Journal of Medicinal Chemistry

VOLUME 12, NUMBER 2 MAY 1969

Correlation of Biological Activity with Chemical Structure. Use of Molar Attraction Constants

JAMES A. OSTRENGA

Institute of Pharmaceutical Sciences, Syntex Research, Stanford Industrial Park, Palo Alto, California 94304

Received September 26, 1968

Molar attraction constants [$F = (EV)^{1/2}$], a physical constant derived from the solubility parameter ($\delta = F/V$), was taken to be indicative of the relative degree of drug–receptor interaction for related compounds. For several different groups of structurally related compounds, a high degree of correlation of F with biological activity was obtained in each case from a least-squares fit of the data to a first-order equation indicating that F may be a useful parameter for correlating chemical structure with activity. These molar attraction constants are available for most organic functional groups, are apparently additive on a constitutive basis, and are thus readily obtained for many compounds by simple calculation. Although F may merely be related to the relative lipophilicity of a compound in a series, its use in correlating biological activity with chemical structure may be of value since no experimentation is required. The proposed method of correlation appears to have predictive capability.

A great deal of evaluating and screening of potentially active drugs within numerous pharmacological classes has been done, but unfortunately many of these data have not been published. Moreover, the reported data have been obtained from semiquantitative and relatively imprecise bioassays. It is apparent that one of the main impediments to progress with structure–activity correlations is the lack of quantitative pharmacological potency data for structurally related drugs.

In the past, investigators have shown that relationships between observed biological activity and an experimentally determined parameter do indeed exist for some structurally related compounds or for some compounds with common pharmacological actions. Correlations have been made with partition coefficients, degree of binding to various protein fractions, lipid solubility, kinetic rate of biotransformation, and other physical parameters. Often structure–activity relationships have been obtained strictly by the methodical compilation of bioassay data. Conclusions have then been drawn from these data by noticing the relative changes in biological potency due to the location and nature of organic functional groups. It has been common to utilize a stereochemical "fit" hypothesis based on the classical notion that a drug must meet certain stereochemical requirements at a substrate site. Conclusions have then been drawn from these data by observing the effect of change in the structure and location of functional groups on biological potency. That is, the correlation comes after the fact.

Ideally, structure–activity relationships should attempt to explain as well as correlate. They should contribute insight into a mechanism of action at the molecular level as well as help define the chemical and physical nature of the biological structure involved in drug action. Such a method would hopefully also have some predictive capability. It is only recently that this subject has been approached on a theoretical basis.[1] One of the most recent and successful methods for correlating biological activity with the chemical structure and physical properties of drugs is that of ρ-σ-π analysis as introduced by Hansch and coworkers,[1b,2] where a mathematical and theoretical treatment of three parameters has been shown to give quite reasonable correlations.

Many biologically active compounds owe their activity to a capacity for participating in a chemical reaction or physically interacting with cell constituents. Variations in the biological activity of closely related compounds in those cases where absorption is not the controlling factor most probably are due simply to differences in the affinity between the drug molecules and a substrate. In order to evaluate the extent of interaction between the drug molecule and a substrate, it was theorized that a consideration of a parameter which would be a measure of the attractive forces involved in this interaction might have significance. The thermodynamic free energy of interaction is such a parameter but its determination would require experimentation involving equilibrium measurements. It would be more desirable to use a parameter that could be readily calculated.

It was assumed that the stronger the interaction between the drug and the biological substrate, the greater the intrinsic activity. Approaching the problem from this viewpoint led to the consideration of a theory which has been used to explain some solubility phenomena, a situation where electrostatic attractive forces play an important role. A theoretical consideration of the attractive forces and their relationship to the enthalpy of mixing for a two-component regular solu-

(1) (a) E. R. Garrett, O. K. Wright, G. H. Miller, and K. L. Smith, *J. Med. Chem.*, **9**, 203 (1966), (b) C. Hansch and T. Fujita, *J. Am. Chem. Soc.*, **86**, 1616 (1964); (c) A. Cammarata, *J. Med. Chem.*, **10**, 525 (1967).

(2) (a) C. Hansch and R. A. Steward, *ibid.*, **7**, 691 (1964); (b) C. Hansch and S. M. Anderson, *ibid.*, **10**, 745 (1967); (c) C. Hansch, R. M. Muir, T. Fujita, P. P. Maloney, F. Geiger, and M. Streich, *J. Am. Chem. Soc.*, **85**, 2817 (1963).

tion has been given by Hildebrand and Scott.[3] For many organic substances the main contribution to the total attractive forces existing between two molecules is made by dispersion forces, the main exception being those systems where hydrogen bonding forces or dipole interactions predominate. Even for the relatively polar organic liquid acetone, 96% of the total attractive force between two acetone molecules is due to dispersion forces.[4] The energy required for the formation of physical bonds due to dispersion forces is derived from the pairing of dipoles momentarily induced by perturbations in the electron clouds of molecules in close proximity. These forces are nondirectional in nature and are additive over all pairs of molecules in a system.

The over-all distribution and nature of these attractive forces will be reflected in the thermodynamic free energy of interaction, ΔG_i, the ideal quantity to examine and one which is a direct indication of the degree of interaction where a large and negative free-energy change corresponds to a strong interaction or high affinity. Thus, based on the aforementioned assumptions, the most biologically active compound of a series will correspond to the situation where this thermodynamic change is at a maximum. For such an interaction process, the thermodynamic enthalpy (ΔH_i) and entropy (ΔS_i) of the Gibbs equation (1) may be either

$$\Delta G = \Delta H - T\Delta S \tag{1}$$

positive or negative and the sign may depend on the role of the interaction medium. For biological systems, this medium is most often water. In those cases where the enthalpy is positive, one approach may be to define this thermodynamic quantity in terms of other physical parameters by an equation derived by Hildebrand and Scott[5] and Scatchard[6] for two-component regular solutions

$$\Delta H_i \simeq (V_1X_1 + V_1X_1)\phi_1\phi_2(\delta_1 - \delta_2)^2 \tag{2}$$

where V = molar volume, X = mole fraction, ϕ = volume fraction, $\delta = (E/V)^{1/2}$ = solubility parameter, and the subscripts 1 and 2 refer to the two interacting species. In the derivation of eq 2, it was assumed that the geometric mean rule holds for the relationship between the potential energies (E) or the cohesive energy densities (E/V) of the individual components and the interaction mixture, that the interaction forces are central and additive, that there is no volume change ($\Delta V \simeq 0$), and that mixing is random. The solubility parameter has found some practical use in predicting the mutual compatibility of organic solvents and in selecting solvents for various synthetic polymers.[7] Mullins[8] has suggested its usefulness in estimating activity coefficients in an attempt to correlate thermodynamic with biological activity, while others[9] have indicated that a relationship exists between solubility parameters and membrane-transport properties. It may also be defined as

$$\delta = (E/V)^{1/2} = F/V \tag{3}$$

where $F = (EV)^{1/2}$ and is called the molar attraction constant. Substitution for δ in eq 2 and substitution for ΔH_i in the Gibbs equation gives

$$\Delta G_i \simeq (V_1X_1 + V_2X_2)\phi_1\phi_2\left[\frac{F_1}{V_1} - \frac{F_2}{V_2}\right]^2 - T\Delta S_i \tag{4}$$

Equation 4 thus relates the molar attraction constant to the thermodynamic free energy of interaction for those systems where the enthalpy term is positive and where the aforementioned assumptions are approximately true. If one now assumes that the molar volume of that part of the drug molecule which is involved in the interaction is approximately equal to the molar volume of the effective participating biological site, then the free energy at equilibrium will be most favorable and will approach a minimum when ΔH_i approaches zero or when F_1 approaches F_2. Consequently, if the magnitude of the biological response is directly related to the degree of interaction, the most biologically potent compound of a series will correspond to the one where the change in the free energy of interaction is maximized or when the molar attraction constant of the drug and substrate are equal.

By this analysis, the molar attraction constant is taken to be a physical parameter which is a measure of the intermolecular attractive forces of a chemical species relative to a second entity. In this regard, F assumes a physical significance analogous to the van der Waals constant (a)[5,10] where a represents the interaction between two molecular species. The molar attraction constants for various organic compounds may be obtained from a knowledge of their solubility parameters and molar volumes (eq 3). Methods for determining or calculating solubility parameters have been discussed elsewhere.[7a] Since the potential energy and molar volume are extensive thermodynamic quantities, they are additive on a molar basis and thus so are the molar attraction constants. That is, since

$$E_t = n_1E_1 + n_2E_2 - \Delta H_t \tag{5}$$

where E_t = total attractive energy of the system, n = moles of component 1 or 2, and ΔH_t = total enthalpy of the system and since

$$\Delta H_t = (n_1V_t + n_2V_2)\phi_1\phi_1(\delta_1 - \delta_2)^2 \tag{6}$$

then substituting for ΔH_t in eq 5 gives

$$E_t = n_1E_1 + n_2E_2 - (n_1V_1 + n_2V_2)\phi_1\phi_2(\delta_1 - \delta_2)^2 \tag{7}$$

Substituting for volume fraction, $\phi_1 = v_1/(v_1 + v_2)$, and similarly for ϕ, where v is the volume of 1 or 2, yields

$$E_t = n_1E_1 + n_2E_2 - \frac{(n_1V_1 + n_2V_2)v_1v_2}{(v_1 + v_2)^2}(\delta_1 - \delta_2)^2 \tag{8}$$

(3) J. H. Hildebrand and R. L. Scott, "The Solubility of Non-Electrolytes," 3rd ed, Dover Publications, Inc., New York, N. Y., 1964, Chapter III.

(4) P. A. Small, *J. Appl. Chem.*, **3**, 71 (1953).

(5) J. H. Hildebrand and R. L. Scott, ref 3, p 129.

(6) G. Scatchard, *Chem. Rev.*, **8**, 321 (1931).

(7) (a) J. Brandrup and E. H. Immergut, "Polymer Handbook," Interscience Publishers, Inc., New York, N. Y., 1966, p IV, 341; (b) H. Burrell, *Offic. Dig. Federation Soc. Paint Technol.*, **29**, 1069 (1957).

(8) L. J. Mullins, *Chem. Rev.*, **54**, 289 (1954).

(9) S. A. Khalil and A. N. Martin, *J. Pharm. Sci.*, **56**, 1225 (1967).

(10) J. J. Van Laar, *Z. Physik. Chem.*, **72**, 723 (1910).

Letting $v = nV$, substituting $(E/V)^{1/2}$ for δ, and rearranging gives

$$E_tV_t = (n_1E_1 + n_2E_2)(n_1V_1 + n_2V_2) - n_1V_1n_2V_2\left[\frac{E_1}{V_1} - \frac{2(E_1V_1E_2V_2)^{1/2}}{V_1V_2} + \frac{E_2}{V_2}\right] =$$

$$n_1^2E_1V_1 + 2n_1n_2(E_1V_1 + E_2V_2)^{1/2} + n_2^2E_2V_2 =$$

$$[n_1(E_1V_1)^{1/2} + n_2(E_2V_2)^{1/2}]^2(E_tv_t)^{1/2} = n_1(E_1V_1)^{1/2} + n_2(E_2V_2)^{1/2} \text{ or } F_t = n_1F_1 + n_2F_2 \quad (9)$$

Small[4] has demonstrated that besides being additive on a molar basis, the molar attraction constant appears to be additive on a constitutive and atomic basis, and a table giving molar attraction constants for common functional groups has been reported.[4,7a] Thus, F can be calculated for many organic compounds merely by adding the respective values for the functional groups comprising their chemical structure.

Results and Discussion

The use of molar attraction constants to correlate structure with activity was tested using published biological data for six different classes of compounds. An actual tabulation of biological activities and calculated molar attraction constants is given for only one set of data in order to not unduly lengthen this report by completely identifying each structure and its corresponding data values. This tabulation is given for data set III in Table II. For the remainder of the data sets, only the resultant statistical parameters for each correlation are given (Table II). Complete structures for all the compounds employed in the correlations may be found, however, in the corresponding references cited along with each set of data in Table II.

The biological activities for the various systems all represent a constant equivalent response such as ED_{50}, LD_{50}, or isonarcotic concentration. Set I corresponds to (i) the inhibition of rat brain cortex respiration by some barbiturates and (ii) the fraction of barbiturates bound to 1% bovine serum albumin, set II to the barbiturate inhibition of *Arbacia* egg cell division, set III to the toxic action of substituted phenols on *Micrococcus pyogenes* var. *aureus*, set IV to the toxic action of substituted phenols on *Salmonella typhosa*, set V to the activity of penicillins on *Staphylococcus aureus*, and set VI to the isonarcotic concentrations (tadpoles) of some esters, ketones, alcohols, and ethers.

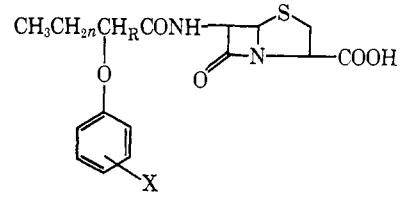
The molar attraction constants (F) were calculated either for the entire molecule (sets III, IV, VI) or only for that part of the molecule where the chemical structure varied (sets I, II, V). The use of partial instead of total molar attraction constants merely displaces a plot of biological activity *vs.* F along the abscissa and does not affect its shape. Partial values were used in those cases where F was not known for some functional group comprising the invariant part of the chemical structure.

Although eq 4 predicts that a plot of activity *vs.* F for compounds in a series will exhibit a maximum (if biological activity is directly related to ΔG_i) it does not necessarily tell one what its shape will be away from the maximum. Such plots, however, suggested that the biological data was linearly related to F for each set and a least-squares fit of each set of data to a first-order equation was performed. The statistical parameters for each set appear in Table II where r is the correlation coefficient, s is the standard deviation, n is the number of data points, and r^* is the correlation coefficient when the same biological data was correlated to π[11] in a comparable manner.

A relationship between the sum of molar attraction constants and observed biological activity appears to exist for each of the data sets examined. In those cases where a comparison can be made and r^* is given, the results suggest that F correlates with the biological data at least as well as π and in some cases slightly better than π. For the data corresponding to the toxic action of phenols on bacteria (sets III and IV), a correlation exists separately for each group of phenols which have substitution at the same ring position and the statistical parameter for each group is given separately. This result may be due to (1) a significant dependence of molar volume on the ring position of the substituted group, (2) the role of the relative partition coefficient, (3) steric effects, or (4) the fact that the molar attraction constant does not evaluate the significance of the relative location of a functional group within an organic molecule.[12] Corrections for the values of F based on the relative position of the substituted group could be made on an empirical basis for each biological system considered. On the other hand, it was found that for the data of sets III and IV a smooth curve exhibiting a maximum was obtained when the observed biological

TABLE I

EFFECTIVE CONCENTRATION FOR THE ACTIVITY OF PENICILLIN DERIVATIVES ON *Staphylococcus aureus* IN MICE



Function	F	Log (1/c) Obsd	Log (1/c) Calcd	Δ log (1/c)
H	1061	5.86	5.75	0.11
4-Cl	1239	5.79	5.43	0.36
4-OCH_3	1253	5.69	5.40	0.29
α-Et (n = 1)	1194	5.54	5.51	0.03
4-NO_2	1256	5.53	5.40	0.13
2-Cl	1239	5.40	5.43	0.03
2,5-Cl_2	1417	5.24	5.11	0.13
α-Pr (n = 2)	1327	5.03	5.27	0.24
3,5-$(CH_3)_2$	1305	5.03	5.31	0.28
α-Bu (n = 3)	1460	5.01	5.03	0.02
2,4-Cl_2	1417	4.97	5.11	0.14
2,4-Br_2	1557	4.87	4.86	0.01
2,3,6-Cl_3	1595	4.72	4.79	0.07
4-Cyclohexyl	1762	4.70	4.49	0.21
4-*t*-Bu	1518	4.67	4.93	0.26
3,4,5-$(CH_3)_3$	1427	4.65	5.09	0.44
4-*t*-Amyl, α-Et	1784	4.57	4.45	0.12
Cl_5	1951	4.25	4.15	0.10

(11) π is the parameter defined by Hansch as the logarithm of the ratio of two partition coefficients.

(12) It should be pointed out here that it appears that some functional groups take on different values for F when they are attached to an aromatic nucleus.

TABLE II

Data set	Equations	r	r^2	S	n	r^2	Comments	Ref
I(i)	Log $(1/c) = 3.58 \times 10^{-3} F - 0.895$	0.955	0.912	0.181	10	0.956		a
(ii)	F B [d] $= 9.1 \times 10^{-4} F - 0.620$	0.867	0.752	0.074	12			b
II	Log $(1/c) = 2.79 \times 10^{-3} F - 0.446$	0.912	0.831	0.236	19	0.960		2b
III(i)	Log $PC' = 3.4 \times 10^{-3} F - 4.36$	0.998	0.996	0.061	9		4-Alkyls	1b
(ii)	Log $PC' = 3.0 \times 10^{-3} F - 4.30$	0.992	0.985	0.122	10		3-Alkoxys	1b
(iii)	Log $PC' = 3.2 \times 10^{-3} F - 4.72$	0.999	0.998	0.026	9		4-Alkoxys	1b
(iv)	Log $PC' = 4.0 \times 10^{-4} \pi F - 0.294$	0.978	0.956	0.206	34	0.925	Complete set	1b
	Log $PC' = 2.5 \times 10^{-8} (\pi F)^2 + 5.9 \times 10^{-4} \pi F + 0.112$	0.995	0.989	0.106	35		Complete set	1b
IV(i)	Log $PC' = 3.45 \times 10^{-3} F - 4.44$	0.998	0.987	0.037	5		4-Alkyls; $F < 1850$	1b
(ii)	Log $PC' = 2.96 \times 10^{-3} F - 4.14$	0.995	0.990	0.057	6		3-Alkoxys; $F < 2025$	1b
(iii)	Log $PC' = 3.12 \times 10^{-3} F - 4.60$	0.991	0.982	0.081	6		4-Alkoxys; $F < 2025$	1b
(iv)	Log $PC' = 4.4 \times 10^{-4} \pi F + 0.245$	0.980	0.960	0.146	26	0.919	Complete set; $F < 4300$	1b
	Log $PC' = 2.1 \times 10^{-4} \pi F + 2.96$	0.973	0.946	0.077	7		Complete set; $F < 4300$	1b
V	Log $(1/c) = 1.78 \times 10^{-3} F + 7.64$	0.892	0.795	0.204	18	0.904		2a
VI	Log $(1/c) = 3.03 \times 10^{-3} F - 1.10$	0.968	0.937	0.212	27	0.965		c

[a] F. Fuhrman and J. Field, *J. Pharmacol. Exptl. Therap.*, **77**, 392 (1943). [b] L. Goldbaum and P. Smith, *ibid.*, **111**, 199 (1954). [c] J. Iwasa, T. Fujita, and C. Hansch, *J. Med. Chem.*, **8**, 150 (1965). [d] Fraction bound.

activity was plotted *vs.* the product of π and F.[13] The values of π employed in sets III and IV were taken from ref 2 and the resultant correlations are included in Table II.

For the systems investigated and in light of the assumptions that were required, the correlations appear to be quite reasonable considering the limited accuracy of the biological data. In the treatment of some of the same data by ρ–σ–π analysis, Hansch has concluded that the relative partition coefficient (π) accounted for most of the differences in biological potency.[1b,2b,c] Thus, it could be concluded that the molar attraction constant is actually only a measure of the relative partition coefficient. This was found not to be the case in the correlation of the structures of a series of related steroids with their relative antiinflammatory potencies where F gave good correlations but π did not.[14] The results in this paper, however, do suggest that F is related to π. Such a relationship is conceivable since partition phenomena as well as physical interactions are dependent on chemical structure.

The proposed method of correlation unquestionably has limitations with regard to its general applicability, and the theoretical basis of this treatment requires perhaps more than the usual number of assumptions. In addition, steric factors involved in drug interactions were neglected while it is likely that changes in molecular size, shape, and rigidity do indeed affect ΔS_i. However, it should be stated that for compounds of a series, the magnitude of these effects are probably approximately equal. The limitation that the enthalpy term as described by eq 1 is directly applicable only to systems where this quantity is positive is perhaps a serious one when coupled with the fact that the sign of ΔH_i is usually not known for most drug interactions. If the drug interaction is of a specific nature, whereby ΔH_i takes on negative values, it is still possible that a correlation between biological activity and molar attraction constants exists. The relationship to thermodynamic quantities would then take a different form than presented here.

In order to broaden the applicability of the proposed method one needs to obtain an expression for the thermodynamic enthalpy or free energy which is related to a molar attraction function as well as a second parameter which allows for those systems where the dispersion forces (f_d) do not follow the geometric mean rule and/or where other attractive forces are significant. Such an expression will permit ΔH_i to take on positive or negative values. In treating F and its apparent additivity it is realized that the validity of eq 2 is dependent on the assumption that the distribution of the dispersion forces follows the geometric mean rule. If this assumption is not a valid one in certain instances, it may be necessary to define a function, $g(f_d)$, for which the distribution of attractive forces would more closely approximate the true situation. An expression for ΔH which would account for specific interaction may take the form

$$\Delta H_i = g(f_d) - g(f_s) \qquad (10)$$

where f_s is an expression containing a parameter which describes the relative magnitude of specific interactions. For the case where ΔH_i is positive and only dispersion forces are operative, $g(f_s)$ would equal 0 and eq 9 would take the form of eq 2.

(13) The criticism offered by one of the reviewers regarding the empirical nature of the product πF is well taken. Its apparent correlation, however, may be rationalized intuitively by considering that the factor π expresses that fraction of the apparent concentration which is made available for interaction and may serve as a corrective factor in assessing the extent of interaction.

(14) Unpublished data.

Acknowledgment.—The author wishes to express thanks to Sheldon Kugler of the Department of Biostatistics, Syntex Research, for his help in the statistical analysis.