tives, is always located a little below the value of experimental pH (*i.e.*, between $pK_A = 6.5$ and 7.7) depending on the experimental conditions and test organisms. In spite of the exception of the N^1-benzoyl derivatives, our over-all results indicate that the pK_A value should be as close to body pH as possible in order to obtain a maximal chemotherapeutic activity.

Although the ρ value for the N^1-heterocyclic derivatives, 0.605, is very similar to that obtained for the sulfanilanilides against *E. coli*, 0.7 (ρ value in eq 15c divided by $\rho_A = 1.88$), the ρ value for the N^1-benzoylsulfanilamides against the same *E. coli*, 2.6, is considerably larger than the other two. As described above, the ρ value for the latter is not highly reliable so that the difference in ρ may not be worth trying to rationalize. However, in our procedure, the ΔpK_A or σ term cannot be assigned only to the contribution of an electronic demand of the drug molecule at the site of action. If the transfer process from outside the cell to the intracellular site of action through many partitionings and adsorption and desorption processes *via* biological membranes is governed to some extent by an electronic effect of the substituent, this effect is contained in the ρ value together with the effect at the site of action. Since we are unable to separate the roles of the ΔpK_A term, the difference in ρ values for different series would not necessarily indicate the difference in the essential electronic demand of the drugs at the site of action.

The above analyses provide another illustration of the great practical advantage of the use of the extrathermodynamic approach[25] to structure–activity problems. The role of the hydrophobic property of the molecule in the bacteriostatic activity and the protein binding is nicely delineated by means of π. The analysis, where the effects of substituent on ionization are separated from other electronic effects of substituents, is able to describe the pK_A dependence of the bacteriostatic activity. It also shows, in a procedure independent from those of earlier workers,[2,3] that the maximal antibacterial activity is exerted by drugs having an optimal pK_A value. This procedure should help in designing new sulfonamide drugs with optimal pK_A and π_0. It should also aid in understanding the pharmacokinetic mechanism underlying sulfonamide chemotherapy when a comprehensive set of biological data and physicochemical constants for *in vivo* properties are available, and an appropriate model can be chosen for *in vivo* phenomena such as curative effect, metabolic process, and renal excretion. Thus, if this procedure could be combined with the recently developed method by Krüger-Thiemer and Bünger,[26] a relationship between dosage schedule and molecular structure of the sulfonamides could be integrated so that an ideal dosage schedule for a new drug could be predicted from structural parameters such as $\log P$ and ΔpK_A.

Acknowledgment.—The authors wish to express their sincere thanks to Professors Tetsuo Mitsui and Minoru Nakajima for their support of this work.

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

Relationships among Current Quantitative Structure–Activity Models[1]

JUDITH A. SINGER AND WILLIAM P. PURCELL

Department of Medicinal Chemistry, University of Tennessee College of Pharmacy, Memphis, Tennessee 38103

Received June 28, 1967

Structure–activity models falling into two categories are compared. One category includes those models in which the observed biological activity is expressed as a function of group contributions to the activity and the other includes the Hansch substituent constant model. It is demonstrated that, if the biological activity is a parabolic function of Hansch's substituent constant, π, the model assuming additive and constant contribution from each group is not appropriate, but a model previously successful in a specific instance is analogous to the Hansch equation. If the π^2 term is not significant, however, the model assuming additive and constant contribution is appropriate when the biological activity is dependent on π and/or σ.

The recent success of attempts to express quantitatively the relationship of chemical structure to biological activity is most encouraging to the medicinal chemist who wishes to approach drug design rationally. The quantitative models for structure–activity relationships of related series of molecules fall into two broad categories. (A) There are mathematical models in which the observed biological activity is expressed as a function of parameters assigned to each substituent group and/or the parent portion of the molecule; the values of these parameters are obtained, after a particular model has been selected, by fitting the experimentally observed activities of a series of molecules using the method of multiple regressions. (B) The second category is comprised of linear free-energy relationships which ascribe the biological activity of a molecule to contributions from various free-energy-related physicochemical parameters of the substituents, the constants associated with each physicochemical parameter being generated by regression analysis for the biologically tested molecules.

Examples of the first approach include those of Free and Wilson[2] and Kopecký and co-workers.[3,4] The method of Free and Wilson[2] is based upon an additive mathematical model in which a particular substituent in a specific position is assumed to make an additive and constant contribution to the biological activity of a molecule in a series of chemically related

(1) This research is being supported by the U. S. Army Medical Research and Development Command (DA-49-193-MD-2779) and the National Science Foundation (GB-4453). This paper is Contribution No. 222 from the Army Research Program on Malaria.

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

(3) K. Boček, J. Kopecký, M. Krivucová, and D. Vlachová, *Experientia*, **20**, 667 (1964).

(4) J. Kopecký, K. Boček, and D. Vlachová, *Nature*, **207**, 981 (1965).

molecules. It was recognized that not all biological activities could be described by this additive model, and failure of the method was suggested to be diagnostic of such instances. The method of Kopecký and coworkers tested four equations for the quantitative expression of the difference in log LD_{50} of *para*-[3] and *meta*-disubstituted[4] benzenes from log LD_{50} of benzene. The biological activity (BA) associated with the substituents in positions X and Y was expressed by the following equations where the x and y subscripts refer

$$BA = a_x + a_y \quad (1)$$

$$BA = d_x d_y \quad (2)$$

$$BA = b_x + b_y + e_x e_y \quad (3)$$

$$BA = b_x + b_y - e_x e_y \quad (4)$$

to the contribution of a particular substituent in, respectively, the X or Y position. Neither the additive model, eq 1, similar to the Free and Wilson approach, the product model, eq 2, nor the combined model described by eq 4 was found appropriate for description of the activity. The combined model described by eq 3, however, gave a statistically significant correlation of the data for both *para*[3] and *meta*[4] compounds. It was thought[3,4] that there could possibly be a relationship between the successful mathematical model (eq 3) and the linear free-energy relationships of the type described in (B). An attractive feature of models in category A is that no physicochemical parameters need be determined for the substituents; a successful correlation of biological activity with the numerical parameters associated with various substituents can rank the structural changes per position, by estimating the amount of biological activity attributed to each change and offer a guide for the future synthesis and testing of other compounds in the series.

An outstanding example of the second approach may be found in the extensive work of Hansch and coworkers.[5] The ρ–σ–π analysis for correlation of biological activity and chemical structure has been successfully applied to problems as varied as enzymatic reaction mechanisms,[6] correlation of localization rates of benzeneboronic acids in brain and tumor tissue,[7] and structure–activity relationships of penicillin derivatives.[8] Equation 5[5] is the basic expression used in such correlations. C_x is the molar concentration of a derivative in a family of related compounds causing an equivalent biological response; π is the free-energy-related substituent constant defined as the logarithm of the partition coefficient of the derivative minus the logarithm of the partition coefficient of the parent compound and is related to hydrophobic bonding of the substituent; σ is the well-known Hammett constant, a free-energy-related electronic-substituent constant. The constants a, b, ρ, and c are generated by regression analysis of the equations for the biologically tested derivatives in a series. For molecules with more than one position of substitution, π and σ values are usually added for the substituents. The basic equation may simplify in some instances[5] to eq 6, 7, 8, or 9.

$$\log 1/C_x = -a\pi^2 + b\pi + \rho\sigma + c \quad (5)$$

$$\log 1/C_x = a\pi + b \quad (6)$$

$$\log 1/C_x = -a\pi^2 + b\pi + c \quad (7)$$

$$\log 1/C_x = \rho\sigma + c \quad (8)$$

$$\log 1/C_x = a\pi + \rho\sigma + c \quad (9)$$

Of these four equations, eq 7, which describes a parabolic dependence of biological activity on π, frequently gives the statistically evaluated best fit,[9] especially in complex systems such as whole animals or cells.[10]

Hansch's ρ–σ–π analysis may serve both to guide the medicinal chemist in future synthesis and testing of other compounds in the series and to untangle the roles of hydrophobic, electronic, and steric factors in drug–receptor interactions. The method does require experimental π and σ values, and, while the approximately additive nature of these values allows prediction of π and σ values for a great many substituents without resort to direct experimental determination, there are limits to this prediction. A series of molecules of biological interest might have complex substituents for which π and σ values are not available. It is conceivable that if the π and σ values required experimental determination to allow application of the ρ–σ–π analysis, the mathematical models described in (A) would be more attractive for use as a guide to further work.

It therefore becomes of interest to compare the two approaches (A and B) and to investigate the implications of the comparisons for subsequent applications of the models.

The basic assumption of the Free and Wilson[2] approach is that the BA contributed by each substituent is additive and constant regardless of substituent variation in the rest of the molecule. In view of the frequency of occurrence of a parabolic relation of π to biological activity (eq 7) found by Hansch and coworkers,[9,10] it becomes of interest to investigate the applicability of the Free and Wilson[2] assumptions in such a situation. The question is, "if the observed BA of the molecules in a series is indeed a parabolic function of π and if the additivity of π values is a valid approximation, does the Free and Wilson[2] assumption of additive and constant contribution for each substituent also apply?"

Consider a molecule with two positions, X and Y, having groups x and y substituted, with π values of π_x and π_y, respectively. Further, assume that eq 7 is applicable, *i.e.*, the biological activity is indeed a parabolic function of π. Following Hansch's assumption of additive π values, the BA of the molecule becomes (from eq 7)

$$BA = -a(\pi_x + \pi_y)^2 + b(\pi_x + \pi_y) + c \quad (10)$$

or

$$BA = (-a\pi_x^2 + b\pi_x) + (-a\pi_y^2 + b\pi_y) - 2a\pi_x\pi_y + c \quad (11)$$

It is immediately apparent on inspection of eq 11 that when biological activity depends parabolically on π, the activity contribution of one substituent is not independent of the π value of the other substituent.

(5) C. Hansch and T. Fujita, *J. Am. Chem. Soc.*, **86**, 1616 (1964).
(6) C. Hansch, E. W. Deutsch, and R. N. Smith, *ibid.*, **87**, 2738 (1965).
(7) C. Hansch, A. R. Steward, and J. Iwasa, *Mol. Pharmacol.*, **1**, 87 (1965).
(8) C. Hansch and A. R. Steward, *J. Med. Chem.*, **7**, 691 (1964).
(9) C. Hansch, A. R. Steward, J. Iwasa, and E. W. Deutsch, *Mol. Pharmacol.*, **1**, 205 (1965).
(10) C. Hansch and E. W. Deutsch, *Biochim. Biophys. Acta*, **112**, 381 (1966).

The BA is composed of four additive components: a constant term, c, associated with the parent portion of the molecule; a term dependent on the substituent at X only, $-a\pi_X^2 + b\pi_X$; another term dependent on the substituent at Y only, $-a\pi_Y^2 + b\pi_Y$; and a cross-product term which shows the mutual dependence upon substitutional variation at X and Y, $-2a\pi_X\pi_Y$. The Free and Wilson[2] assumption of additive and constant activity contribution associated with each substituent is therefore not appropriate for the biological activity of a series of molecules which depends parabolically on π.

It is interesting to point out the analogy between the terms in eq 11 and those in the independently reported expression which Kopecký, *et al.*,[3,4] found to be successful, eq 3; compare[11] $-a\pi_X^2 + b\pi_X$ with b_x, $-a\pi_Y^2 + b\pi_Y$ with b_y, and $-2a\pi_X\pi_Y$ with e_xe_y. This suggests that eq 3 has some physical significance and is related to the linear free-energy models.

Equation 3 may be extended readily to describe a series of molecules with three or more substituent positions, the biological activity of which depends parabolically on π.

Generally, Hansch's eq 7 may be written

$$BA = -a(\Sigma\pi_k)^2 + b\Sigma\pi_k + c \qquad (12)$$

for n substitutional positions, which can be expanded

$$BA = -a(\pi_1^2 + \pi_2^2 + \pi_3^2 + \ldots + 2\pi_1\pi_2 + 2\pi_2\pi_3 + 2\pi_1\pi_3 + \ldots) + b(\pi_1 + \pi_2 + \pi_3 + \ldots) + c \qquad (13)$$

and rewritten

$$BA = (-a\pi_1^2 + b\pi_1) + (-a\pi_2^2 + b\pi_2) + (-a\pi_3^2 + b\pi_3) + \ldots + 2\pi_1\pi_2 + 2\pi_2\pi_3 + 2\pi_1\pi_3 + \ldots + c \qquad (14)$$

One can see from eq 14 that, generally, the biological activity of a molecule is a function of n independent terms, a constant c, and n mutually dependent cross products. In Kopecký's notation, eq 14 could be written

$$BA = b_1 + b_2 + b_3 + \ldots + c_1c_2 + c_2c_3 + c_1c_3 + \ldots \qquad (15)$$

Corresponding mathematical models may be similarly derived for molecules with biological activity dependent upon π and σ according to eq 5, 6, 8, and 9 as described by Hansch, *et al.*[5]

In particular, when the most general of these equations (eq 5) applies, it will be seen that the biological activity of a molecule with substituent positions X and Y may be described in terms of four components: c, $-2a\pi_X\pi_Y$, $-a\pi_X^2 + b\pi_X + \rho\sigma_X$, and $-a\pi_Y^2 + b\pi_Y + \rho\sigma_Y$. The first two components are identical with their counterparts derived from eq 7, while the last two terms, $\rho\sigma_Y$, being determined independently by substitution at X and Y, respectively, are similar to the terms $-a\pi_X^2 + b\pi_X$ and $-a\pi_Y^2 + b\pi_Y$ derived from eq 7 in that they are dependent only on the substituent at X or Y, respectively, although they are different in that they each contain a $\rho\sigma$ term.

Equation 15, then, should also be an appropriate mathematical model for describing the biological activity of a series of molecules which satisfies eq 5, as well as for activities satisfying eq 7.

One cannot, of course, distinguish merely from the success of eq 15 in correlating biological activities whether eq 5 or 7 best describes the biological activity of the series, or indeed whether eq 5 or 7 would describe the activity at all. In view of the success of eq 5 and 7 in correlating biological activity, however, application of eq 15 when the π and σ values necessary for a ρ-σ-π analysis are not readily at hand would seem to be justified.

Equations 6, 8, and 9, respectively, yield for a molecule substituted at X and Y the following expressions of activity: $a\pi_X + a\pi_Y + b$, $\rho\sigma_X + \rho\sigma_Y + c$, and $a\pi_X + a\pi_Y + \rho\sigma_X + \rho\sigma_Y + c$. It is seen that none of the individual terms in these expressions depend on more than one substituent. The original assumption of constant, additive contribution of each substituent made in the Free and Wilson[2] method and implicitly in the additive model (eq 1) of the Kopecký[3,4] approach seems appropriate, therefore, for biological activities which satisfy eq 6, 8, or 9.

A slight nonlinear dependence of BA on the electron density (corresponding to the σ term) of the nitrogen atom of amines in their enzymatic demethylation was shown to be significant.[12] Although examples of correlations between BA and a σ^2 term are rare, one should recognize that an expression analogous to eq 11 may be developed by simply including appropriate σ and σ^2 terms. The corresponding Kopecký[3,4] equation would then be

$$BA = b_x + b_y + c_xe_y + f_xf_y \qquad (16)$$

where f_xf_y represents an additional cross product.

In applying these multiple-regression analyses, one should bear in mind that, when the BA is a parabolic function of π, one should not expect the Free and Wilson[2] method to hold, but Kopecký's[3,4] model should apply; the rare instance of the dependence of the BA on σ^2, as well as on π^2, π, and/or σ, may be accomodated by eq 16. On the other hand, when the π^2 term is not significant, but one of the other ρ-σ-π equations expresses the biological activity, the Free and Wilson[2] model seems to be a reasonable one. All of these models require some degrees of freedom (*i.e.*, more equations than unknowns) and they should be statistically evaluated[13] to determine the significance of the correlation.

In view of the relationship between the mathematical models of biological activity and those based on linear free-energy relationships, the biological-response parameter chosen for correlation with the mathematical models might best be selected by the criteria applied to those selected for the linear free-energy relationships. These parameters as selected by Hansch and coworkers[5] are usually negative logarithms of a molar concentration necessary to achieve a constant equivalent response.

Acknowledgment.—The authors wish to thank K. Sundaram for reading and commenting upon the manuscript.

(11) The constant term c of eq 11 has no counterpart in eq 3, since the parent compound's biological activity was compensated for in the expression of biological activity chosen for the model.

(12) C. Hansch, A. R. Steward, and J. Iwasa, *J. Med. Chem.*, **8**, 868 (1965).

(13) C. K. Hancock, *J. Chem. Educ.*, **42**, 608 (1965).