Acknowledgment.—The authors wish to acknowledge support for this research from National Institutes of Health Grant GM-16312.

Appendix

The simulation of Br and I orbitals by 3s and 3p functions may at first appear somewhat controversial. Therefore, a detailed description is given here, along with additional data indicating that this procedure appears to be valid and should give reasonable predictions for the trends noted in this paper.

The orbitals to be fitted were chosen to be Slater functions with orbital exponents given by Slater's rules. The radial part of these orbitals is of the form

$$N_g r^{n_g - 1} \exp(-\zeta_g r),$$

where $N_g = (2\zeta_g)^{n_g + \frac{1}{2}} / \Gamma(2n_g + 1)^{\frac{1}{2}}$, $\Gamma(x)$ being the γ function. It is necessary to use this form since n_g for bromine is nonintegral. The radial part of the simulating orbital is

$$N_s r^2 \exp(-\zeta_s r),$$

where $N_s = (2\zeta_s)^{\frac{5}{2}} / (720)^{\frac{1}{2}}$.

The overlap integral between the above two functions is

$$S = N_g N_s \Gamma(n_g + 4) / (\zeta_s + \zeta_g)^{\frac{n_g + 4}{2}}$$

differentiating S with respect to ζ_s, and setting the result to zero leads to

$$\zeta_s = 7\zeta_g / (2n_g + 1).$$

The overlap integral between the actual Slater orbital and the simulating orbital is 0.998 for bromine and 0.996 for iodine, showing that there is a negligible difference between them.

Table II gives computed and experimental values for the barrier to internal rotation in C_2H_6, EtCl, EtBr, and EtI. The computed barriers are all exaggerated, which is not uncommon in EHT calculations. However, the use of simulated orbitals for bromine and iodine has led to an additional error of only about 1 kcal/mole. Therefore, even if the computed barriers in the thyronine derivatives are in error by a factor of 5, we feel the results for the trends predicted cannot be disputed.

TABLE II
COMPUTED AND EXPERIMENTAL BARRIERS TO INTERNAL ROTATION

Molecule	EHT barrier[a]	Experimental barrier[a,b]
C_2H_6	4.0	2.8
EtCl	5.7	3.7
EtBr	6.8	3.7
EtI	6.8	3.2 ± 0.5

[a] All values are in kcal/mole. [b] Experimental results taken from the tabulation given by J. P. Lowe, *Progr. Phys. Org. Chem.*, **6**, 1 (1968).

Quantitative Structure–Activity Models. Some Conditions for Application and Statistical Interpretation[1]

DONNA R. HUDSON, GEORGE E. BASS, AND WILLIAM P. PURCELL*

Department of Molecular and Quantum Biology, College of Pharmacy, University of Tennessee Medical Units, Memphis, Tennessee 38103

Received May 28, 1970

The basis for a statistical analysis of the Free-Wilson structure–activity model is presented along with an explanation and interpretation of the multiple correlation coefficient, the F test of coefficient significance, and the explained variance. Ill conditioning is discussed as a problem of data suitability and a method for detecting this property is suggested. Two examples illustrate the methodology and interpretation.

With increasing emphasis on the application of mathematical and linear free energy related models to quantitative understanding of structure–activity relationships,[2,3] it becomes necessary to investigate the limitations and utilities of these methods. This work represents an attempt to establish criteria for the validity of the application of the Free-Wilson model.[4] The proposed model and the development of the basis for a statistical analysis of the data will be presented here. The problem of ill conditioning which sometimes occurs with this model will be presented along with two examples using the developed analysis.

The mathematical model developed by Free and Wilson to study structure activity relationships is applicable when dealing with an analogous series of compounds with corresponding biological activity data. A basic assumption of this model is that the activity of each compound can be resolved as the sum of contributions associated with the separate segments of the molecule. As a result, the activity of each compound in a series can be represented in the form of a linear equation as follows:

$$\text{biological activity} = \text{overall average} + \text{contribution of segment 1} + \ldots\ldots + \text{contribution of segment } n \quad (1)$$

* To whom correspondence should be addressed.

(1) This research is being supported by the U.S. Army Medical Research and Development Command (DA-49-193-MD-2779), the Cotton Producers Institute (through the National Cotton Council of America), the National Science Foundation (GB-7383), and a grant from Eli Lilly and Company. This paper is Contribution No. 821 from the Army Research Program on Malaria. Computer facilities were provided through Grant HE-09495 from the National Institutes of Health.

(2) W. P. Purcell, J. A. Singer, K. Sundaram, and G. L. Parks, "Medicinal Chemistry," A. Burger, Wiley, New York, N. Y., Chapter 10, 1970, pp 164–192.

(3) J. M. Clayton, O. E. Millber, Jr., and W. P. Purcell, *Annu. Rep. Med. Chem.*, **1969**, Chapter 27 (1970).

(4) S. M. Free, Jr., and J. W. Wilson, *J. Med. Chem.*, **7**, 395 (1964).

The different segments are those portions of the general structure that change between any two compounds in the series. A number of different substituents appear at each segment; thus, for each molecule, one will have a different linear equation for the activity. The activity contributions of the substituents are the unknowns in the system of independent, linear, inhomogeneous equations which are set up for the series of compounds under consideration.

An arbitrary assumption of the Free–Wilson model is that the total contribution of the substituents of a given segment over the entire series is zero. For example, at segment k one might have four substituents (*e.g.*, H, CH_3, C_2H_5, C_3H_7) with the corresponding activity contributions S_a, S_b, S_c, and S_d and thus require that

$$\sum^{m} (A_i S_a + B_i S_b + C_i S_c + D_i S_d) = 0 \qquad (2)$$

where the summation index i runs over the m compounds in the series and the coefficients, A_i, B_i, C_i, and D_i, are either 1 or 0, depending on whether the corresponding substituent is present in the ith molecule. From this treatment one can see that the substituent contributions at a segment are not linearly independent; one is algebraically dependent and can be expressed as a linear combination of the others. This algebraic relationship is referred to as a symmetry equation.[4] Equation 2 can be solved for one of the substituent contributions, *e.g.*,

$$-S_d = \sum^{m} \left(\frac{A_i}{D_i} S_a + \frac{B_i}{D_i} S_b + \frac{C_i}{D_i} S_c \right) \qquad (3)$$

Thus, for a compound which contains the substituent with contribution S_d, the linear equation is written in terms of S_a, S_b, and S_c, instead of S_d, using eq 3. In this way, the assumption of symmetry is incorporated into the solution.

Under these conditions, a series of m compounds with p segments and a total of n substituents transforms into a system of m equations with $n-p$ independent variables, or unknowns. To solve this system of equations, one must have $m \geq n-p$. The solution yields the activity contributions of the $n-p$ substituents explicitly treated. The symmetry equations can then be used to obtain the p remaining contributions.

As Free and Wilson show,[4] once a solution has been found for a system of equations, the substituents at each segment can then be ranked according to their individual contributions and the values of the substituent activity contributions can be used to predict the activities of untested compounds.

Statistical Analysis.—To determine the success or failure of any series of compounds to fit this additive model, a statistical analysis of the data should be considered. The Free–Wilson model lends itself well to regression analysis, which fits the data to a general linear equation. In order to apply a regression analysis to a system of equations, certain mathematical requirements must be fulfilled.[5] This model does, in fact, meet these requirements.

First, the independent variables are fixed variates and the dependent variables are randomly produced. In the Free–Wilson model, it is obvious that the substituent groups are fixed for any series of compounds tested and, since biological responses are not determined *a priori* to experimentation, they may be considered randomly produced.

Second, for any fixed set of independent variables, the dependent variables associated with this set are normally and independently distributed. If one were to measure repeatedly the biological response of any one compound (assuming identical experimental conditions), this set of responses would indeed be normally distributed with no one measurement affecting any of the others (independence). If one uses an already averaged value for the biological response, this requirement is still met since the sample means of a normally distributed population are also normally distributed.[6]

Finally, for any set of independent variables, the variance of the dependent variables must be the same. Since there is an underlying normal population of biological responses for each set of substituents, the total population of the biological responses will be normally distributed; therefore, the variances can be considered equal for all of the responses.

After performing a regression analysis using the Free–Wilson model, several statistics indicative of the "goodness" of fit are appropriate for consideration; among these are the multiple correlation coefficient, the overall F value for the test of coefficient significance, and the explained variance.

The multiple correlation coefficient, R, gives an indication of the degree of correspondence between the experimentally observed biological responses and those calculated with the proposed linear equation resulting from the regression analysis ($R = 1.0$ indicates perfect correlation). This correlation coefficient is usually used in terms of its square, R^2, because of the similarity in formulas with other statistics. The mathematical formula for R^2 is $\Sigma\hat{y}^2/\Sigma y^2$, where $\hat{y}$ = (calculated response − mean response), and y = (observed response − mean response). As such, R^2 is interpreted as the fraction of the sum of squares of the deviations of observed responses from the mean responses that is attributable to the regression.[7]

The F value is the decision statistic of the F test of significance. The overall F test with this model is a test of the null hypothesis that all of the substituent coefficients (activity contributions) are equal to zero; in other words, the mean biological response would be as good an estimate of the actual response as the response calculated from the linear regression equation. Thus, this value, after tabular interpretation, indicates the significance of the substituent contributions to the activity in a series of compounds. The basic assumption that validates the use of the F test is that the dependent variables are normally and independently distributed, which in fact holds true for the biological responses in the Free–Wilson model. The formula for the F statistic is $[\Sigma\hat{y}^2/(k-1)]/[\Sigma d^2/(n-k)]$, where d = (observed response − calculated response), k = total number of variables (or unknowns) used in the regression, and n =

(5) R. L. Anderson and T. A. Bancroft, "Statistical Theory in Research," McGraw-Hill Book Company, Inc., New York, N. Y., 1952, p 168.

(6) R. R. Sokal and F. J. Rohlf, "Biometry. The Principles and Practice of Statistics in Biological Research," W. H. Freeman and Company, San Francisco, Calif., 1969, p 130.

(7) G. W. Snedecor and W. G. Cochran, "Statistical Methods," 6th ed, The Iowa State University Press, Ames, Iowa, 1967, pp 385–387, 400–402.

total number of data points (compounds) used in the regression. The corresponding level of significance for an F statistic can be found in any table of f distribution values under $(k-1)$ and $(n-k)$ degrees of freedom.[7]

The explained variance gives the fraction of the variance of the biological responses which is attributed to the linear relationship of those substituent contributions, or unknowns, included in the analysis. The formula for calculating this quantity is $1 - |\Sigma d^2/(n-k)|/|\Sigma y^2/(n-1)|$ (where d, n, k, and y are defined above). Even though regression coefficients may prove to be statistically significant with the F test, it is not uncommon to find that the fraction of explained variance is quite small. This would indicate that most of the variance in the responses must be attributed to variables not included in the regression.[7]

The biological response data for a series of compounds can fail an analysis with the Free–Wilson model not only on the basis of the statistical results but also as a result of the system of linear equations being ill conditioned, or unstable. When applying the Free–Wilson model to biological studies, it has been found that it is not uncommon for these systems of equations to be ill conditioned. Ill conditioning can be caused by the disappearance of significant figures during solution of the system and can lead to extreme inaccuracy in calculated coefficients. It is thought that this situation occurs when inverse matrix elements or initial matrix coefficients are extremely small. The coefficients of the matrix to be solved in this example (cross product matrix) are:

$$a_{ij} = \sum_{k=1}^{m} x_{ki}x_{kj} - \frac{\left(\sum_{k=1}^{m} x_{ki}\right)\left(\sum_{k=1}^{m} x_{kj}\right)}{m} \qquad (4)$$

where m = the number of equations (compounds) and x_{ki} = the coefficient of the ith variate of the kth compound. Also, inaccuracy in the initial data may itself produce ill conditioning. Faddeev and Faddeeva state that "it is clear that a system with a matrix possessing such a property [ill conditioning] cannot be solved with any sort of confidence."[8]

It has been found that a reliable way to detect ill conditioning is to perform the regression analysis on the system twice, changing the identity of the dependent substituent at at least one of the segments. If the values of the substituent contributions from these two solutions do not agree to several significant figures, one can consider the system to be unstable and, therefore, the contribution values and relative rankings would be unreliable. It should be noted that rounding errors one might introduce through the symmetry equations will cause uniform but substantial changes in the contribution values of the system; this could lead to an incorrect decision of instability.

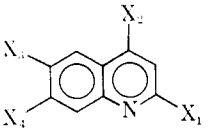
Examples.—To demonstrate the use of the statistical indicators of fit and the problem of ill conditioning, two applications of the Free–Wilson model are presented.

Example 1.—The data used in this example (series I) were selected from those reported by Coatney, *et al.*,[9]

(8) D. K. Faddeev and V. N. Faddeeva, "Computational Methods of Linear Algebra," W. H. Freeman and Company, San Francisco, Calif., 1963, p 121.

(9) G. R. Coatney, W. C. Cooper, N. B. Eddy, and J. Greenberg, "Survey of Antimalarial Agents," Public Health Monograph No. 9, U. S. Government Printing Office, Washington, D. C., 1953, pp 64–74.

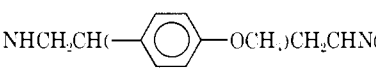
TABLE I
FREE–WILSON DESIGN FOR SELECTED CHLOROQUINE DERIVATIVES EVALUATED AGAINST *Plasmodium gallinaceum*—SUBSTITUTIONAL VARIATION AT FOUR RING POSITIONS



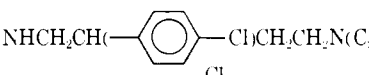
Series I

X_1 = H (A) or CH_3 (B)

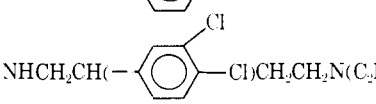
X_2 = $NHCH_2CHOHCH_2N(C_2H_5)_2$ (C),



(D),



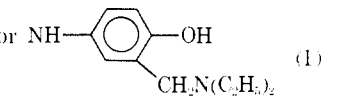
(E),



(F),

$NH(CH_2)_2S(CH_2)_2N(C_2H_5)_2$ (G), $NH(CH_2)_3S(CH_2)_2N(C_2H_5)_2$ (H),

or



(I)

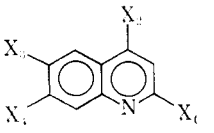
X_3 = H (J) or Cl (K)

X_4 = H (L) or Cl (M)

X_1	X_2							X_3	X_4	log (0.1/METD)[a]
B	C	D	E	F	G	H		J	M	
−0.75	−1	−1	−1	−1	−1	−1		1	1	1.699
−0.75		1						1	1	1.301
−0.75	1							−6	1	1.398
1					1			−6	−13	1.398
1							1	1	1	1.301
−0.75							1	1	1	1.301
1		1						1	1	1.097
−0.75					1			1	1	1.699
1				1				1	1	0.824
1			1					1	1	1.097
−0.75			1					1	1	1.602
−0.75				1				1	1	1.097
−0.75	1							1	1	2.000
1	−1	−1	−1	−1	−1	−1		1	1	1.699

[a] See ref 9.

for the activities of chloroquine derivatives against *Plasmodium gallinaceum*. The activity reported is the minimum effective therapeutic dose (METD), which is the dose required to reduce parasitemia to 25% or less of controls. For analysis, log (0.1/METD) is used as the biological response. A logarithmic form of the activity is used because of the nature of dose–response activity data. The general structure of the compounds is

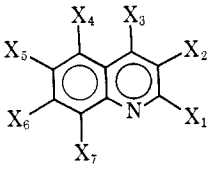


series I

where X_1, X_2, X_3, and X_4 are the segments at which substituent variations occur. The Free–Wilson model

TABLE II

FREE-WILSON DESIGN FOR SELECTED CHLOROQUINE DERIVATIVES EVALUATED AGAINST *P. gallinaceum*—SUBSTITUTIONAL VARIATION AT SEVEN RING POSITIONS

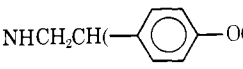


Series II

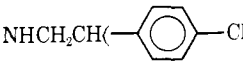
X_1 = H (A) or CH_3 (B)

X_2 = H (C), CH_3 (D)

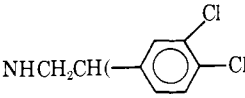
X_3 = $NH(CH_2)_3N(C_2H_5)_2$ (E), $NHCH_2CHOHCH_2N(C_2H_5)_2$ (F), $NHCH[CH_2N(CH_3)_2]_2$ (G),



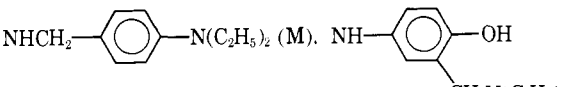
(H),



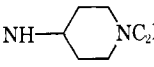
(I),



(J), $NH(CH_2)_2S(CH_2)_2N(C_2H_5)_2$ (K), $NH(CH_2)_3S(CH_2)_2N(C_2H_5)_2$ (L),



(M),



(O)

X_4 = H (P), Cl (Q). X_5 = H (R), Cl (S), OH (T), OCH_3 (U), OCH_2CH_2OH (V)

X_6 = H (W), Cl (X). X_7 = H (Y), Cl (Z), CH_3 (AA)

X_1	X_2	X_3										X_4	X_5				X_6	X_7		Log (0.1/ METD)[a]
A	C	E	F	H	I	J	K	L	M	N	O	P	R	S	U	V	W	Y	AA	
1	1	1										1	1				−0.438	1		1.921
1	1									1		1	1				−0.438	1		1.699
−10.5	1								1			1	1				1	1		0.699
−10.5	1								1			1			1		1	1		0.301
1	1	−1	−3	−2	−2	−2	−2	−2	−2	−4	−2	1	−17	−2	−2	−1	1	1		−0.494
1	1		1									−22	1				1	−21	−1	0.301
1	1			1								1	1				−0.438	1		1.301
1	1										1	1	1				−0.438	1		1.699
1	1		1									1		1			−0.438	1		1.398
1	1									1		1				1	1	1		0.796
1	−2.83						1					1		1			1	1		1.398
1	1									1		1			1		1	1		1.496
1	−2.83							1				1	1				−0.438	1		1.301
1	1							1				1	1				−0.438	1		1.301
1	−2.83			1								1	1				−0.438	1		1.097
1	1						1					1	1				−0.438	1		1.699
1	−2.83					1						1	1				−0.438	1		0.824
1	−2.83				1							1	1				−0.438	1		1.097
1	1				1							1	1				−0.438	1		1.601
1	1					1						1	1				−0.438	1		1.097
1	1										1	1	1				−0.438		1	0.432
1	1		1									1	1				−0.438	1		2.000
1	−2.83									1		1	1				−0.438	1		1.699

[a] See ref 9.

input data for the series is given in Table I. This is a system of 14 equations and 10 unknowns with 4 dependent substituents (corresponding to 4 segments). The symmetry equations (eq 2) for the segments are

Segment	Equation	
X_1	$8A + 6B = 0$	(5)
X_2	$2C + 2D + 2E + 2F + 2G + 2H + 2I = 0$	(6)
X_3	$12J + 2K = 0$	(7)
X_4	$1L + 13M = 0$	(8)

where A, B, etc. denote the substituents. The dependent substituents used in this example are A, I, K, and L. Therefore, when substituent A appears at segment X_1, it is represented as $-(6/8)B$ (from eq 5), when substituent I appears at segment X_2, it is represented as $-1C -1D -1E -1F -1G -1H$ (from eq 6), and so on.

For example, the first compound in series I has substituents A, I, J, and M appearing at segments X_1, X_2, X_3, and X_4, respectively. In the Free–Wilson model, this compound is represented by the equation

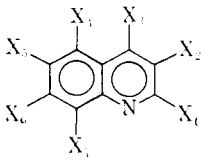
$$-3/4B -1C -1D -1E -1F -1G -1H +1J +1M = 1.699 \quad (9)$$

(since A and I are dependent substituents, their equivalence in terms of the other substituents at their respective segments is used). The coefficients of eq 9 appear in the first row of the model matrix in Table I.

The multiple correlation coefficient of the regression is 0.97, the F test proves significant at a 95% confidence level, and 78% of the variance is explained. Respectively, these statistical values are interpreted as indicating that, in this series of compounds, (a) there is a definite linear relationship between substituent contributions and biological activity, (b) at least one of the substituent contributions substantially affects the biological activity, and (c) most (78%) of the variance in the biological responses is explained by this treatment.

To check the stability of this system, the dependent substituent at segment X_2 was changed (from "I" to "H") and the system was solved again. The change in the identity of a dependent substituent had no substantial affect on the solution values. Therefore, this system is considered to be stable and reliable for a study of the relative effects of different substituents on the biological activity.

Example 2.—The data used in this example were also selected from the activity data reported by Coatney, *et al.*,[9] and the biological response used for analysis was log (0.1 METD) as before. Table II lists the compounds and data for this series which has the general structure



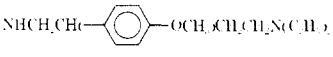
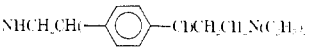
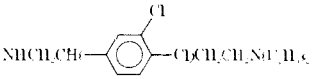
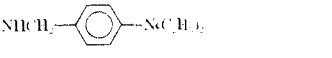
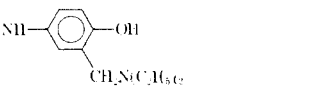
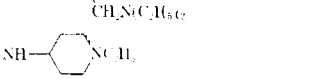
series II

The Free–Wilson model for this series gives rise to a system of 23 equations and 21 unknowns with 7 dependent substituents (Table II). Two solutions for this system were obtained (Table III) by changing the dependent substituents at segments X_3, X_5, and X_7. If one compares these results, it is seen that the change in dependent substituents does have an appreciable affect on the contribution values. This illustrates the effect of an ill conditioned system of equations; there is no unique solution set. Without a unique solution, the substituent contributions are unreliable and, therefore, no sound conclusion can be reached about the resulting connection between changes in structure and changes in activity.

Statistically, an analysis might appear good or satisfactory even though the system is ill conditioned. Each set of contribution values is indeed a valid solution of the system. For this example, the multiple correlation coefficient is 0.99, the overall F test is significant at the 85% confidence level, and the amount of explained variance is 85%! These statistics imply that there is a good linear relationship between structure and activity in this series, but these values are meaningless in this example since the system itself has no unique solution.

In conclusion, when using the Free–Wilson model for structure-activity studies, it is advisable to submit the model to a regression analysis in order to test the basic assumption of activity additivity, using the multiple correlation coefficient, the overall F test for coefficient significance, and the explained variance as statistical

Table III
Contribution Values for Substituents of Series II (Table II)

Segment substituents	Solution set 1 [log (0.1 METD)]	Solution set 2 [log (0.1 METD)]
$(X_1)H$	0.0589	0.0855
CH_3	−0.6183	−0.3978
$(X_2)H$	0.0433	0.0433
CH_3	−0.1225	−0.1225
$(X_3)NH(CH_2)_3N(C_2H_5)_2$	0.5137	0.7314
$NHCH_2CHOHCH_2N(C_2H_5)_2$	0.5430	0.7597
$NHCH[CH_2N(CH_3)_2]_2$	−0.3367	−8.7470
	−0.1243	0.0924
	0.0264	0.2431
	−0.3628	−0.1461
$NH(CH_2)_2S(CH_2)_2N(C_2H_5)_2$	0.3436	0.5603
$NH(CH_2)_3S(CH_2)_2N(C_2H_5)_2$	−0.0221	0.1946
	−0.2424	0.2804
	0.3503	0.5670
	0.2927	0.5094
$(X_4)H$	0.0915	−0.2595
Cl	−2.0132	5.7085
$(X_5)H$	0.0879	−0.1555
Cl	−0.4123	−0.6557
OH	1.2909	6.6447
OCH_3	−0.4120	−0.6553
OCH_2CH_2OH	−1.1359	−1.3793
$(X_6)H$	0.1832	0.1832
Cl	−0.0823	−0.0823
$(X_7)H$	0.0467	0.3977
Cl	0.2398	−7.4818
CH_3	−1.2201	−0.8692

indicators. It is also advisable, with any series of compounds, to check the stability of the system by changing dependent substituents at various segments, solving the system of equations again, and comparing the two sets of solution values. With an unstable system of equations, there is no unique set of solution coefficients; thus, the substituent contributions are unreliable and no sound conclusion can be reached about the resulting connection between changes in structure and changes in activity.

Acknowledgment.—The authors would like to express their gratitude to Mr. Walter Lafferty of the University of Tennessee Medical Units Biometric Computer Center for fruitful discussions during the early stages of this work.

Structure–Activity Correlations for Anticonvulsant Drugs

ERIC J. LIEN

School of Pharmacy, University of Southern California, University Park, Los Angeles, California 90007

Received April 10, 1970

The anticonvulsant activity of series of drugs in mice and in rats against electroshock and pentylenetetrazole-induced seizures has been found to be highly correlated with the log P values of the drugs, where P is the 1-octanol-water partition coefficient. From the data on hand, linear dependence on log P is found for the anti-electroshock test in mice and the pentylenetetrazole protection test in rats, where the slope of the regression line associated with log P is about 0.6 ± 0.2. Parabolic dependence on log P is found for the antielectroshock activity in rats with an optimum lipophilic character (log P_0) of 1.75.

It was estimated that more than 20,000 compounds had been screened for anticonvulsant action in the last 10 years,[1] but many of them were not active or had very low activity. The need for better anticonvulsants to cope with epileptic seizures is reflected by continuous publications in this field. Unfortunately, not only is the mechanism of anticonvulsant action unknown, but also, few guide lines are available to help medicinal chemists in searching for better and safer anticonvulsants. The "common denominator" of clinically useful anticonvulsants has been known for some time.[2,3] However, no quantitative correlation of the relative potency of these drugs with the chemical structure has been satisfactory.

Recently Andrews examined the anticonvulsant activity of a number of potent anticonvulsants and tried to correlate it with the atomic charges of the so-called "biological active center" obtained from MO calculations and with the dipole moments of the drugs.[4] No significant correlation was obtained. The H-bonding atoms, although common to all the drugs studied, were not proven responsible for variations in activity.

In view of the fact that the anticonvulsant activity was studied *in vivo* and that the availability of the drug at the biophase and the receptor site must be considered before any meaningful structure–activity correlation can be obtained,[5] the author wishes to show that the variation in the anticonvulsant activity of series of potent drugs in 4 different tests can be correlated satisfactorily with log P (P = 1-octanol–H_2O partition coefficient).

Methods

The antisupramaximal-electroshock data in mice, the atomic charge and the dipole moments were taken from Andrews' paper.[4] The antielectroshock data in rats and in mice were from the work of Chen and Ensor.[6] The data of pentylenetetrazole protection were from the report of Swinyard.[7] For the details of the biological tests the original articles should be consulted. The log P values of 4 compounds were experimentally determined by Hansch's group and the others were calculated from the log P values of the parent molecules and the π constants of the substituents[8–11] (see Table I). The following log P of π values were used in the calculation of the log P values: π of oxazolidine-2,4-dione = $\pi_{OCO} + \pi_{CH_3CON} = (-1.14) + (-0.79) = -1.93$; $\pi_{CH_3CO} = -0.55$; $\pi_{NHCONH_2} = -1.01$; $\pi_{Br(aliphatic)} = 0.60$; $\pi_{hydantoin}$ = log P of 5-ethyl-5-phenylhydantoin $- (\pi_{Et} + \pi_{Ph}) = 1.53 - (1.00 + 1.77) = -1.24$; $\pi_{succinimide}$ = log P of 2-ethyl-2-phenylglutarimide $- (\pi_{Et} + \pi_{Ph} + \pi_{1/6\ cyclohexane}) = 1.90 - (1.00 + 1.77) - 1/6(2.51) = -1.29$; $\pi_{Ph} = 1.77$ (on the heterocyclic ring); $\pi_{Ph} = 2.13$ (for terminal substituents); π_{Me} (on N) = 0.56; π_{Me} (on C) = 0.50.

The equations correlating the antielectroshock and the antipentylenetetrazole activity in mice and rats with the physicochemical constants (see Table II) were derived *via* the method of least squares using an IBM 360/65 computer.

Results and Discussion

The equations obtained from the regression analysis are summarized in Table II. The results are not presented where no better correlation coefficient than 0.85 could be obtained. From eq 1–3 it is clear that neither the dipole moment nor the charge on the "biological activity center" (EHT, CNDO/2) can account for the variations in the anticonvulsant activity ($r < 0.4$).

(1) R. K. Richards, *Clin. Pharmacol. Ther.*, **10**, 602 (1969).
(2) T. C. Daniels and E. C. Jorgensen in "Textbook of Organic Medicinal and Pharmaceutical Chemistry," C. O. Wilson, Gisvold, and R. F. Doerge, Ed., 5th ed, Lippincott Co., Philadelphia, Pa., 1966, p 403.
(3) W. C. Cutting, "Handbook of Pharmacology," 4th ed, Appleton-Century-Crofts, New York, N. Y., 1969, p 669.
(4) P. R. Andrews, *J. Med. Chem.*, **12**, 761 (1969).
(5) E. J. Lien, *J. Amer. Pharm. Educ.* **33**, 368 (1969).
(6) G. Chen and C. R. Ensor, *Arch. Neurol. Psychiat.*, **63**, 56 (1950).
(7) E. A. Swinyard, *J. Amer. Pharm. Ass.*, **38**, 201 (1949).
(8) T. Fujita, J. Iwasa, and C. Hansch, *J. Amer. Chem. Soc.*, **86**, 5175 (1964).
(9) J. Iwasa, T. Fujita, and C. Hansch, *J. Med. Chem.*, **8**, 150 (1965).
(10) C. Hansch, private communication.
(11) C. Hansch, A. R. Steward, S. M. Anderson, and D. Bentley, *J. Med. Chem.*, **11**, 1 (1968).