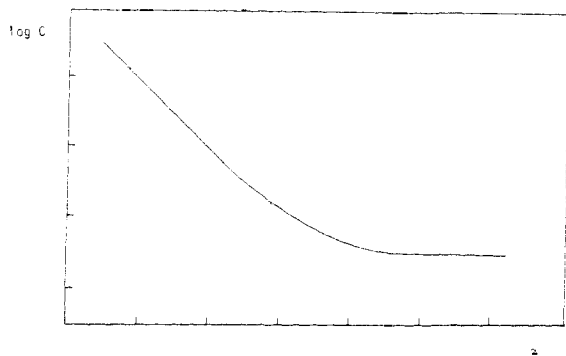


Figure 1. Plot of $\log C$ against α based on eq 4. Both axes are represented on the same scale. It can be seen that the gradient is initially equal to -1 and is ultimately equal to zero as α becomes more positive. The scales are in logarithmic units, one division corresponding to one decade.

must be taken to ensure that they are of the correct form. Thus, the Hammett equation would dictate the use of the product $\sigma\rho$. The parameter π is defined by an equation which does not involve any reaction constant but there is no evidence, theoretical or practical, to suggest that a serious error would be introduced by setting $\alpha = \pi$, *i.e.*, by assuming a "reaction constant" of unity, provided that π pertains to a series which is closely relevant to the drug molecules being studied.

Substitution of $K = k \times 10^\alpha$ into the above equation now leads to

$$\log C = \log [DR] - \log (R' - [DR]) + \log (1 + k(R' - [DR]) \times 10^\alpha) - \log k - \alpha \quad (2)$$

The biological activity is defined as the logarithm of the total molar concentration C of drug which will produce a given effect, and it is assumed that, for any drug in the series, the standard effect will result from the same concentration of drug-receptor complex DR. The activity of each drug is therefore the value of $\log C$ which produces this concentration of the complex. Since R' and k remain constant for a series of drugs acting on a given receptor, the relationship between the biological activity and a single physicochemical parameter reduces to the simple equation

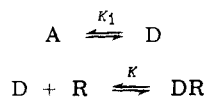
$$\log C = \text{constant} + \log (1 + a \times 10^\alpha) - \alpha \quad (3)$$

where a is a constant. This equation simplifies still further to

$$\log C = \text{constant} + \log (a + 10^\alpha) \quad (4)$$

Figure 1 shows the general shape of the relationship between $\log C$ and α . When a real physicochemical parameter such as $\log P$ is used, one may set α equal to $+\log P$ or $-\log P$. Hence, the present treatment is equally applicable to the case in which the curve slopes up or down.

Model 2. Model 2 is an extension of model 1, in which it is assumed that the drug can exist in two unbound forms, A and D, but only D can interact with the receptor.



K_1 is the equilibrium constant relating concentrations of A and D and might be a partition coefficient, for example. In general it is concerned with the process by which drug reaches receptor and has often been cited as responsible

for any observed relationship between biological and chemical properties.

By the same arguments that were used in model 1, the following relationship may be obtained

$$\log C = \text{constant} + \log (10^{-(\alpha+\beta)} + a \times 10^{-\alpha} + b) \quad (5)$$

where $\alpha = \log K - \log k$ and $\beta = \log K_1 - \log k_1$ and a and b are both constants. Equation 5 can be regarded as the general form of the relationship between a biological activity and the two physicochemical parameters on this model. If K (the constant for formation of DR) is independent of substituents in the series of compounds, eq 5 reduces to

$$\log C = \text{constant} + \log (c + 10^{-\beta}) \quad (6)$$

where c is a constant for the series. This is analogous to eq 4 as derived for the first model. Hence, the same algebraic expression is obtained whether the crucial equilibrium constant is directly involved in the interaction of the drug with receptor or relates to an earlier stage such as access to the receptor.

A review has been made of published results in order to see whether observations consistent with the predictions of this equilibrium model are commonly found. The general form of the relationship which is now predicted between biological activity and a single physicochemical parameter, α , has been illustrated in Figure 1. However, with a limited number of compounds in any series, one would often find that observations were only available for a part of this curve. Whether a straight line of unit gradient or a nonlinear relationship was found would depend on the values of the coefficients in eq 4 and the experimental range of the parameter, α .

While no attempt is made to present a comprehensive survey of the practical use of this model, some examples showing the sort of fit that has been obtained using eq 4 are quoted in Table I. The table is based on sets of data for which linear or parabolic relationships between biological activity and the logarithm of a partition coefficient have previously been quoted^{8,13} since these are plentiful. A regression analysis program has been used to fit the function $\log (a + 10^{-\log P}) + \text{constant}$ to the experimental values of $\log C$ for each example shown in Table I. The values of the coefficient a and of the constant have been chosen so as to minimize the sum of squares of the residuals. The multiple correlation coefficient (MCC) was obtained from the following relationship

$$\text{MCC} = \left(\frac{(\text{total sum of squares} - \text{sum of squares of residuals})}{\text{total sum of squares}} \right)^{1/2}$$

The sum of squares of residuals is a measure of the squares of the differences between predicted and observed activity values summed over all the data points. For a situation such as this in which the number of degrees of freedom is known, comparison of MCC's can safely be undertaken. However, in comparing the equations, allowance should be made for the fact that the old linear and the present equilibrium models require the determination of only two coefficients, whereas the old parabolic relationship required the determination of three. Furthermore, the need to choose between the linear and parabolic equations is obviated when the present equilibrium model is applied.

In a surprising number of cases the linear relationship published in the literature has a gradient which does not differ significantly from unity, and under these circumstances the present model can provide a comparable fit to

Table I.^a Examples of Data Showing the Applicability of the Equation $-\log(1/C) = \log(a + 10^{-\pi}) + \text{Constant}$ for Comparison with the Originally Published Equation

Eq no.	Orig-inal ref	Chemical type	Activity against	No. of data points	Published eq:		Equilibrium model eq				
					(i) $\log(1/C)$ $= a\pi^2 + b\pi + c;$	(ii) $\log(1/C)$ $= a\pi + c$	$-\log(1/C) = \log(a + 10^{-\pi})$ + constant	a	Constant	Multiple correlation coeff	Deg of freedom
29 ^b	10	Phenols	<i>A. niger</i>	26	i	0.927	23	5.9×10^{-4}	-0.672	0.908	24
49 ^b	9	R COO^-	<i>T. interdigitale</i>	14	i	0.972	11	5.6	-4.417	0.970	12
51 ^c	14	Phenyl methacrylates	<i>Strep. faecalis</i>	10	i	0.861	7	3.0×10^{-4}	-0.514	0.817	8
6 ^b	9	R COO^-	<i>A. niger</i>	8	ii	0.974	6	3.9	-2.738	0.990	6
7 ^b	9	R COO^-	<i>T. interdigitale</i>	14	ii	0.994	12	4.3×10^{-1}	-2.854	0.984	12
14 ^b	12	2,4-Bis(XC ₆ H ₄ NH)-pyrimidines	<i>C. albicans</i>	8	ii	0.957	6	2.4×10^{-2}	-3.973	0.932	6
41 ^c	14	Phenyl methacrylates	<i>S. aureus</i>	10	ii	0.966	8	1.0×10^{-5}	-0.305	0.948	8
42 ^c	15	Alkyl- β -naphthols	<i>S. aureus</i>	22	ii	0.898	20	5.0×10^{-6}	0.156	0.908	20
56 ^c	14	Phenyl methacrylates	<i>B. subtilis</i>	10	ii	0.976	8	1.8×10^{-5}	-0.362	0.933	8
57 ^c	14	Phenyl methacrylates	<i>B. cereus</i>	10	ii	0.815	8	1.4×10^{-4}	-0.469	0.847	8
60 ^c	14	Phenyl methacrylates	<i>Sarcina lutea</i>	10	ii	0.849	8	2.4×10^{-3}	-0.858	0.824	8

^aThe symbol π in this table denotes the logarithm of a measured or predicted partition coefficient. The multiple correlation coefficients in the eleventh column are derived from values of proportion explained variation. ^bRefer to ref 8 in the text which contains the published equation. ^cRefer to ref 13 in the text which contains the published equation.

the data. For some of these cases, the above fit to the data would not be observed if the assumption that α could be set directly equal to π were not valid. Examples demonstrating unit gradient have been omitted from Table I, which illustrates results where a less satisfactory fit might be expected using the equilibrium model, either because the originally calculated gradient was not unity or because the original fit included an extra square term. Although these few examples were therefore selected as being unfavorable to the newly proposed equations, one may see that its performance in calculating multiple correlation coefficients is not greatly inferior to the traditional approach.

A broader study of the literature shows that a parabolic function has often been fitted to results where the distribution of the experimental points over the full curve is so limited that its form is ambiguous. For this reason the present function can also provide an adequate or better fit, which it could not do if the original observations had fully defined the parabolic form. It may be significant that such clearly demonstrated parabolas have not been commonly reported.

There is an important implication of eq 4 which can be investigated by inspection of the literature.^{5,8,13} For straight line relationships between activity expressed as $\log(1/C)$ and the logarithm of partition coefficient, it is predicted that the gradient should not be significantly greater than unity under any circumstances. The evidence from the published equations so far studied thoroughly endorses this prediction.

The present study has practical implications when regression equations are to be used to predict the activity of further compounds in systems comprising one of the present models. If the slope does not differ significantly from

unity, it may be possible to extrapolate linearly out of the experimental range already studied. On the other hand, such an extrapolation would be risky if the slope was less than unity, since the present models predict that the gradient should increase to unity in one direction and fall away to zero in the other. In such circumstances a linear extrapolation might have a poorer chance of success.

Acknowledgment. I thank Dr. P. J. Goodford for his advice and criticism. I also thank Mrs. N. Trist for her invaluable assistance and Mr. G. C. Sheppey and Dr. C. R. Beddell for helpful discussions.

References

- (1) P. J. Goodford, *Advan. Pharmacol. Chemother.*, **11**, 51 (1973).
- (2) C. Hansch and T. Fujita, *J. Amer. Chem. Soc.*, **86**, 1616 (1964).
- (3) L. P. Hammett, *J. Amer. Chem. Soc.*, **59**, 96 (1937).
- (4) C. G. Swain and E. C. Lupton, *J. Amer. Chem. Soc.*, **90**, 4328 (1968).
- (5) C. Hansch, *Med. Chem., Ser. Monogr.*, **11**, 296 (1971).
- (6) T. Higuchi and S. S. Davis, *J. Pharm. Sci.*, **59**, 1376 (1970).
- (7) A. J. Clark, *J. Physiol.*, **61**, 530 (1926).
- (8) C. Hansch and E. J. Lien, *J. Med. Chem.*, **14**, 653 (1971).
- (9) O. Wyss, B. J. Ludwig, and R. R. Joiner, *Arch. Biochem.*, **7**, 415 (1945).
- (10) H. G. Shirk and R. R. Corey, *Arch. Biochem. Biophys.*, **38**, 417 (1952).
- (11) W. B. Geiger, *Arch. Biochem.*, **16**, 423 (1948).
- (12) D. Ghosh, *J. Med. Chem.*, **9**, 423 (1966).
- (13) E. J. Lien, C. Hansch, and S. M. Anderson, *J. Med. Chem.*, **11**, 430 (1968).
- (14) R. Woodside, M. Zief, and G. Sumrell, *Antibiot. Chemother.*, **9**, 470 (1959).
- (15) M. L. Khorana, S. Y. Pandit, and A. D. Pishawikar, *J. Pharm. Sci.*, **56**, 993 (1967).