Theory and Application of Molecular Potential Energy Fields in Molecular Shape Analysis: A Quantitative Structure–Activity Relationship Study of 2,4-Diamino-5-benzylpyrimidines as Dihydrofolate Reductase Inhibitors

A. J. Hopfinger

Department of Drug Design, Searle Research and Development, Skokie, Illinois 60077. *Received October 25, 1982*

A general formalism, based upon molecular mechanics pairwise potential functions, has been developed to compute the molecular potential energy fields inherent to a given molecule in a given conformation. Molecular descriptors are derived from the potential energy fields, which can be used in QSAR studies based upon molecular shape analysis. These descriptors have been computed for a set of 2,4-diamino-5-benzylpyrimidines that are dihydrofolate reductase (DHFR) inhibitors. A QSAR is derived in which DHFR inhibition activity can be explained in terms of molecular shape, as represented by differences in molecular potential energy fields between pairs of superimposed molecules, and the sum of the π constants of substituents on the 3- and 4-position of the benzyl ring. This QSAR is superior to one developed earlier (Hopfinger, A. J. *J. Med. Chem.* **1981**, *24*, 818) in which molecular shape is described by common overlap steric volume. Ancillary information defining the "active" conformation and electrostatic nature of the binding site are realized in the construction of the QSAR.

In an earlier paper[1] a quantitative structure–activity relationship (QSAR) to describe the bovine liver dihydrofolate reductase (DHFR) inhibition activity of a set of substituted 5-benzyl-2,4-diaminopyrimidines (I) was developed by molecular shape analysis (MSA) (eq 1). In

$$\log (1/C) = -21.31V_0 + 2.39V_0^2 + 0.44(\pi_3 + \pi_4) + 52.23 \quad (1)$$
$$n = 23,\ r = 0.931,\ s = 0.137$$

eq 1, V_0 is the common overlap steric volume between each molecule in the data base and the X = *cis*-4-$NHCOCH_3$ compound that serves as the shape reference standard.[2] The values of V_0 are based upon identical atomic superposition of the respective pyrimidine rings of pairs of compounds. V_0 is interpreted as a quantitative measure of relative shape similarity. The term $\pi_3 + \pi_4$ represents the sum of the hydrophobic constants of the 3- and 4-position substituents with the fragment values given by Blaney et al.,[3] who constructed the parent structure–activity data base. C is the molar concentration of inhibitor that produces 50% inhibition.

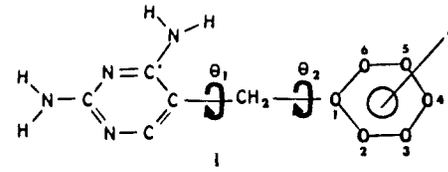
Blaney et al. also carried out a QSAR analysis of the 5-benzyl-2,4-diaminopyrimidine data base using Hansch analysis.[3] They formulated a correlation equation (eq 2)

$$\log (1/C) = 0.622\pi_3 + 0.322\sum\sigma + 4.99 \quad (2)$$
$$n = 23,\ r = 0.931,\ s = 0.146$$

to explain the DHFR inhibition activity. In this equation, $\sum\sigma$ represents the summed electronic effect of 3-, 4-, and 5-substituents on position 1. Equations 1 and 2 are of identical quality. However, $\sum\sigma$ is a somewhat surprising feature to find related to activity, especially since it does not appear in any of the other QSARs for DHFR inhibition developed by Hansch and co-workers for the triazines[4,5] and quinazolines.[6] Moreover, each of the QSARs developed by Hansch and co-workers for DHFR inhibition contain relatively diverse descriptors. Conversely, equivalent QSARs, in terms of correlation descriptors, were developed using MSA for the triazines[7,8] and quinazolines.[9]

In fact, the analysis of the benzylpyrimidines was undertaken to see if this consistency in correlation descriptors could be extended. Equation 1 was the fourth QSAR developed for DHFR inhibitors in which activity could be explained in terms of shape, as measured by common overlap steric volume, and substituent lipophilicity.

This foundation for quantitatively characterizing molecular shape, as well as successfully using resultant measures of shape similarity as activity correlates, has prompted further investigation of molecular shape. The results of constructing a new set of molecular shape descriptors, derived from the potential energy field of a molecule, and using or evaluating these descriptors in a MSA-based QSAR mode is presented here. The set of substituted 5-benzyl-2,4-diaminopyrimidines (I) used to develop eq 1 and 2 is the test data base selected for this investigation.



I

One final point to note is that all the past MSA–QSAR studies of DHFR inhibition,[1,7-9] as well as that reported here, have been carried out *without* using the geometric information available from the crystal structures of DHFR and DHFR–inhibitor binary complexes.[10-13] The goal in the development of MSA is to have a set of procedures that can be used to quantitatively predict (1) "active" molecular conformations and, consequently, (2) biological activities without independent knowledge of the geometric properties of the receptor. The crystal structures of binary

(1) Hopfinger, A. J. *J. Med. Chem.* **1981**, *24*, 818.
(2) Hopfinger, A. J.; Potenzone, Jr., R., *Mol. Pharmacol.* **1982**, *21*, 187.
(3) Blaney, J. M.; Dietrich, S. W.; Reynols, M. A.; Hansch, C. *J. Med. Chem.* **1979**, *22*, 614.
(4) Silipo, C.; Hansch, C. *J. Am. Chem. Soc.* **1975**, *97*, 6849.
(5) Dietrich, S. W.; Smith, R. N.; Fukunaga, J. Y.; Olney, M.; Hansch, C. *Arch. Biochem. Biophys.* **1979**, *194*, 600.
(6) Fukunaga, J. Y.; Hansch, C.; Steller, E. E. *J. Med. Chem.* **1976**, *19*, 605.
(7) Hopfinger, A. J. *J. Am. Chem. Soc.* **1980**, *102*, 7196.
(8) Hopfinger, A. J. *Arch. Biochem. Biophys.* **1981**, *206*, 153.
(9) Battershell, C.; Malhotra, D.; Hopfinger, A. J. *J. Med. Chem.* **1981**, *24*, 812.
(10) Matthews, D. A.; Alden, R. A.; Bolin, J. T.; Freer, S. T.; Hamlin, R.; Xuong, N.; Kraut, J.; Poe, M.; Williams, M.; Hoogsteen, K. *Science* **1977**, *197*, 452.
(11) Matthews, D. A.; Alden, R. A.; Bolin, J. T.; Filman, D. A.; Freer, S. T.; Hamlin, R.; Hol, W. G. J.; Kisliuk, R. L.; Pastore, E. J.; Plante, L. T.; Xuong, N.; Kraut, J. *J. Biol. Chem.* **1978**, *253*, 6946.
(12) Matthews, D. A.; Alden, R. A.; Freer, S. T.; Xuong, N.; Kraut, J. *J. Biol. Chem.* **1979**, *254*, 4144.
(13) Kuyper, L. F.; Roth, B.; Baccanari, D. P.; Ferone, R.; Beddel, C. R.; Champness, J. N.; Stammers, D. K.; Cann, J. G.; Norrington, F. E. A.; Baker, D. J.; Goodford, P. J. *J. Med. Chem.* **1982**, *25*, 1122.

0022-2623/83/1826-0990$01.50/0

Table I. The QSAR Data Base. Log $[\Delta P_{\alpha,\beta}(R, 2)^{1/2}]$ for CH_3, O^-, and H^+ Potential Field Test Probes

no.	substituent	obsd log (1/C)	log $[\Delta P_{\alpha,\beta}(R, 2)^{1/2}]$ CH_3	O^-	H^+	V_0,[a] Å^3	$\pi_3 + \pi_4$[b]	pred eq 9 log (1/C)	eq 9 Δ log (1/C)	pred eq 1 log (1/C)	eq 1 Δ log (1/C)
1	3,4-$(OH)_2$	4.30	−0.69	4.57	4.57	421.6	−1.24	4.30	0.00	4.36	−0.06
2	4-NH_2	4.57	−0.78	3.12	3.13	414.0	−0.86	4.59	−0.02	4.63	−0.06
3	4-$N(CH_3)_2$	4.76	−0.65	3.46	3.59	446.5	0.29	4.84	−0.08	4.90	−0.14
4	4-CH_3	4.80	−0.76	3.76	3.77	433.8	0.56	4.90	−0.10	5.05	−0.25
5	4-OCH_3	4.92	−0.87	2.91	2.91	425.3	−0.02	5.03	−0.11	4.86	0.06
6	4-OCF_3	4.99	−0.92	2.75	2.75	426.2	0.34	5.22	−0.23	5.01	−0.02
7	3-OCH_3	5.02	−0.44	2.86	2.86	410.1	−0.02	5.07	−0.05	5.06	−0.04
8	4-NO_2	5.02	−0.94	2.95	2.95	416.7	−0.09	4.98	0.04	4.93	0.09
9	4-$NHCOCH_3$	5.09	[c]			486.9	−0.35			5.03	0.06
10	4-Cl	5.10	−0.76	2.85	2.87	437.1	0.71	5.29	−0.19	5.10	0.00
11	3,4,5-$(OCH_3)_3$	5.10	−0.58	3.06	3.13	405.6	0.04	4.93	0.17	5.17	0.07
12	3,4-$(OCH_3)_2$	5.15	−0.49	2.90	2.92	405.6	0.04	5.05	0.10	5.17	0.02
13	3-NO_2, 4-$NHCOCH_3$	5.16	−0.74	2.78	2.91	414.9	−0.38	5.02	0.14	4.83	0.33
14	4-Br	5.17	−0.75	3.75	3.72	436.8	0.86	5.03	0.14	5.17	0.00
15	4-F	5.18	−0.51	2.85	2.86	417.3	0.14	5.13	0.05	5.02	0.16
16	H	5.19	−0.48	2.68	2.71	395.7	0.00	5.18	0.01	5.36	−0.17
17	3-CH_3	5.22	−0.63	2.84	2.82	406.5	0.56	5.31	−0.09	5.38	−0.16
18	3-F	5.33	−0.59	2.50	2.50	401.6	0.14	5.41	−0.08	5.29	0.04
19	3-Cl	5.47	−0.60	2.77	2.78	402.4	0.71	5.40	0.07	5.53	−0.06
20	3-CF_3	5.53	−0.61	2.81	2.82	408.3	0.88	5.43	0.10	5.49	0.04
21	3-Br	5.54	−0.79	2.70	2.77	401.5	0.86	5.46	0.08	5.62	−0.08
22	3-CF_3, 4-OCH_3	5.79	−0.87	2.39	2.39	409.6	0.86	5.78	0.01	5.46	0.33
23	3-$OCH_2C_6H_5$	6.10	−0.71	2.60	2.69	402.7	1.89	6.08	0.02	6.05	0.05

[a] V_0 is the common overlap steric volume shape parameter. [b] π_3 and π_4 are the hydrophobic fragments of substituents on the 3 and 4 sites. [c] 4-$NHCOCH_3$ is the reference compound in the MSA analyses.

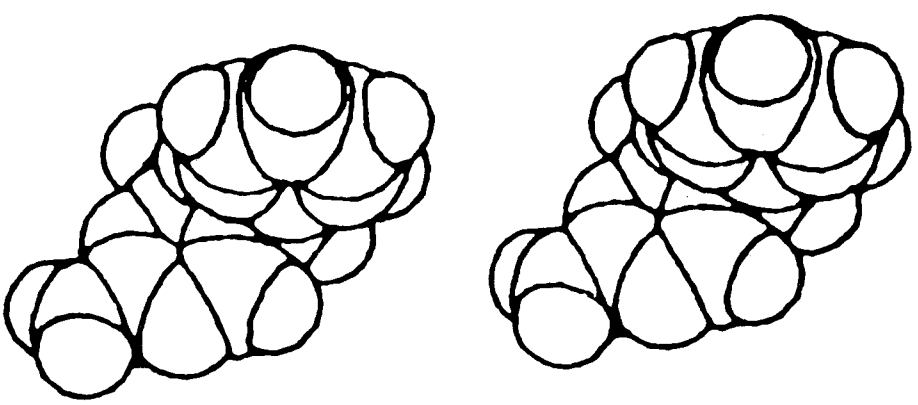


Figure 1. Space-filling stereo representation of compound **16** (X = H) (Table I) in the postualted active conformation.

DHFR–inhibitor complexes have been examined *after the fact* to evaluate the accuracy of MSA-predicted "active" conformations. This is discussed for the benzylpyrimidines under Methods.

Methods

Data Base. Blaney et al. have constructed a data base containing 23 5-(substituted-benzyl)-2,4-diaminopyrimidines whose inhibitory effect on bovine liver DHFR has been measured.[3] The activity is expressed as log (1/C), where C is the molar concentration of inhibitor that produces 50% inhibition. The compounds and measured inhibition activities are given as part of Table I.

Conformational and Molecular Shape Analyses. The "active" conformation, with respect to θ_1 and θ_2 (see I), determined in the MSA study employing common steric overlap volume as the shape descriptor[1] was adopted in this investigation. The substituent conformations were also those used in constructing eq 1. In other words, the conformations hypothesized as the inhibitor-bound molecular geometries on the basis of eq 1 have also been selected in this MSA study, which focuses upon molecular potential fields.

There is support for making this critical assumption. Recently, Kuyper et al.[13] have reported the crystal structures of binary complexes of *E. coli* DHFR and trimethoprim (compound 11 of Table I) and some of its derivatives. The bound conformation of trimethoprim in the crystal differs from our predicted active conformer[1] by $\theta_1 \approx 7°$, $\theta_2 \approx 12°$ (see Figure 1). This is quite good agreement, especially when one considers that trimethoprim is conformationally quite flexible with respect to θ_1 and θ_2. The conformational freedom can be descerned from the conformational energy map for trimethoprim shown in Figure 2. However, the inhibitory activity of DHFR can be species specific,[8,14,15] which, in turn, could reflect species-dependent binding geometries. Nevertheless, we have considered the conformational agreement between theory and experiment of sufficient merit so as to retain the postulated active conformers reported in ref 1.

Molecular Potential Energy Fields as Shape Descriptors. It is conceptually clear that every molecule generates a potential field about itself that is a function of both its chemical and geometric structure. Further, it is generally believed that recognition and binding of an inhibitor to an enzyme are functions of the complementary nature of the potential fields of the two molecules, e.g., the intermolecular energy. However, quantitative modeling of these working concepts poses a very substantial problem. The formalism developed here represents an attempt to model molecular potential energy fields and to derive therefrom a second-generation set of MSA descriptors that can be used in place of common overlap steric volume in QSAR applications. The following three specific issues had to be addressed in constructing the formalism.

(14) Ho, Y. K.; Hakala, M. T.; Zakrzewski, S. F. *Cancer Res.* **1972**, *32*, 1023.
(15) Greco, W. R.; Hakala, M. T. *Mol. Pharmacol.* **1980**, *18*, 521.

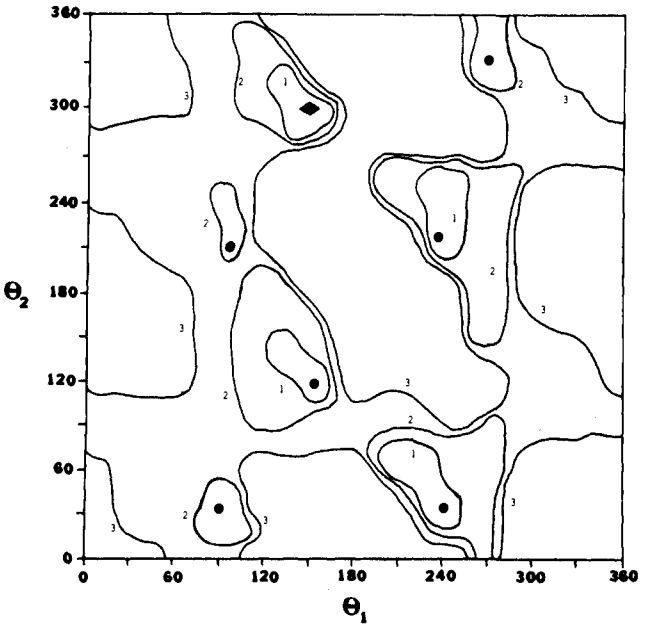


Figure 2. Conformational energy map of trimethoprim. θ_1 and θ_2 are defined in I. Energy contours are in kilocalories per mole above global minimum denoted by ♦. Relative energy minima are denoted by ●.

(1) Choice of Potential Energy Field Probe. There is no definitive answer, other than using the receptor, for the "best" choice of field probe. Three field probes have been adopted in the work reported here. H^+ and O^- are used to monitor positve and negative charge response characteristics, respectively, of the potential energy field. A CH_3 unit is used to characterize the dispersion component of the field. Additions and deletions of field probes can be made on an as need or as interest basis.

(2) Computational Feasibility. Quantum chemists are increasingly using molecular electrostatic potential theory as a means of probing both the physical and chemical reactivity of a molecular species.[16] Weinstein[17] pioneered in employing the molecular electrostatic potential as a means of explaining common biological activity among compounds of limited structural homology. Thus, the molecular electrostatic potentials realized through molecular orbital theory represent one avenue of computing MSA descriptors based upon molecular potential energy fields. Unfortunately, the computation time required to generate all of the field potentials for data bases like that in Table I using multiple field probes is prohibitively high. This has prompted the employment of molecular mechanics (MM) potentials as a means of estimating the molecular potential energy fields. Hence, the molecular potential energy field, $P_u(R, \theta, \phi)$, at any point (R, θ, ϕ) for molecule u is given by eq 3. In eq 3 $a(T)_i$ and $b(T)_i$ are the attractive and

$$P_u(R, \theta, \phi) = \sum_{i=1}^{n}\left[\frac{a(T)_i}{r_i^6} + \frac{b(T)_i}{r_i^{12}} + \frac{Q_iQ(T)}{\epsilon(r_i)r_i}\right] \quad (3)$$

repulsive potential energy coefficients, respectively, of atom i of molecule u interacting with the test probe, T, treated as a single force center. The $a(T)_i$ and $b(T)_i$ can be taken from any of a number of different reported potential energy sets. Those of Hopfinger[18] were used in this work. Q_i and $Q(T)$ are, respectively, the charge densities of the ith atom and the test probe. $\epsilon(r_i)$ is the dielectric term. In the current study, $\epsilon(r_i) = 3.5$. A discussion of the nature and role of $\epsilon(r_i)$ is given below. Lastly, n is the number of atoms in u, and r_i is the distance between atom i and the test probe. The coordinate system is spherical and, by convention, located at the center of mass (of the shape reference standard compound when pairs of fields are being compared).

(16) See "Chemical Applications of Atomic and Molecular Electrostatic Potentials"; Politzer, P.; Truhlar, D. G., Eds.; Plenum Press: New York, 1981.
(17) See ref 16, p 309.

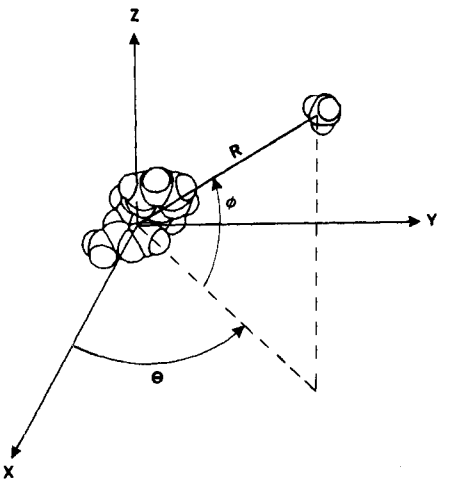


Figure 3. Intermolecular coordinates R, θ, and ϕ defining field probe relative to compound **16**.

Figure 3 illustrates the geometry used to compute $P_u(R, \theta, \phi)$ when u is compound **16** (Table I), in its active conformation, and the test probe is CH_3.

(3) Construction of MSA Descriptors Based upon the Potential Energy Field. A variety of different descriptors can be generated from the molecular potential energy field. Six descriptors hve been constructed and considered in this study.

(i) A graphical display of the isoenergy contour curves of the molecular potential energy field in space, which is superimposed upon a molecular model of the corresponding compound. This form is most useful when represented by real-time computer graphics. A corresponding table of $P_u(R, \theta\, \phi)$ values as a function of (R, θ, ϕ) can also be considered. These representations of molecular potential energy field are difficult to use in a quantitative mode.

(ii) Extrema in $P_u(R, \theta, \phi)$, e.g., eq 4, where $X_i = R, \theta, \phi$. The

$$\frac{\partial}{\partial X_i}[P_u(R, \theta, \phi)] = 0 \quad (4)$$

location of maxima and minima for the molecular potential energy field in space provides direct information regarding the least, and most, interactive sites for molecule u interacting with a particular test probe. In so far as a test probe represents a critical receptor binding group, the spatial and energetic properties of these extrema might be expected to correlate with binding (activity).

(iii) The radial distribution function in the molecular potential energy field (eq 5). This function provides a profile of the total

$$P_u(R)dR = \int_\theta\int_\phi P_u(R, \theta, \phi)R^2 \sin\theta\, d\theta d\phi dR \quad (5)$$

angular potential energy field at any distance from a preselected coordinate frame center, normally the center of mass of u. The numerical integrations carried out in this investigation have been done at a 10° resolution in θ and ϕ.

(iv) The total molecular potential energy field (eq 6). In eq

$$P_u = \int_R P_u(R)dR \quad (6)$$

6, P_u provides a simple scalar measure of the overall molecular potential energy field of u with the test probe. A comparison of the P_u and $P_u(R)$ for various test probes can provide information regarding how different types of receptor interactions and/or groups influence the binding propensity of u. For example, a comparison of P_u for H^+ and O^- probes could be used to discriminate between positive and negative, respectively, charge receptor sites. The resolution in R, ΔR, for the numerical integration is 0.2 Å.

(v) Difference radial distribution function in the molecular potential energy field (eq 7). The value of m is normally 2 but

$$\Delta P_{\alpha,\beta}(R, m)dR = \int_\theta\int_\phi[P_\alpha(R, \theta, \phi) - P_\beta(R, \theta, \phi)]^m R^2 \sin\theta\, d\theta d\phi dR \quad (7)$$

could be 1 depending upon the type of difference profile required.

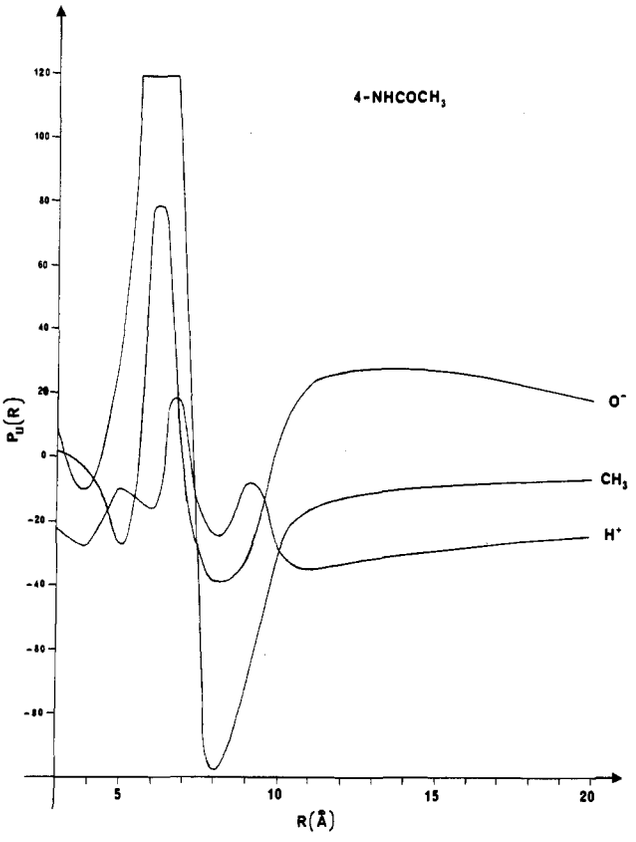


Figure 4. $P_u(R)$ vs. R for compound **9** (X = 4-NHCOCH$_3$) (Table I) using each of the field probes. $P_u(R)$ is in kilocalories per mole.

By convention, α is the reference molecule; i.e., the coordinate frame is referenced to α and, in a series of compounds, all molecules are compared to α.

(vi) The total difference in molecular potential energy fields (eq 8). This descriptor can be considered a generalized form of

$$\Delta P_{\alpha,\beta} = \int_R \Delta P_{\alpha,\beta}(R, m)dR \qquad (8)$$

the common overlap steric volume, V_0. $\Delta P_{\alpha,b}$ is a measure of similarity or difference in total spatial potential energy fields, which is an energetic representation of molecular shape. The ability to use different field probes makes it possible to compute different $\Delta P_{\alpha,\beta}$ for a given pair of α and β. In this way it is possible to evaluate how different substituents in α and/or β make these compounds similar or different in total field potential, i.e., shape.

Results

The MSA desciptors derived from molecular potential energy fields were determined for each of the 23 5-(substituted-benzyl)-2,4-diaminopyrimidines reported in Table I. All calculations were carried out on a TRS-80 Model I microcomputer. Examination of the set of $P_u(R, \theta, \phi)$ and associated extrema did not suggest any correlative applications to measured inhibition activities.

The descriptors $P_u(R)$ and P_u provide interesting information regarding the characteristic, and intrinsic, molecular potential energy field behavior of each compound. Figure 4 illustrates $P_u(R)$ vs. R for X = *cis*-4-NHCOCH$_3$ using each of the three test probes O$^-$, H$^+$, and CH$_3$. At short distances, each of the $P_u(R)$ exhibit erratic behavior, reflecting repulsive steric intermolecular interactions. In the range of R = 8 Å, both the CH$_3$ and O$^-$ $P_u(R)$ exhibit potential energy minima, reflecting net stable intermolecular binding. The H$^+$ probe appears to be too "small" to exhibit a definitive net binding potential energy minimum. The behavior of $P_u(R)$ at large R is characteristic of the nature of the test probe employed. The CH$_3$ probe generates a $P_u(R)$ that most rapidly converges to zero potential as R increases. This behavior is indicative of the short-range behavior of dispersive forces. The $P_u(R)$ of both O$^-$ and H$^+$ converge much more slowly than that of CH$_3$ and are essentially mirror-image functions of one another at large R. This slow convergence may be artificial to some extent and due to the choice of a low value, fixed "molecular dielectric constant" of 3.5. In an aqueous medium the actual effective dielectric factor may enhance the convergence of these functions. Nevertheless, both the O$^-$ and H$^+$ $P_u(R)$ can be expected to converge less rapidly than that of the CH$_3$ probe. The dielectric contribution can also be expected to exert an equal effect, at a common distance, on the $P_u(R)$ of O$^-$ and H$^+$. Thus, the compound X = 4-NHCOCH$_3$ will always be repelled by a net negative charge source and attracted by a positive charge site. These repulsions and attractions, due to electrostatic interactions, will be realized at larger interaction distances than corresponding dispersion-based attractions.

Each compound in the data base was selected as the reference molecule in the determination of the pairwise field-difference descriptors, $\Delta P_{\alpha,\beta}(R, m)$ and $\Delta P_{\alpha,\beta}$. Thus, like V_0 (the common overlap steric volume), 23 sets of $\Delta P_{\alpha,\beta}(R, m)$ and $\Delta P_{\alpha,\beta}$ values are available as possible correlation features. Figure 5 illustrates the log $[\Delta P_{\alpha,\beta}(R, 2)^{1/2}]$ vs. R for five compounds from Table I using (a) CH$_3$, (b) O$^-$, and (c) H$^+$. The choice of using the logarithms of the square root of $\Delta P_{\alpha,\beta}(R, 2)$ is for numerical convenience. The difference functions generated via the CH$_3$ probe converge rapidly relative to those constructed using O$^-$ and H$^+$. All of the difference functions, for each of the three probes, exhibit erratic behavior at R < 8 Å. This is again indicative of steric repulsive interactions. In order to eliminate the spurious contributions from these short-range interactions in the calculation of P_u and $\Delta P_{\alpha,\beta}$, the radial integration was begun at R = 8 Å. Unfortunately, the numerical integrations had to be carried out to 40 Å in order to realize reasonable convergence in the $P_u(R)$ and $\Delta P_{\alpha,\beta}(R, 2)$ for the O$^-$ and H$^+$ test probes. This necessitates large numbers of computations on potential energy functions whose forms may not be valid at such large distances. Nevertheless, $\Delta P_{\alpha,\beta}$ was found to be a successful replacement for V_0 in eq 1. It was also gratifying to find that the set of $\Delta P_{\alpha,\beta}$ generated with X = *cis*-4-NHCOCH$_3$ as the reference molecule yielded the most significant QSAR as was the case with V_0 based upon X = *cis*-4-NHcOCH$_3$ as the reference molecule.[1] The $\Delta P_{\alpha,\beta}$ for H$^+$ and O$^-$ yielded superior correlation equations to that obtained with CH$_3$ as the test probe. The correlation equation based upon H$^+$ was slightly better than that generated using O$^-$. Attempts to generate QSARs using the P_u as correlation descriptors were not successful.

The slow convergence of the $\Delta P_{\alpha,\beta}(R, 2)$ to zero, as a function of increasing R, poses a serious computational liability to the determination of the $\Delta P_{\alpha,\beta}$. However, the functional behavior of the $\Delta P_{\alpha,\beta}(R, 2)$ for O$^-$ and H$^+$ at $R \geq 8$ Å (see Figure 5b,c) indicate that the curves are, essentially, a set of parallel lines sloping to zero. Thus, the relative areas under the curves, the set of $\Delta P_{\alpha,\beta}$, are directly proportional to the relative *heights* of the curves at some common value of R. In so far as this observation can be generalized to other classes of compounds, the time-consuming calculation of the $\Delta P_{\alpha,\beta}$ can be replaced by a simple calculation of the set of $\Delta P_{\alpha,\beta}(C, 2)$. C is the common value of R (>8 Å in this case) selected.

The best correlation equation was found using C = 12 Å for the benzylpyrimidines. The corresponding log $[\Delta P_{\alpha,\beta}(12, 2)^{1/2}]$ for each of the test probes are reported as part of Table I. The reference compound is the X =

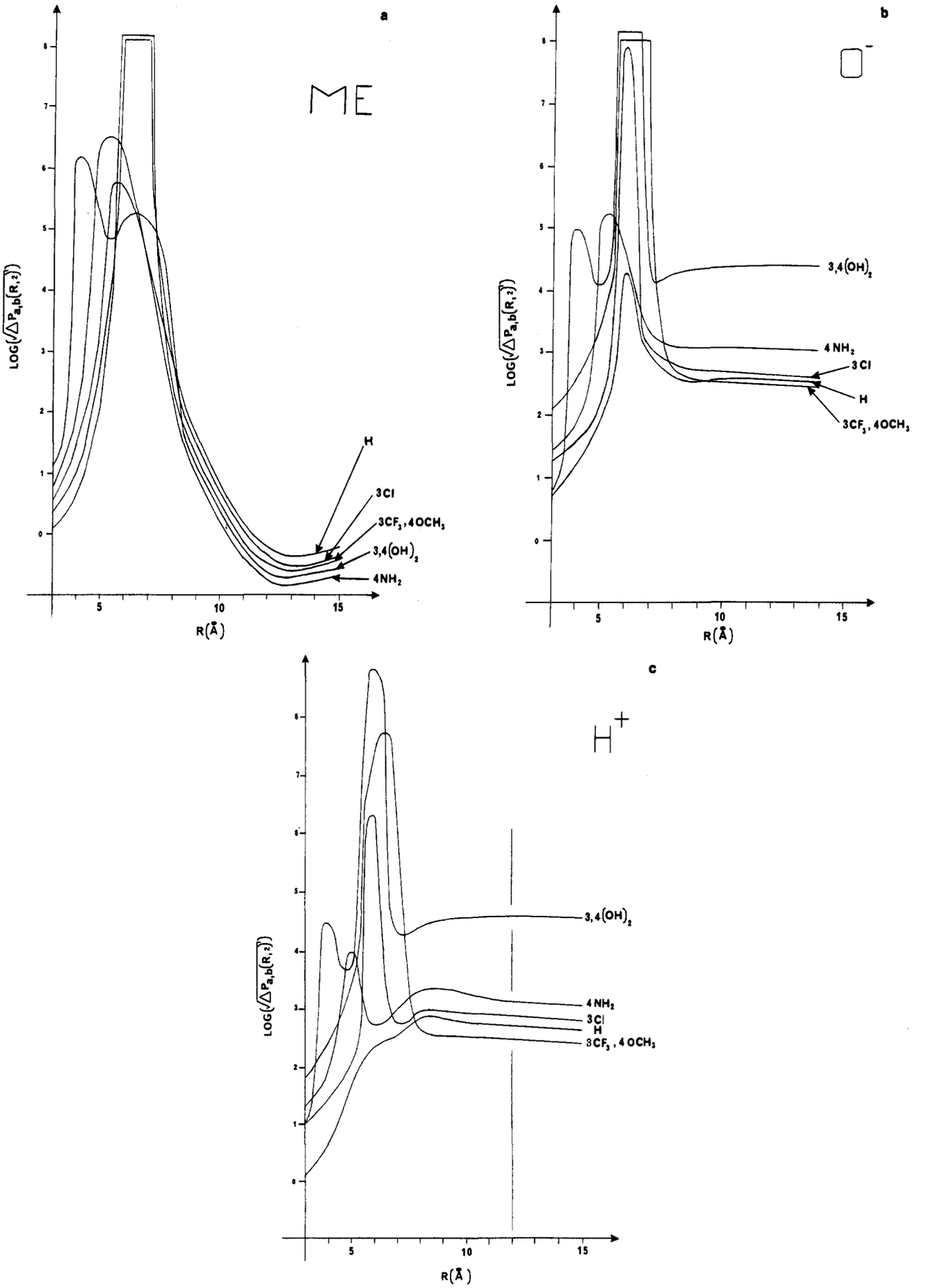


Figure 5. Log $[\Delta P_{\alpha,\beta}(R, 2)^{1/2}]$ vs. R for five compounds in Table I relative to compound **9**, X = 4-$NHCOCH_3$, used as the shape reference standard: (a) CH_3 field probe; (b) O^- field probe; (c) H^+ field probe.

Table II. Effect of Descriptor and Field Probe on the Significance of the Correlation Equations Analogous to Equation 1[a]

molecular field descriptor	r		
	CH_3	O^-	H^+
$P_u(12)$	0.805	0.783	0.785
$P_u(14)$	0.818	0.806	0.797
P_u	0.820	0.850	0.843
$\Delta P_{\alpha,\beta}(12, 1)$	0.815	0.860	0.835
$\Delta P_{\alpha,\beta}(12, 2)$	0.809	0.942	0.961
$\Delta P_{\alpha,\beta}(14, 1)$	0.822	0.851	0.840
$\Delta P_{\alpha,\beta}(14, 2)$	0.811	0.937	0.952
$\Delta P_{\alpha,\beta}$	0.792	0.929	0.963

cis-4-$NHCOCH_3$ molecule, since it yields the most significant QSAR as described above. The optimum correlation equation analogous to eq 1 using $F = \log [\Delta P_{\alpha,\beta}(12, 2)^{1/2})$ for the H^+ test probe is eq 9. Equation 9 is superior

$$\log (1/C) = -2.34\ (0.67)\ F + 0.29\ (0.09)\ F^2 + 0.37\ (0.05)\ (\pi_3 + \pi_4) + 9.39 \quad (9)$$

$$m = 22,\ r = 0.961,\ s = 0.105$$

to eq 1, as reflected by the increase in the correlation coefficient r and decrease in the standard deviation of fit, s. Table II reports the significance of the correlation equations, in terms of r, when various molecular potential energy field descriptors are used in place of V_0 in eq 1. The quality of the QSAR depends moderately on the choice of the reference compound. The range in r is 0.878 (X = 3-F) to 0.961 (X = 4-$NHCOH_3$). This variation is similar to that observed with V_0 for a set of triazine DHFR inhibitors.[7]

Discussion

The molecular potential energy fields about the benzylpyrimidines indicate some surprising relative properties about these congeners that might be of a general nature. First, the difference in field potential between two homologues can be quite large. For example, the total angular difference in the molecular potential energy fields between X = 3,4-$(OH)_2$ and X = H, with H^+ as a probe, at R = 12 Å (see Figure 5c) is about 100 kcal/mol! Of course, the dielectric effect of an aqueous medium could be expected to reduce this difference. With an effective molecular dielectric of 80, the difference drops to about 8 kcal/mol. This is still a significant amount of potential energy in terms of differentiating physical binding behavior between two similar molecules.

An analysis of Figure 4 suggests that long-range recognition of a binding site is first achieved through electrostatic interactions. The $P_u(R)$ for CH_3 converges to zero much more rapidly than those generated for the H^+ and O^- probes. Again the relative convergence rate is a function of molecular dielectric. Nevertheless, the $P_u(R)$ derived from H^+ and O^- will converge slower than that from CH_3 for any reasonable choice of molecular dielectric.

Further, Figure 4 indicates that a binding site possessing an effective negative charge will repel the benzylpyrimidine, while a site with a composite positive charge will be attracted. There is experimental evidence that supports a net positively charged binding site. Kuyper et al.[13] have observed that the DHFR binding affinity of benzylpyrimidines can be markedly enhanced by the addition of negatively charged meta side chains. More generally, the data in Figure 4 may be indicative that the *effective* net charge at a binding site controls the recognition process.

The $\Delta P_{\alpha,\beta}$ and $\Delta P_{\alpha,\beta}(12, 2)$ derived from H^+ yield, respectively, somewhat better correlations with activity than those derived with O^- and far superior correlations to those generated employing CH_3 as the test probe. This empirical finding may reflect the assertion in the preceding paragraph that a positive effective charge over the inhibitor binding site is critical to inhibition specificity. Further, the calculated variations in the potential energy fields as a function of field probe demonstrates that the mechanistic specificity of the intermolecular interaction can be explored via choice in field probe. In this case, the relative differences in electrostatic interactions can be judged most important in defining intermolecular specificity, as evidenced by the fact that only molecular potential energy field difference descriptors based upon H^+ and O^- yield significant QSARs (see Table II). Still, it must be kept in mind that all the molecular potential energy field based QSARs have been generated using the postulated active conformation.[1] Thus, it may be that steric and/or dispersion interactions are already intrinsically accounted for in the QSAR through the "proper" selection of molecular conformation.

The molecular potential energy fields generated by the empirical potential functions represented by eq 3 can only be applied to physical binding. Moreover, the energy fields are probably most representative to describing the initial, long-range recognition between a ligand and a biomacromolecule. If chemical reactions are involved in the intermolecular interaction, that is, irreversible binding, electrostatic potentials must be generated by molecular orbital methods.[16]

Irrespective of the source of the potential set, the role of molecular dielectric upon the intermolecular energetics remains ill-defined. More generally, the lack of understanding of how to treat local dielectric behavior in molecular structure calculations plagues much of theoretical chemistry. The consequences of inadequate dielectric modeling are most acute in systems involving charged species,[18] which, unfortunately, are typical of enzyme receptor sites. Given the current lack of understanding of molecular dielectric, it is probably most prudent to treat this factor as a scaling variable in molecular structure calculations. If the molecular dielectric constant is treated as a scaling variable, then consideration should also be given to varying the magnitude of the dielectric as a function of interaction distance. Some distance-dependent dielectric representations have been reported[18,19] and will be considered in further studies of molecular potential energy fields in this laboratory. The general consensus is that the molecular dielectric approaches that of the bulk solvent as the interaction distance increases. For water, which has a large ϵ ($\sim$80), such a distance-dependent dielectric model should lead to a more rapid convergence in the $\Delta P_{\alpha,\beta}$ for charged field probes.

The results of this investigation, even with the limitations imposed by the molecular dielectric, indicate that molecular potential energy field calculations can be used to generate MSA descriptors that may be used to develop QSARs superior to those based upon common overlap steric volume, V_0. Equation 9 is superior to eq 1 as a result of F being a more representative measure of shape differentiation than V_0. Still, the empirical nature of using simple molecular entities as molecular shape generating descriptors should be kept in mind. Within the QSAR framework, the identification of a mechanism of action associated with a correlation equation usually involves major untested assumptions. Thus the finding that $\Delta P_{\alpha,\beta}$, or its mathematical equivalent, $\Delta P_{\alpha,\beta}(C, 2)$, using an H^+

(18) Hopfinger, A. J. "Conformational Properties of Macromolecules"; Academic Press: New York, 1973.
(19) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640.

probe, is a superior replacement for V_0 in this study must be viewed with reserved mechanistic interpretation.

However, in conjunction with the QSAR given by eq 9 it has been possible to generate the following satellite information on the benzylpyrimidines, which can be used in a design mode: (1) the active conformation with respect to θ_1 and θ_2; (2) identification of preferred binding to a site having a net positive charge; and (3) realization that a H^+ test probe yields relative intermolecular energetics that strongly correlate with biological activity. The "trick" of replacing $\Delta P_{\alpha,\beta}$ with $\Delta P_{\alpha,\beta}(C, 2)$ is critical to the practical and reliable usage of MSA based upon molecular potential energy fields. The calculation of potential energy fields in the 30- to 40-Å range is quite time consuming and of limited physical meaning. The form of the potentials at such distances are unknown, and, again, dielectric effects are expected to alter (decrease) field strength. The $\Delta P_{\alpha,\beta}(C, 2)$ for C in the range just beyond intermolecular steric contacts, C = 8 to 15 Å for the benzylpyrimidines, may, in fact, be the descriptors reflecting actual binding. These distances correspond to direct engagement with the receptor site. In any event, evaluating the general applicability of $\Delta P_{\alpha,\beta}(C, 2)$ as the comparative shape descriptor in future studies is a high priority in our work, as is understanding what it measures.

Acknowledgment. The author acknowledges the helpful discussions with R. A. Pearlstein and D. Malhotra of Case Western Reserve University during the course of this study.

Registry No. 1, 71525-05-8; 2, 69945-50-2; 3, 69945-51-3; 4, 46726-70-9; 5, 20285-70-5; 6, 49561-94-6; 7, 59481-28-6; 8, 69945-52-4; 9, 69945-53-5; 10, 18588-43-7; 11, 738-70-5; 12, 5355-16-8; 13, 69945-54-6; 14, 69945-55-7; 15, 836-06-6; 16, 7319-45-1; 17, 69945-56-8; 18, 69945-57-9; 19, 69945-58-0; 20, 50823-94-4; 21, 69945-59-1; 22, 50823-96-6; 23, 69945-60-4; DHFR, 9002-03-3.

Combined Distance Geometry Analysis of Dihydrofolate Reductase Inhibition by Quinazolines and Triazines

Arup K. Ghose and Gordon M. Crippen*

Department of Chemistry, Texas A&M University, College Station, Texas 77843. *Received July 30, 1982*

Guided by the success of distance geometry in explaining the inhibition of dihydrofolate reductase by 68 quinazolines, we have made a combined analysis on the inhibition of rat liver dihydrofolate reductase by 33 triazines and 15 quinazolines. The model gave a fit having the correlation coefficient 0.892 and root mean square (rms) deviation 0.596 in log $(1/C_{50})$ units. The model was applied to predict the biological activity of 91 compounds. The predicted values showed an rms deviation of 0.907 and a correlation coefficient of 0.790. The present study suggested the synthesis of some triazines as possible potent dihydrofolate inhibitors. The site geometry was compared with the crystal structure of a triazine bound to chicken liver dihydrofolate reductase, and a good correlation has been found.

The medicinal importance of dihydrofolate reductase has stimulated a great deal of work in several fields, and a large number of molecules have been evaluated[1] as its inhibitors; the species specificity[2] of the enzyme inhibitors has been examined; crystallographers have not only studied the structures of the inhibitors[3] but also the structure of inhibitor–enzyme complexes.[4,5] Such extensive data have attracted quantitative structure–activity relationship (QSAR) workers. Hansch et al.[6-8] made an extensive analysis using their physicochemical parameter dependent QSAR. On the other hand, the three-dimensional structure directed QSAR has been investigated by Crippen et al.,[9-11] Hopfinger et al.[12,13] and Simon et al.[14]

Guided by the success[10,11] of the distance geometry approach with even as few as six site points to explain the inhibition of *Streptococcus faecium* dihydrofolate reductase by 68 quinazolines, we have undertaken in this work a combined QSAR analysis of the triazines and quinazolines. Although in our previous studies we took the inhibition data of 68 quinazolines against dihydrofolate reductase from *S. faecium*, in the present study we selected the dihydrofolate reductase from rat liver, since a large variety of inhibitors have been investigated for this mammalian enzyme. We took all 33 3′- and 4′-substituted triazines having the general structure I (shown in Table I), reported by Dietrich et al.[15] and Hansch et al.[16] In order to keep the data set to a reasonable size and hold

(1) B. R. Baker, *Ann. N.Y. Acad. Sci.*, 186, 214 (1971).
(2) B. R. Baker, "Design of Active Site Directed Irreversible Enzyme Inhibitors", Wiley, New York, 1967, pp 192–266.
(3) A. Camerman, H. W. Smith, and N. Camerman, *Acta Crystallogr., Sect. B*, 35, 2113 (1979).
(4) D. A. Matthews, R. A. Alden, J. T. Bolin, S. T. Freer, R. Hamlin, N. Xuong, J. Kraut, M. P. M. Williams, and K. Hoogsteen, *Science*, 197, 452 (1977).
(5) D. A. Matthews, R. A. Alden, S. T. Freer, N. Xuong, and J. Kraut, *J. Biol. Chem.*, 254, 4144 (1979).
(6) S. W. Dietrich, R. N. Smith, S. Brendler, and C. Hansch, *Arch. Biochem. Biophys.*, 194, 612 (1979).
(7) S. W. Dietrich, J. M. Blaney, M. A. Reynolds, P. Y. C. Jow, and C. Hansch, *J. Med. Chem.*, 23, 1205 (1980).
(8) C. Hansch, J. Y. Fukunaga, P. Y. C. Jow, and J. B. Hynes, *J. Med. Chem.*, 20, 96 (1977).
(9) G. M. Crippen, *J. Med. Chem.*, 22, 988 (1979).
(10) A. K. Ghose and G. M. Crippen, *J. Med. Chem.*, 25, 892 (1982).
(11) A. K. Ghose and G. M. Crippen, "Proceedings of the 4th European Symposium on Chemical Structure and Biological Activity: Quantitative Approaches", Elsevier, Amsterdam, 1982, in press.
(12) A. J. Hopfinger, *J. Am. Chem. Soc.*, 102, 7196 (1980).
(13) C. Battershell, D. Malhotra, and A. J. Hopfinger, *J. Med. Chem.*, 24, 812 (1981).
(14) Z. Simon, I. Badilescu, and T. Racovital, *J. Theor. Biol.*, 66, 485 (1977).
(15) S. W. Dietrich, R. N. Smith, J. Y. Fukunaga, M. Olney, and C. Hansch, *Arch. Biochem. Biophys.*, 194, 600 (1979).
(16) C. Hansch, S. W. Dietrich, and J. Y. Fukunaga, *J. Med. Chem.*, 24, 544 (1981).

0022-2623/83/1826-0996$01.50/0