between the total available surface area of particles of respectively 16 and 1000 Å in diameter, assuming spherical shapes for all particles as a rough approximation) and in catalysts surface "concentrations" (going from small to large particle) may introduce some kinetics effects, especially in cases where there are

short-lived intermediates.

A QSAR Investigation of Dihydrofolate Reductase Inhibition by Baker Triazines Based upon Molecular Shape Analysis

A. J. Hopfinger

Contribution from the Department of Macromolecular Science, Case Institute of Technology, Case Western Reserve University, Cleveland, Ohio 44106. Received March 3, 1980

Abstract: Three-dimensional molecular shape descriptors have been defined for a set of Baker triazines which are dihydrofolate reductase (DHFR) inhibitors. Two shape descriptors, and one physicochemical feature, the sum of the 3- and 4-substituent octanol/water fragment constants, are needed to explain a range in activity in the data base of 5.11 log $(1/C)$ units. C is the molar concentration of triazine for 50% in vitro inhibition. One of the molecular shape descriptors, S_0, is proportional to the common overlap steric volume between pairs of molecules. The other shape descriptor, D_4, is a measure of the length of the 4-substituent in a triazine molecule. The compound data base consists of 27 compounds selected to represent each possible conformational class of compounds within the set of Baker triazines. No compound in the data base had to be eliminated as an outlyer. The quantitative structure–activity relationships (QSARs) generated in this study are tested on seven known compounds and used to predict five active compounds. This study constitutes, perhaps, the clearest instance to date where the consideration of molecular shape/conformation leads to an improved quantitative description of drug potencies.

Introduction

Two major obstacles have impeded quantitatively employing conformational analysis in drug design. First there is the problem of being able to identify the stable conformers of flexible molecules. A significant breakthrough in this area has been the development of molecular mechanics.[1,2] The decisive practical advantage of doing conformational analysis using molecular mechanics methods, as opposed to molecular orbital theory,[3] was first shown by our laboratory for some phenethylamines.[4] The general utilization of molecular mechanics calculations for molecular structure determination has been realized in our software package, currently called CAMSEQ-II.[5] Technical progress to explore larger and more flexible molecules has come from a new generation of low-cost, high-speed "super-mini" computers.

The other obstacle to conformation-dependent, quantitative drug design is the characterization of molecular shape. At this point let a formal distinction be made between shape and conformation. Conformation is a component of shape in that conformation defines the location of atoms in space. The properties of these atoms, most notably their "sizes", represent an additional set of factors needed to specify molecular shape. A theory of molecular shape analysis is outlined in this paper.

Some structure–activity studies have employed conformational features. These include: (1) an interatomic distance within a molecule,[3,6,7] (2) a set of interatomic distances within a molecule,[3,8–10] (3) a set of atomic coordinates within a molecule,[4,11]

and (4) a set of critical intermolecular binding distances.[12] Simple shape descriptors, which include (1) molecular volume,[13] (2) molecular surface area,[14] and (3) spatial potential surfaces of a molecule with respect to a test species[15,16] have also been considered in establishing structure–activity relationships.

Each of these descriptors, except the last, only provides a *partial* representation of the overall shape of a molecule. Spatial potential surfaces do reflect overall molecular shape, but a means representing this information for quantitative comparison is not reported.

A general theory of comparing molecular shapes is being developed in our laboratory. Molecular shape descriptors derived from this theory have been applied to quantitatively explain the observed 50% in vitro inhibition activity of a dihydrofolate reductase enzyme assay by a set of Baker triazines.[17] The purpose of this paper is to report this study which represents a clear instance where a systematic consideration of molecular shape leads to a quantitative structure–activity relationship (QSAR) superior to a QSAR derived only using physicochemical and substructural features.[18]

Silipo and Hansch pioneered in the structure–activity analysis of the Baker triazines.[18] Hansch and co-workers have continued in the structure–activity development of triazine inhibitors of dihydrofolate reductases (DHFR)s.[19,20] Silipo and Hansch suggest that the Baker triazine data base serve as a "standard" for evaluating quantitative structure–activity methods.[18] We adopted this suggestion and carried out our first study of some Baker

(1) A. J. Hopfinger, "Conformational Properties of Macromolecules", Academic Press, New York, 1973.
(2) A. J. Hopfinger and S. K. Tripathy In "Critical Review of Solid State Physics". D. E. Schuele and R. W. Hoffmann, Eds., CRC-Press, Boca Raton, Fl., Vol. 14, in press.
(3) L. B. Kier, "Molecular Orbital Theory in Drug Research", Academic Press, New York, 1971.
(4) H. J. R. Weintraub and A. J. Hopfinger, *J. Theor. Biol.*, **41**, 53 (1973).
(5) R. Potenzone, Jr., E. Cavicchi, H. J. R. Weintraug, and A. J. Hopfinger, *Comput. Chem.*, **1**, 187 (1977).
(6) L. B. Kier, *J. Pharmacol. Exp. Ther.* **164**, 75 (1968).
(7) N. S. Ham In "Molecular and Quantum Pharmacology", E. D. Bergmann and B. Pullman, Eds., Reidel-Dordrecht, Holland, 1974, p 261.
(8) D. C. Rohrer, D. S. Fullerton, A. H. L. From, and K. Ahmed In "Computer Assisted Drug Design". E. C. Olson and R. E. Christoffersen, Eds., ACS Monograph Series 112, Washington, D.C., 1979, p 259.
(9) L. B. Kier In "Fundamental Concepts in Drug-Receptor Interactions", J. Danielli, J. Moran, and D. Triggle, Eds., Academic Press, New York, 1970, p 15.

(10) G. Loew, D. Berkowitz, H. Weinstein, and S. Srebrenik In "Quantum and Molecular Pharmacology", E. D. Bergmann and B. Pullman, Eds., Reidel-Dordrecht, Holland, 1974, p 355.
(11) H. Weinstein, B. Z. Apfelderfer, S. Cohen, S. Maayani, and M. Sokolovsky In "Conformation of Biological Molecules and Polymers", E. D. Bergmann and B. Pullman, Eds., Academic Press, New York, 1973, p 531.
(12) G. M. Crippen, *J. Med. Chem.*, **22**, 988 (1979).
(13) K. Yamamoto, *J. Biochem.*, **76**, 385 (1974).
(14) D. Pensak, unpublished work, 1978.
(15) H. Weinstein, *Int. J. Quantum Chem.*, **2**, 59 (1975).
(16) J. Bartlett and H. Weinstein, *Chem. Phys. Lett.*, **30**, 441 (1975).
(17) See Table I of ref 18.
(18) C. Silipo and C. Hansch, *J. Am. Chem. Soc.*, **97**, 6849 (1975).
(19) S. W. Dietrich, R. Nelson Smith, J. Y. Fukunaga, M. Olney, and C. Hansch, *Arch. Biochem. Biophys.*, **194**, 600 (1979).
(20) S. W. Dietrich, R. Nelson Smith, S. Brendler, and C. Hansch, *Arch. Biochem. Biophys.*, 612 (1979).

Table I. van der Waal Radii, $R_{\alpha,\beta}$, and Arbitrary $t_{\alpha,\beta}$ Atom Type Assignments

| atom	$R_{\alpha,\beta}$, Å	$t_{\alpha,\beta}$ [a]
C (sp^2)	1.50	4
C (sp^3)	1.65	4
N	1.40	9
O (sp)	1.30	11
O (sp^2)	1.35	11
F	1.40	14
Cl	1.80	15
S	2.05	19
Br	1.95	20

[a] These assignments are consistent with those in the CAMSEQ-II system for molecular structure determination.

triazines about 3 years ago. This study led to an incorrect model. However, these early calculations did provide a base for the work presented here. A report of this early study was prematurely published as part of a symposium proceedings by taping of a lecture.[21]

Methods

1. Molecular Shape Descriptors. The detailed theory of molecular shape analysis is not reported here. However, the shape descriptors related to the common overlap steric volume between pairs of molecules are defined in detail. These shape descriptors are essential to the shape-dependent QSAR analysis of the Baker triazines.
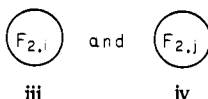
Let molecules i and j be of the shape form

F$_{1,i}$ — F$_{2,i}$ and   F$_{1,j}$ — F$_{2,j}$
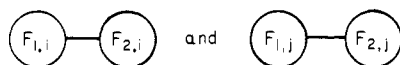
respectively, such that fragments

F$_{1,i}$ and   F$_{1,j}$

i   ii
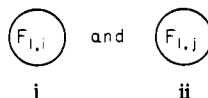
contain the same number of atoms which can be identically superimposed in space with respect to their atomic coordinates through some translation T and rotation **R**. Conversely

F$_{2,i}$ and   F$_{2,j}$

iii   iv

do not superimpose with respect to atomic positions under T and **R**, and/or the numbers of atom in iii and iv are different, and/or the overall set of atom types (sizes) in each fragment is different.

After performing operations T and **R** on j, let iv now be

F'$_{2,j}$

v

to denote i and ii are now identical. iii will contain n_i atoms each uniquely referred to by index k.

The "size" of atom (i,k) will be expressed in terms of a steric sphere whose size is determined by the van der Waals radius, $R_{i,k}$. The $R_{\alpha,\beta}$ used in this study are reported in Table I. Correspondingly for v the index is l and we define n_j and $R_{j,l}$.

The common overlap steric volume associated with atom (i,k) due to atom $l*$ of molecule j (see Figure 1 for a two-dimensional representation) is

$$v_0(i,k) \equiv v_{i,k} \cap v_{j,l}* \tag{1}$$

such that $l*$ is that atom in v for which

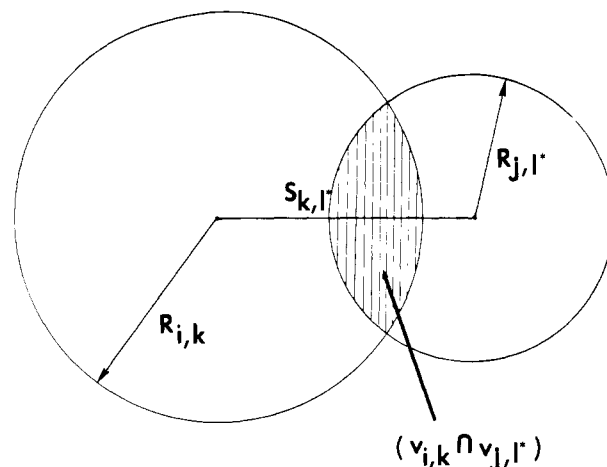$$s_{k,l} = [(x_{i,k} - x_{j,l})^2 + (y_{i,k} - y_{j,l})^2 + (z_{i,k} - z_{j,l})^2]^{1/2} \tag{2}$$

(21) A. J. Hopfinger In "Structural Correlates of Carcinogenesis and Mutagenesis", I. M. Asher and C. Zervos, Eds., The Office of Science. FDA, Washington, D.C., 1978, p 74.

Figure 1. A two-dimensional representation of the volume overlap of atom k of molecule i with atom $l*$ of molecule j. The overlap volume is denoted by the area having the vertical lines.

yields the minimum interatomic distance, $s_{k,l*}$. The total common overlap steric volume is then defined as

$$V_0(i,j) \equiv \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} (v_{i,k} \cap v_{j,l*}) \tag{3}$$

with $l*$ denoting that l yields $s_{k,l*}$ for *each* k. The sums over all atoms in each molecule make the results, e.g., $V_0(i,j)$, independent of the index assignments. The explicit form of $v_0(i,k)$ for $|R_{i,k} - R_{j,l*}| < s_{k,l*} < |R_{i,k} + R_{j,l*}|$ is

$$(v_{i,k} \cap v_{j,l*}) = (1/3)\pi[2R_{i,k}^3 + 2R_{j,l*}^3 + s_{k,l*}^3] - \pi[R_{i,k}^2 X + (s_{k,l*} - X)(R_{j,l*}^2 + s_{k,l*}X)] \tag{4}$$

$$X = (R_{i,k}^2 - R_{j,l*}^2 + s_{k,l*})/2s_{k,l*} \tag{5}$$

When $s_{k,l*} < |R_{i,k} - R_{j,l*}|$, the overlap volume is

$$(v_{i,k} \cap v_{j,l*}) = (4/3)\pi R_s^3 \tag{6}$$

where R_s is the smaller of $R_{i,k}$ and $R_{j,l*}$.

$V_0(i,j)$, as computed using eq 3, will be larger than the actual joint steric volume shared by molecular fragments i and j. The reason for this is that the overlap volume due to three, and more, spheres intersecting is neglected. It is also to be noted that $V_0(i,j)$ is a *pairwise* property which only has a meaning relative to i and j.

Consequently, a quantitative relationship between some observed property and shape similarity may not be best expressed through a cubic dependence in the $s_{k,l*}$, as is inherent to the common overlap steric volume function $V_0(i,j)$. Therefore, we arbitrarily define two other functions

$$S_0(i,j) = [V_0(i,j)]^{2/3} \tag{7a}$$

and

$$L_0(i,j) = [V_0(i,j)]^{1/3} \tag{7b}$$

which, respectively, establish second- and first-order dependencies in $s_{k,l*}$. S_0 has the dimensions of area, but is *not* a physical measure of common atomic surface areas between atoms in i and j. L_0 has the dimensions of length, but is *not* a cumulative measure of distances between atoms in i and j. S_0 and L_0 are to be viewed as alternate mathematical representations of V_0 that may be advantageous to use in establishing empirical QSARs.

2. Representation of Linear Regression Equations for Pairwise Shape Descriptors. The inclusion of pairwise molecular shape properties does *not* allow one to *directly* establish a correlation between activity and molecular descriptors. A pairwise shape descriptor is a relative internal measure between *two* molecules. This is in contrast to other physicochemical descriptors (like the water/octanol partition coefficient, P) that are absolute measures employing relative external scales and, consequently, are independent of the properties of other molecules in a data base. The

pairwise shape descriptors can be incorporated into a multidimensional linear regression formalism between molecular features and activity by adopting a corollary to a fundamental postulate of quantitative drug design. This postulate is:

Activity, A, is a function of the molecular properties, p_μ:

$$A = f_1(p_1, p_2, ..., p_\mu, ..., p_\psi) \qquad (8)$$

The corollary postulate, which we designate as (P-I), is:

(P-I) *The difference in activity, ΔA, between any two molecules is a function of the difference in measures of the corresponding molecular properties, Δp_μ:*

$$\Delta A = f_2(\Delta p_1, \Delta p_2, ..., \Delta p_\mu, ..., \Delta p_\psi) \qquad (9)$$

The pairwise shape descriptors are formally identified as members of the set of Δp_μ. This constitutes the second critical assumption, (P-II), of the theory:

(P-II) *Pairwise shape descriptors are measures of differences in shape between two molecules.*

A consequence of (P-I) is that N different multidimensional linear regression equations, each using N compounds, can be constructed to establish the functional relationship f_2. Each equation is based upon adopting the measures of the activity and molecular descriptors of a different member compound as a relative internal standard of comparison. These N different regression equations will be degenerate, with respect to one another, for all descriptors other than the pairwise shape factors. This follows directly from the fact that a pairwise shape factor is an internal measure between two molecules, while each set of the other like descriptors are measured relative to their respective common external scale. Consequently, the regression equation which yields the best fit, as measured by the correlation coefficient, R, and standard deviation, S, ultimately corresponds to that specific compound in the data base whose shape, relative to its pairwise comparison to the shapes of all other compounds in the data base, provides the most information for the regression fit.

If $(p_a)_b$ and $(p_a)_c$ are the measures of property a for, respectively, molecules b and c, then,

$$(\Delta p_a)_{b,c} = (p_a)_b - (p_a)_c \qquad (10)$$

except for the pairwise shape factors which are defined above. ΔA will represent the difference in activities. Let us adopt conventional procedures and include linear and squared descriptor value terms. The set of equations, relative to compound i, to be solved results from the minimization of:

$$G = ((\Delta A)_i - \sum_{\mu=1}^{\psi} [C_{1,\mu}(\Delta p_\mu)_i + C_{2,\mu}(\Delta p_\mu^2)_i])^2 \qquad (11)$$

with respect to the $C_{1,\mu}$ and $C_{2,\mu}$, which are the regression coefficients. ψ is the number of properties (independent variables). The solution of this set of equations, reexpressed directly in terms of the actual activity, A, of any arbitrary compound k (not necessarily part of parent data base), is

$$A_k = C_{1,1}(p_s)_{i,k} + C_{2,1}(p_s)_{i,k}^2 - \sum_{\mu=2}^{\psi}(C_{1,\mu}(p_\mu)_k + C_{2,\mu}(p_\mu^2)_k) +$$
$$\sum_{\mu=2}^{\psi}(C_{1,\mu}(p_\mu)_i + C_{2,\mu}(p_\mu^2)_i) + \Phi_i \qquad (12)$$

In eq 12 p_s is a pairwise shape descriptor, measured relative to some "standard" compound i, and arbitrarily associated with regression coefficients $C_{1,1}$ and $C_{2,1}$. Φ_i is the constant in the regression fit. The second summation term in eq 12 is also a constant. Therefore, these two constants can be combined as Φ_i', and eq 12 can ultimately be expressed as

$$A_k = C_{1,1}(p_s)_{i,k} + C_{2,1}(p_s)_{i,k}^2 - \sum_{\mu=2}^{\psi}(C_{1,\mu}(p_\mu)_k + C_{2,\mu}(p_\mu^2)_k) + \Phi_i' \qquad (13)$$

This expression differs in form from the "usual" structure-activity correlation equation only through the p_s terms. It is worth noting that the application of eq 13 may not be restricted to
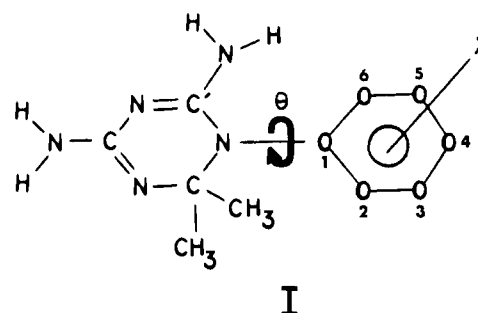
BAKER TRIAZINES

I

Figure 2. The general form of the Baker triazines used in the shape-dependent structure–activity analysis. θ defines the principal torsional rotation angle responsible for conformational freedom.

compounds which are structurally congeneric to those in the parent data base. The pairwise shape descriptor may provide the linkage information necessary to coanalyze structurally diverse compounds exhibiting common biological action.

Results and Discussion

1. Compound Selection and Conformation Analysis. The general form of the Baker triazines included in the study is shown as **I** in Figure 2. Substitutions on the 2 to 5 positions were considered. The principal degree of conformational freedom is torsional angle θ defined in Figure 2.

The CAMSEQ-II software system was employed to determine the preferred molecular conformations of **I**.[5] The valence geometry of **I** with X = H was arrived at using Allinger's consistent-force-field method as formulated in program MMI.[22,23] This program has been partially parameterized in our laboratory and is a component of the CAMSEQ-II system. The valence geometry of **I** with X = H was held fixed for all substituent modifications. The valence geometry of the substituents was based upon usage of "standard" bond lengths and angles.[1]

Conformational analysis, using a fixed valence geometry molecular mechanics formalism, is about two orders of magnitude faster, computation-wise, than any other means of three-dimensional molecular structure determination. However, it still was not practical to analyze all 256 compounds reported by Silipo and Hansch.[18] Therefore, some filtering technique was required. Fortunately, the Baker triazines can be divided into seven classes with respect to conformational freedom. The successful inclusion of members of each class in a structure–activity analysis gives implicit confidence that any compound in the data base could have been selected. These classes are defined, in terms of there respective torsional rotation angles (conformational properties), in Table IIA. The footnotes in Table II define the manner in which the conformational space of each compound was scanned. The calculated atomic coordinates for the θ = 310° conformation of the unsubstituted compound are given in Table IIB. The principal point to make is that the conformational scans are uniform with respect to structural class. The compounds of each structural class used in the structure–activity analysis are denoted by numbers in Table II. The number assignment is given as part of Table III.

The observed activities, as measured by log $(1/C)$, where C is the molar concentrations of the drug for 50% in vitro inhibition of DHFR from either Walker 256 tumor or L1210 leukemia cells,[17] are reported as part of Table III. The range in activity is 5.11 log $(1/C)$ units or 128 825 molar concentration units. This is 96% of the total range in activity of the parent data which have been constructed by Silipo and Hansch.[18] The activities of the compounds given in Table III are quite uniformily distributed over the activity scale. However, from a shape-analysis point of view,

(22) N. L. Allinger, J. T. Sprague. *J. Am. Chem. Soc.*, **95**, 3893 (1973).
(23) "Quantum Chemistry Program Exchange", program No. 318, Chemistry Department. Indiana University.
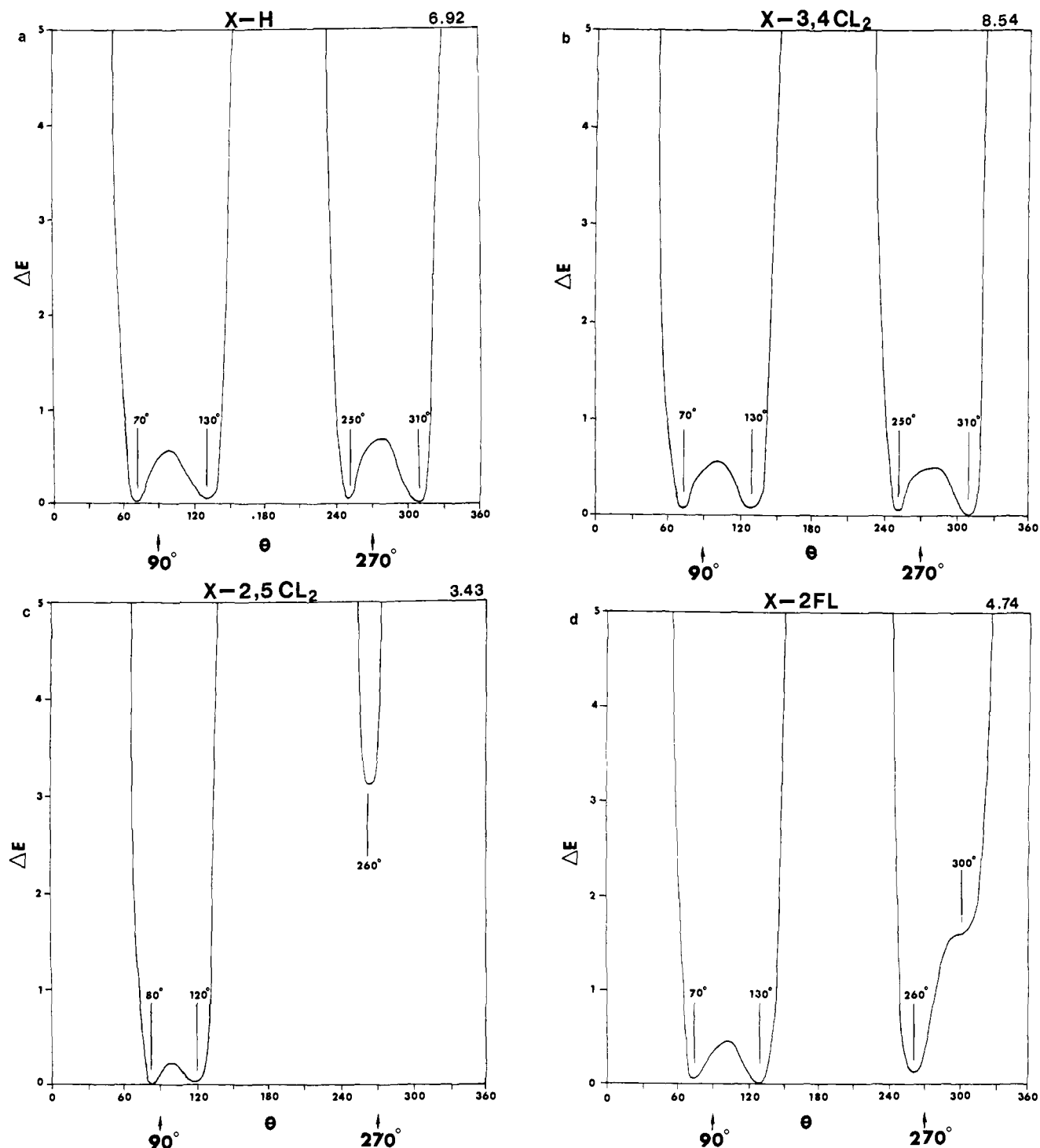
Molecular Shape Analysis for Quantitative Drug Design

J. Am. Chem. Soc., Vol. 102, No. 24, 1980 7199

Figure 3. The relative intramolecular energy ΔE, reported in kcal/mol, versus torsional rotation angle θ. $E \equiv 0$ is the global energy minimum. The 90 and 270° major energy wells are defined as are individual local energy minima. (a) X = H, (b) X = 3,4-Cl$_2$, (c) X = 2,5-Cl$_2$ and (d) X = 2-Fl.

the most and least active compounds in the data base of Table III are central to this work. The most active molecule has a 3,4-Cl$_2$ substitution and the least active has 2,5-Cl$_2$ substituents. All total molecular physicochemical descriptors, except molecular shape, are nearly identical for these two congeners.

In order to initiate a shape term into a structure–activity model it is necessary to first define a conformation critical to activity. We did this in the following way.

A conformational analysis was done for the unsubstituted compound (no. 13). The relative intramolecular conformational energy, ΔE, as a function of θ is shown in Figure 3a. Two major potential wells are found centered around θ = 90 and 270°. Each of these major wells possesses two subwells so that four conformational energy minima are realized at θ = 70, 130, 250, and

310°. Stereographic stick models of each of the these four conformer states are shown in Figure 4 for the 3-Cl compound (no. 7).

The conformational analysis of the 3,4-Cl$_2$ (no. 1) structure was next done. The ΔE vs. θ plot is virtually identical with that of the unsubstituted compound; see Figure 3b. Thus the increased activity of (no. 1) over (no. 13) cannot be attributed to conformational preference, although the shapes are different owing to the Cl's as opposed to the H's. However, the ΔE vs. θ plot for the inactive 2,5-Cl$_2$ compound (no. 27), shown in Figure 3c, is markedly different from the other two conformational energy plots. Specifically, the major well centered at 270° is shrunken to a minor well centered at θ = 260°, which is about 3.1 kcal/mol higher than the global energy minimum conformation, (θ = 120°). We

Table II

A. Types of Baker Triazines Based Upon Primary Structure		
type	structure	examples (compd no.)[c]
I		1, 7, 13, 17, 21, 22, 23, 25, 26,[d] 27
II		6, 9, 10,[e] 11, 14
III		10,[e] 23
IV		4, 15, 16, 19, 20
V		3, 12, 18
VI		2
VII		5, 8

B. Relative Atomic Coordinates for the $\theta = 310°$ Conformer State of the Unsubstituted Compound (no. 13 of Table III)

atom	X, Å	Y, Å	Z, Å	atom	X, Å	Y, Å	Z, Å
H_1	0.89	1.01	-2.09	H_{17}	1.52	-2.40	-1.85
N_2	0.32	0.98	-1.25	H_{18}	2.40	-3.61	-0.88
H_3	-0.50	1.58	-1.37	H_{19}	0.64	-3.75	-1.10
C_4	0.22	-0.16	-0.53	N_{20}	1.31	-0.75	-0.03
N_5	-0.94	-0.72	-0.47	C_{21}	2.50	-0.08	-0.05
C_6	-1.11	-1.96	0.26	C_{22}	3.65	-0.63	-0.52
N_7	-2.32	-2.45	0.48	H_{23}	3.61	-1.56	-0.96
H_8	-3.11	-1.97	0.15	C_{24}	4.84	0.02	-0.44
H_9	-2.43	-3.31	0.98	H_{25}	5.70	-0.46	-0.73
N_{10}	-0.10	-2.51	0.77	C_{26}	4.90	1.29	0.03
C_{11}	1.23	-2.17	0.35	H_{27}	5.80	1.79	0.05
C_{12}	2.20	-2.59	1.48	C_{28}	3.77	1.90	0.46
H_{13}	2.43	-3.65	1.39	H_{29}	3.79	2.87	0.79
H_{14}	3.12	-2.01	1.40	C_{30}	2.61	1.18	0.44
H_{15}	1.74	-2.40	2.45	C_{31}	1.77	1.62	0.83
C_{16}	1.47	-3.05	-0.98				

[a] This primary bond rotation is denoted as θ. The reference state for $\theta = 0°$ has C'N of the triazine ring cis to C_1C_6 of the phenyl ring. See Figure 1 for atom designations. In the conformational analysis of type I compounds the increment in θ, $\theta\Delta$, is 10° from 0° to 350°. In the types II and III compounds $\Delta\theta$ is 10° for the ranges 60° to 140° and 230° to 330°. For types IV and V $\Delta\theta$ is 10° for the ranges 60-80°, 120-140°, 240-160°, and 300-320°. For types VI and VII compounds θ was fixed at 70°, 130°, 250°, and 310°, respectively. [b] Rotations about single bonds of the 3 and/or 4 substituent groups are denoted as $\phi_{3,i}$ and $\phi_{4,i}$, respectively. $i = 1, 2, 3$ depending on whether the bond rotation is first, second, or third "closest" to the parent ring. All $\Delta\phi_{k,i} = 30°$ and the reference conformation is always chosen as the extended trans-planar conformation being co-planar to the aromatic ring. Whenever possible, symmetry is used to reduce the number of nonunique conformer states. [c] Compound number (no.) in Table III. [d] There are two ϕ rotations considered for this compound. [e] This compound has both $\phi_{3,1}$ and $\phi_{4,1}$ rotations.

therefore assume the 270° major well to be involved in specifying activity. However, it is not possible to deduce whether the 250°, or 310°, or some other conformer in the 270° well is specific to activity. Figure 3d is a ΔE vs. θ energy plot of the 2-F1 compound (no. 22). This compound is about 20 times more active than 2,5-Cl₂, but less active by a factor of 150 than the unsubstituted molecule (no. 13). Figure 3d indicates that the 310° conformer state is destabilized by about 1.60 kcal/mol while the $\theta = 250°$ minima shifts to 260°, but is only about 0.2 kcal/mol above the global minimum at $\theta = 130°$. Therefore, we conclude and introduce as the third critical postulate in this paper, P-III;

(P-III) *The conformer state, corresponding to $\theta = 310°$, and shown in Figure 4d, is the active conformation for I. If this conformer cannot be sterically realized, the conformer found as an energy minimum in the 270° major well represents this active conformer.*

Figure 5a is a space-filling shaded graphics (red) picture of the 3,4-Cl₂ substituted phenyl ring of compound (no. 1) in the $\theta = 310°$ conformer state. The $\theta = 250°$ conformer of the 2,5-Cl₂ substituted phenyl ring of compound (no. 2) is shown in green in Figure 5b from the same external reference point. Figure 5c shows the spatial superposition of these two molecular fragments. The common overlap steric volume appears as light yellow. V_0 is the approximate measure of this common overlap steric volume. Dissimilarity in shape corresponds to the red and green "left over" volumes. The pink line in each figure denotes the N–C bond about

which the θ rotation takes place.

The significance of the empirical structure–activity relationships reported in this paper depends upon the validity of P-III. This
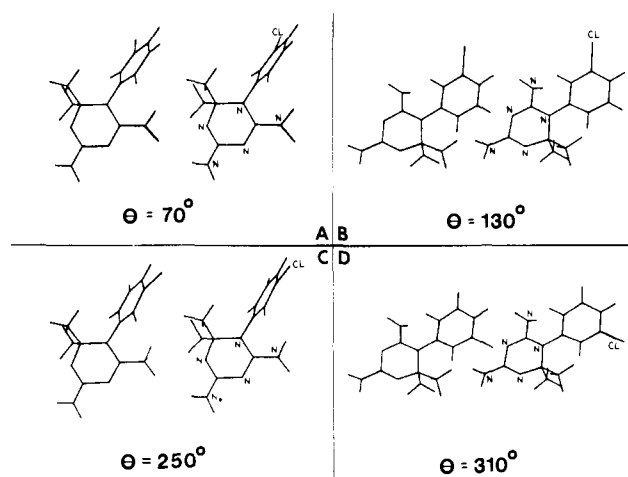

Figure 4. Stereographic stick models of the four local minimum energy conformers of the X = 3 and/or 4 substituted triazines. The X = 3-Cl molecule is used here. (a) $\theta = 70°$, (b) $\theta = 130°$, (c) $\theta = 250°$, and (d) $\theta = 310°$.

Table III. Absolute Parameter Data Table

compd	no.	obsd[c] activity	ΔE^{*d}	$\pi_3{}^e$	$\pi_4{}^f$	$\Sigma\pi^g$	$D_4{}^h$	Scheme I		Scheme II	
								pred activity	pred − obsd activity	pred activity	pred − obsd activity
3,4-Cl$_2$,^{j}	1	8.54	−0.11	0.94	0.94	1.88	0	7.79	0.75	7.66	0.88
3-(CH$_2$)$_2$-Ph	2	8.19	−1.52	3.05	0.23	3.28	0	8.32	−0.13	8.27	−0.08
4-CH$_2$-Ph	3	8.05	0.12	0.23	2.62	2.85	−0.25	7.95	0.10	8.08	−0.03
3-CH$_2$-Ph	4	8.00	−0.86	2.62	0.23	2.85	0	7.80	0.20	7.80	0.20
4-(CH$_2$)$_2$-Ph	5	7.89	−2.47	0.23	3.05	3.28	−0.25	7.66	0.23	7.72	0.17
3-CF$_3$	6	7.76	−0.35	1.33	0.23	1.56	0	7.44	0.32	7.35	0.41
3-Cl	7	7.76	−0.18	0.94	0.23	1.17	0	7.28	0.48	7.17	0.59
3-Cl, 4-OCH$_2$-Ph	8	7.52	0.12	0.94	1.89	2.83	−0.37	8.08	−0.56	8.23	−0.72
3-SO$_2$F	9	7.27	0.00	0.23	0.61	0.84	0	7.27	0.00	7.05	0.22
3-Ph,4-OH^k	10	7.14	−1.17	1.89	−0.44	1.45	−0.46	7.59	−0.45	7.78	−0.64
3-NO$_2$	11	7.07	−0.57	−0.03	0.23	0.20	0	6.94	0.13	6.78	0.30
4-CH$_2$CN	12	6.92	0.18	0.23	−0.61	−0.38	−0.27	6.56	0.36	6.56	0.36
H	13	6.92	−0.01	0.23	0.23	0.46	0	6.34	0.58	6.27	0.65
3-Ph	14	6.85	−0.80	1.89	0.23	2.12	0	7.61	−0.76	7.55	−0.70
3-COCH$_3$	15	6.79	−0.95	−0.20	0.23	0.03	0	6.79	0.00	6.62	0.18
3-DL-CH(OH)^a	16	6.71^i	−0.46	0.68	0.23	0.91	0	6.93	−0.22	6.84	−0.13
2,3-Cl$_2$	17	6.52	3.29	0.94	0.23	1.17	0	6.09	0.43	6.11	0.39
4-COCH$_2$Cl-Ph	18	6.45	−0.52	0.23	−0.34	−0.11	−0.32	6.76	−0.31	6.80	−0.35
3-COCH$_2$Cl	19	6.21	−1.23	−0.34	0.23	−0.11	0	6.61	−0.40	6.46	−0.25
3-OCH$_3$	20	6.17	−1.01	0.28	0.23	0.51	0	6.69	−0.52	6.59	−0.42
4-CN^b	21	5.14	0.05	0.23	−0.34	−0.11	0.91			5.91	−0.77
2-Fl	22	4.74	0.23	0.23	0.23	0.46	0	4.24	0.49	4.36	0.38
4-Ph^b	23	4.70	0.00	0.23	1.89	2.12	2.41			4.63	0.07
2-Br	24	4.25	4.66	0.23	0.23	0.46	0	4.35	−0.10	4.46	−0.21
2-Cl	25	4.15	3.32	0.23	0.23	0.46	0	3.92	0.22	4.07	0.08
2-OCH$_3$	26	3.68	2.43	0.23	0.23	0.46	0	4.09	−0.41	4.22	−0.54
2,5-Cl$_2$	27	3.43	3.30	0.23	0.23	0.46	0.72	3.87	−0.44	3.45	−0.02

^{a} D form yields best correlations. ^{b} Not used to construct Schemes I and III. ^{c} As given by Silipo and Hansch in ref 18. ^{d} ΔE^* is the difference in energy between the lowest conformer state in the 270° well from the lowest energy conformer in the 90° well. Individual conformer energies are <0. Units are kcal/mol. ^{e} π value of the 3 substituent. The f-fragment values were used, ref 26. ^{f} π value of the 4 substituent. The f-fragment values were used, ref 26. ^{g} $\Sigma\pi$ – the sum of the π values for the 3 and 4 substituents; default is the hydrogen π value of 0.23. ^{h} $D_4 = D$(C-X) − D(C-Cl), where D(C-X) is the maximum distance as defined by atom X that a 4-substituent group traverses along a vector defined by C$_1$C$_4$. D(C-Cl) is the bond length, 1.77 A, of the C-Cl bond. Units are A. ^{i} This compound was made at Smith-Kline Laboratories and independently tested by Hansch and co-workers. The measured activity has been "adjusted" to the Baker activity scale. ^{j} Schemes I and II were developed using this compound as the relative standard. ^{k} Schemes III and IV were developed using this compound as the relative standard.

is an important point to emphasize because the crystal structures of Baker's soluble and insoluble antifol triazines have been determined.[24] The conformation of these two molecules with respect to θ are virtually identical. A stereographic drawing of the soluble antifol is shown in Figure 6.

The triazine ring puckering at the carbon to which the two methyls are bonded can be "above" or "below" the "plane" of the ring. The conformational energy is degenerate with regard to the choice of puckering. The crystal structures of the two antifols have the opposite puckering to the triazine ring structure used in the conformational analyses reported here. When this is taken into consideration the crystal structures are found to correspond to the θ = 70° conformational state. Atomic coordinates and torsional angles are not reported for the crystal structures. However, stereographic model-building studies suggest that $\theta \simeq$ 76° in both crystal structures.

Puckering at the saturated triazine ring carbon results in the two substituent methyl groups being axial and equatorial to the ring. The θ = 310° conformer state is destabilized for X substitutions at the 2 position because the (2-X) group and the axial methyl from the triazine ring are very close to one another. Only (2-X) = H can be sterically tolerated.

Compound (no. 8)-[3-Cl,4-OCH$_2$Ph] is a good model for the soluble Baker antifol with respect to the conformational preference of θ. The global energy minimum of each of the four θ conformers (70°, 130°, 250°, 310°) has been rigorously minimized, using the energy minimizer in CAMSEQ-II. The 4-OCH$_2$Ph substituent was found to be virtually identical in conformation to the crystal structure for *all* four θ conformers. The θ = 70° conformer minimized to θ = 72°. More significant, however, the difference

in the relative energy, ΔE, for each of the four energy minimized conformers is ΔE_{70} = 0.16, ΔE_{130} = 0.00, ΔE_{250} = 0.12, and ΔE_{310} = 0.09, where the energy units are kcal/mol. This suggests that there is little *intramolecular* conformational preference for any one of these four θ states. In turn, it is quite reasonable to hypothesize that the θ state realized in the crystal structure is dictated by "packing" forces, while that adopted at the enzyme receptor is controlled by "binding" forces. It is also quite possible that the "packing" and "binding" forces are different. Consequently, the respective θ conformer states may be different. Obviously, we are trying to make the case that the crystal conformation may not be the active conformation.

An attempt has been made to use the θ = 70° conformational state as the "active" conformation in shape-analysis dependent structure–activity correlations. The resulting structure–activity equations are marginal, especially when compared to those using the θ = 310° conformer as the active state. The details are reported below.

2. Selection of Descriptors and Generation of the Correlation Equations. Each of the pairwise shape descriptors V_0, S_0, and L_0 were individually considered as representative measures of shape. Linear and squared shape terms were considered in the construction of the correlation equations.

Hansch and co-workers have reported that selected π and molar refractivity, MR, substituent constants of the 3 and/or 4 substitution groups exhibit a correlation with log $(1/C)$.[18-20] Consequently, these descriptors were included in our studies. The π values[29] were computed with the use of the f-fragment model.[25] MR, as well as aqueous activity coefficient, aqueous solubility

(24) A. Camerman, H. W. Smith, and N. Camerman, *Biochem. Biophys. Res. Commun.*, **83**, 87 (1978).

(25) C. Hansch and A. Leo In "Substituent Constants for Correlation Analysis in Chemistry and Biology". Wiley-Interscience, New York, 1979.
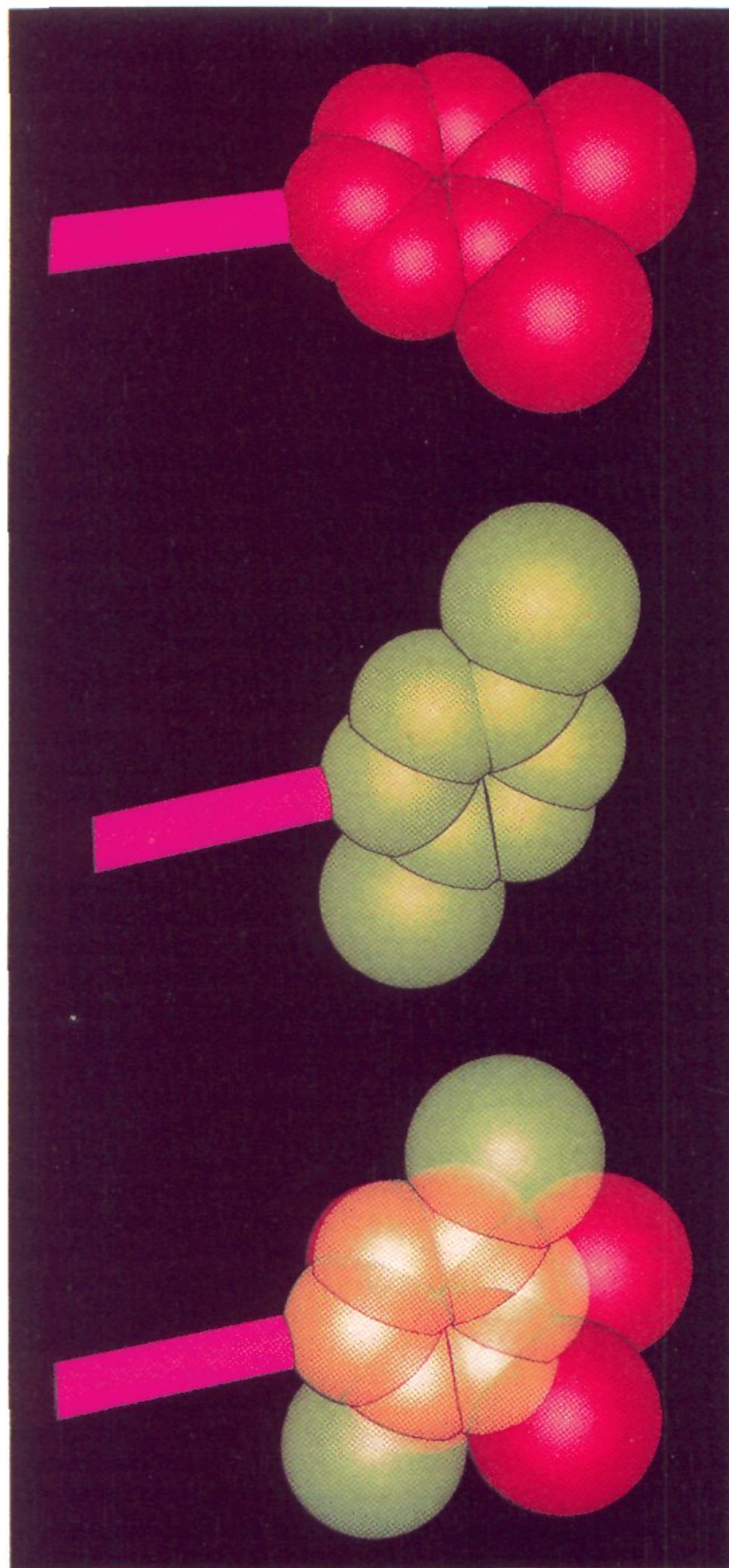(26) R. D. Cramer, III. *J. Am. Chem. Soc.*, **102**, 1837 (1980).

Figure 5. Space-filling shaded graphics representations of (a, top) the $3,4$-Cl$_2$ substituted phenyl ring and (b, middle) the $2,5$-Cl$_2$ substituted phenyl ring, illustrating the van der Waals shapes for the $\theta = 310$ and $250°$ conformers, respectively. Part c (bottom) illustrates the overlap volume (in light yellow) and volume differences (in red and green) for the two conformer fragments.

parameter, log of the dielectric constant, and dipole moment, were computed with the use of the BC(DEF) model of Cramer.[27,28] The difference in conformational energy, ΔE^*, between the minimum energy conformers in the 270 and 90° wells was also used as a descriptor. ΔE^* was included to see if the relative intramolecular stability of the conformer states had any influence on activity. The set of compounds, reference numbers, absolute activities, and some absolute nonshape descriptor values are given in Table III.

(27) R. D. Cramer, III, *J. Am. Chem. Soc.*, **102**, 1849 (1980).
(28) R. F. Rekker, "The Hydrophobic Fragment Constant", Pharmaco-chemistry Library, Vol. 1, Elsevier, New York, 1977.
(29) These are termed *f* values (fundamental fragments) in the revised formalisms.

Each of the compounds listed in Table III was used as the reference state in the set of shape analyses. The optimum correlation was found with the use of (no. 1)-[3,4-Cl$_2$] as the reference compound. The *relative* properties for the data base are given in Table IV, using (no. 1) as the reference. The corresponding structure–activity is shown in Scheme I.

Scheme I

$$\Delta \log (1/C) = -1.474[S_0] + 0.02240[S_0]^2 + 0.378\,[\Delta\Sigma\pi] + 24.98 \text{ (A)}$$

$$\log (1/C) = 1.474[S_0] - 0.02240[S_0]^2 + 0.378[\Sigma\pi] - 17.15 \text{ (B)}$$

$$N = 25 \qquad R = 0.960 \qquad S = 0.40$$

Equation A in Scheme I is the relative-activity to relative-property representation, while eq B is the absolute-activity to absolute-property correlation. A couple of points are in order concerning Scheme I. First, $\Delta\Sigma\pi$ ($\Sigma\pi$), the sum of the π values of the 3 and 4 substituents, is the only nonshape descriptor which significantly enhances an activity-shape correlation. Only the linear term is necessary. This is the reason only π values are reported in Table III and not other descriptors. The ΔE^* are included in Table III as a point of information regarding intramolecular conformational energetics. Second, both V_0 and L_0, as well as S_0, correlate well with $\Delta \log (1/C)$. The correlation coefficient, using V_0 and L_0 with $\Delta\Sigma\pi$, is about $R \simeq 0.920$ for both shape descriptors.

Last, two compounds (no. 21)-[4-CN] and (no. 23)-[4-Ph] could not be handled with the use of equations of the form of Scheme I. Equation A predicts both these compounds to be more active than is observed. Compound (no. 21) is predicted to have $\log (1/C) = 6.88$ (50 times more active than observed) and (no. 23) has a predicted $\log (1/C) = 7.83$ (1500 times more active than observed). These two compounds share one common shape property: The spatial location of the X = 4-CN group and the $4'$-CH of X = 4-C$_6$H$_5$ are independent of all torsional rotations including θ and $\phi_{4,1}$. The geometric center of the nitrogen atom of the 4-CN, and the $4'$-carbon of the 4-C$_6$H$_5$, lie on a vector defined by C$_1$C$_4$ of the substituted benzyl ring such that both atoms are more than 2.4 Å from 4-C. This is illustrated in Figure 7 for the $\theta = 310°$ conformer state. No other compounds in the set of 27 have atoms which occupy the space common to the 4-CN and 4-C$_6$H$_5$ groups.

We make the assumption that this space is occupied by receptor site groups of DHFR which causes steric overlaps and diminishes drug binding and, consequently, activity. The loss in activity is introduced into Scheme I through a parabolic dependence on a distance term ΔD_4. ΔD_4 is the maximum difference in length that some particular atom of a 4-substituent group traverses along the $\overline{C_1C_4}$ of the substituted benzyl ring as compared to the length of a (C–Cl) bond (1.77 Å). In doing this we further speculate that a 4-Cl group represents the near-optimum in size for a substituent centered along the $\overline{C_1C_4}$ vector. The ΔD_4 are given in Table IV and the corresponding D_4 in Table III. It was also assumed the 5-Cl of compound (no. 27)-[2,5Cl$_2$] follows the ΔD_4 dependence with a (C–H) = 1.05 Å bond length representing the optimum value. Silipo and Hansch[18] suggest a "semirigid area" between the 3 and 4 positions of the substituted benzyl ring. This may be an alternate representation of the ΔD_4 model.

The series of linear regression analyses were repeated, using the functional form of Scheme I plus the ΔD_4 terms. The resulting correlation equation, with compound (no. 1) as the shape reference compound, is shown in Scheme II.

Scheme II

$$\Delta \log (1/C) = -1.384[S_0] + 0.02127[S_0]^2 + 0.434[\Delta\Sigma\pi] - 0.574[\Delta D_4] + 0.294[\Delta D_4]^2 + 23.38 \text{ (A)}$$

$$\log (1/C) = 1.384[S_0] - 0.02127[S_0]^2 + 0.434[\Sigma\pi] - 0.574[D_4] - 0.294[D_4]^2 - 15.66 \text{ (B)}$$

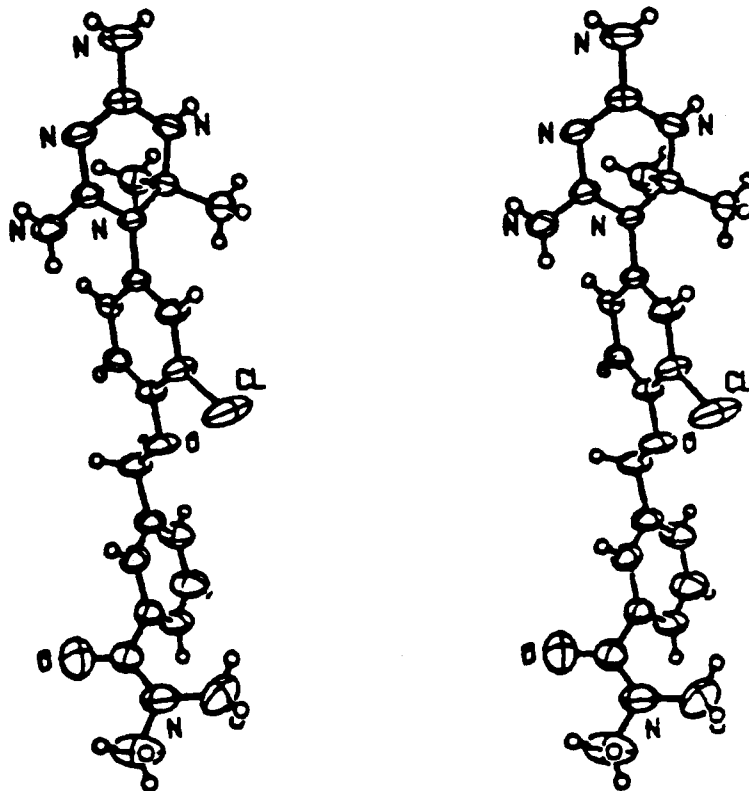$$N = 27 \qquad R = 0.953 \qquad S = 0.44$$

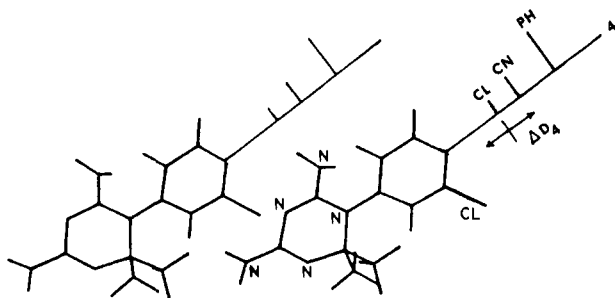**Figure 6.** Stereoscopic drawing of the crystal structure of Baker's soluble antifol. Taken from ref 24.

Figure 7. Definition of the steric descriptor ΔD_4 to account for the apparent loss in activity of X = 4-CN and 4-C$_6$H$_5$ compounds.
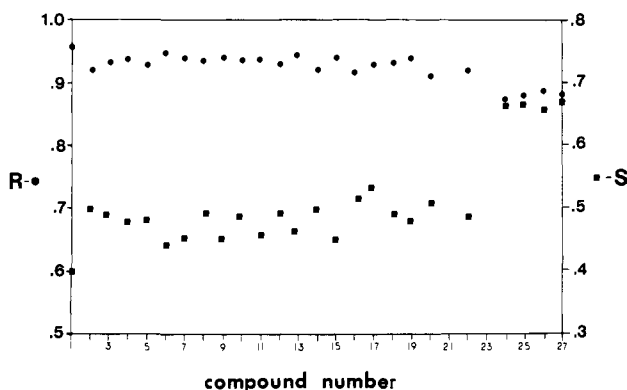

compound number

Figure 8. Correlation coefficient, R (●), and standard deviation, S (■), of the linear regression equations analgous to Scheme I versus the compound (as defined by number in Table III) selected as shape reference molecule to compute S_0.

Scheme II meaningfully accounts for the activity of all 27 compounds. Table III contains absolute predicted activities and Table IV predicted differences in activities relative to (no. 1).

Figure 8 contains plots of the correlation coefficient, R, and standard deviation, S, as a function of the choice of shape reference compound for linear regression equations analogous to Scheme I. The compound numbers are based upon the assignments made

in Table III. As described above, compound (no. 1) gives the highest R and lowest S. However, all equations have $R > 0.85$, and only equations employing 2-X substituted compounds as the shape reference "standard" have S significantly greater than 0.50. Several correlation equations are about as significant as Scheme I with respect to R, but not S. If pairwise shape descriptors are not used in the regression fit, the highest value is $R < 0.55$. Terms involving ΔE^* do not make meaningful contributions to the structure–activity correlation equations.

As a point of information and perspective, an example of a "good" structure–activity equation, found in constructing Figure 8, results from using compound (no. 10)-[3-ϕ,4-OH] as the pairwise shape-reference structure. The relative properties, including relative activities, are given in Table V. The regression equations, analogous to Schemes I and II, eq A, are, respectively, Schemes III and IV.

Scheme III

$$\Delta \log (1/C) = -1.143[S_0] + 0.01671[S_0]^2 + 0.442[\Delta\Sigma\pi] + 18.99$$

$$N = 25 \qquad R = 0.938 \qquad S = 0.49$$

Scheme IV

$$\Delta \log (1/C) = -0.996[S_0] + 0.01450[S_0]^2 + 0.481[\Delta\Sigma\pi] + 0.296[\Delta D_4] + 0.528[\Delta D_4]^2 + 16.66$$

$$N = 27 \qquad R = 0.940 \qquad S = 0.49$$

The ability to take into account the loss in activity as a result of ortho substitution (groups at the 2 position of the benzyl ring) is critical to generating a meaningful correlation equation.

All S_0 values used to establish Schemes I–IV employed the P-III postulate that the $\theta = 310°$ conformer is the "active" state. The $\theta = 70°$ conformer state was also postulated to be the active conformer in alternate pairwise shape analyses. The pairwise shape analyses in this latter model were carried out against the $\theta = 130°$ state, it it was lower in energy than $\theta = 70°$. The global energy minimum was also used in place of the $\theta = 70°$ conformer state in yet another study. Thus V_0, S_0, and L_0 values were generated for θ conformer pairs (70° or 130°) and, second (70° or global energy minimum θ). These additional structure–activity analyses

Table IV. Relative Properties of the Triazines Using the 3,4-Cl$_2$ Compound (no. 1) as the Reference

									Scheme I		Scheme II	
Δ no.'s[g]	$V_0/10^{2\,a}$	$S_0{}^b$	$S_0{}^2$	$\Delta\Sigma\pi^c$	$\Delta\pi_4{}^d$	$\Delta D_4{}^e$	$\Delta(D_4{}^2)$	obsd Δact.	pred Δact.	pred – obsd Δact.	pred Δact.	pred – obsd Δact.
(1-1)	1.854	32.5	1057.5	0	0	0	0	0	0.75	0.75	0.88	0.88
(1-2)	1.849	32.5	1056.3	-1.40	0.71	0	0	0.35	0.22	-0.21	0.27	-0.08
(1-3)	1.623	29.8	888.0	-0.97	-1.68	0.25	0.063	0.49	0.59	0.10	0.46	-0.03
(1-4)	1.547	28.8	829.4	-0.97	0.71	0	0	0.54	0.74	0.20	0.74	0.20
(1-5)	2.363	38.3	1466.9	-1.40	-2.11	0.25	0.063	0.65	0.88	0.23	0.83	0.17
(1-6)	1.617	29.7	882.1	0.32	0.71	0	0	0.78	1.10	0.32	1.19	0.41
(1-7)	1.610	29.6	876.2	0.71	0.71	0	0	0.78	1.26	0.48	1.37	0.59
(1-8)	2.039	34.6	1197.2	-0.95	-0.95	0.37	0.137	1.02	0.46	-0.56	0.31	-0.72
(1-9)	2.100	35.3	1246.1	1.04	0.33	0	0	1.27	1.27	0.00	1.49	0.21
(1-10)	1.779	31.6	998.6	0.43	1.38	0.46	0.212	1.40	0.95	-0.45	0.76	-0.64
(1-11)	1.627	29.8	888.0	1.68	0.71	0	0	1.47	1.60	0.13	1.76	0.29
(1-12)	1.549	28.8	829.4	2.26	1.55	0.27	0.073	1.62	1.98	0.36	1.98	0.36
(1-13)	1.366	26.5	702.3	1.42	0.71	0	0	1.62	2.20	0.58	2.27	0.65
(1-14)	1.592	29.4	864.5	-0.24	0.71	0	0	1.69	0.93	-0.76	0.99	-0.70
(1-15)	1.579	29.2	852.6	1.85	0.71	0	0	1.75	1.75	0.00	1.92	0.18
(1-16)	1.494	28.2	795.2	0.97	0.71	0	0	1.83	1.61	-0.22	1.70	-0.13
(1-17)	1.237	24.9	620.1	0.71	0.71	0	0	2.02	2.45	0.43	2.43	0.39
(1-18)	1.595	29.4	864.4	1.99	1.27	0.32	0.102	2.09	1.78	-0.31	1.74	-0.35
(1-19)	1.495	28.3	798.8	1.99	0.71	0	0	2.33	1.93	-0.40	2.08	-0.25
(1-20)	1.468	27.8	772.8	1.37	0.71	0	0	2.37	1.85	-0.52	1.95	-0.42
(1-21)	1.606	29.6	873.2	1.99	1.28	-0.91	0.828	3.40			2.63	-0.77
(1-22)	0.984	21.3	453.7	1.42	0.71	0	0	3.80	4.30	0.49	4.18	0.38
(1-23)	1.684	30.5	927.5	-0.24	-0.95	-2.41	5.808	3.84			3.91	0.07
(1-24)	0.993	21.5	462.3	1.42	0.71	0	0	4.29	4.19	-0.10	4.09	-0.20
(1-25)	0.939	20.7	428.5	1.42	0.71	0	0	4.39	4.62	0.22	4.47	0.08
(1-26)	0.965	21.0	441.0	1.42	0.71	0	0	4.86	4.45	-0.41	4.33	-0.54
(1-27)	0.939	20.6	424.4	1.42	0.71	-0.72[f]	0.518	5.11	4.67	-0.44	5.09	-0.02

[a] V_0 is the common overlap steric volume computed using eq 2–6 (see text). Units are A^3. [b] $S_0 = [V_0]^{2/3}$, units are A^2. [c] $\Delta\Sigma\pi = \Sigma\pi_i - \Sigma\pi_j$ for $j = 1, 27$ denoting compound no. as defined in Table III as in $\Sigma\pi$. [d] $\Delta\pi_4 = \pi_4(1) - \pi_4(j)$ for $j = 1, 27$. See Table III for definition of π_4 and j. [e] $\Delta D_4 = D_4(1) - D_4(j)$ for $j = 1, 27$, units are A. See Table III for definition of D_4. [f] It is assumed that substitutions at the 5 position can be represented by the D_4 function. The value of -0.72 is the difference between the C–Cl bond length and the C–H bond length. Units are A. [g] Note the numbers used in the first column of this table are as defined in Table III.

Table V. Relative Properties of the Triazines Using the 3-Ph, 4-OH Compound as the Reference

									Scheme III		Scheme IV	
Δ no.'s	$V_0/10^2$	S_0	$S_0{}^2$	$\Delta\Sigma\pi$	$(\Delta\Sigma\pi)^2$	ΔD_4	$\Delta D_4{}^2$	obsd Δact.	pred Δact.	pred – obsd Δact.	pred Δact.	pred – obsd Δact.
(10-1)	1.779	31.6	979.7	-0.43	0.18	-0.46	0.21	-1.40	-0.94	0.46	-0.83	0.57
(10-2)	1.893	32.9	1082.4	-1.83	3.35	-0.46	0.21	-1.05	-1.33	-0.28	-1.31	-0.26
(10-3)	1.499	28.2	795.2	-1.40	1.96	-0.21	0.04	-0.91	-0.57	0.34	-0.60	0.31
(10-4)	1.800	31.9	1017.6	-1.40	1.96	-0.46	0.21	-0.86	-1.08	-0.22	-1.05	-0.19
(10-5)	1.566	29.1	846.8	-1.83	3.35	-0.21	0.04	-0.75	-0.92	-0.17	-0.96	-0.21
(10-6)	192.2	33.3	1108.9	-0.11	0.01	-0.46	0.21	-0.62	-0.59	0.03	-0.50	0.12
(10-7)	167.6	30.4	924.1	0.28	0.08	-0.46	0.21	-0.62	-0.19	0.43	-0.10	0.52
(10-8)	2.713	41.9	1755.6	-1.38	1.90	-0.09	0.01	-0.38	-0.17	0.21	-0.29	0.09
(10-9)	2.626	41.0	1681.0	0.61	0.37	-0.46	0.21	-0.13	0.49	0.62	0.48	0.61
(10-10)	2.549	40.3	1624.1	0	0	0	0	0	0.07	0.07	0.08	0.08
(10-11)	1.959	33.7	1135.7	1.25	1.56	-0.46	0.21	0.07	0.01	-0.06	0.15	0.08
(10-12)	1.463	27.8	772.8	1.83	3.35	-0.19	0.04	0.22	0.94	0.72	1.03	0.81
(10-13)	1.366	26.4	697.0	0.99	0.98	-0.46	0.21	0.22	0.90	0.68	0.93	0.71
(10-14)	2.528	40.0	1600.0	-0.67	0.45	-0.46	0.21	0.29	-0.28	-0.57	-0.32	-0.61
(10-15)	1.938	33.5	1122.3	1.42	2.02	-0.46	0.21	0.35	0.09	-0.26	0.23	-0.11
(10-16)	1.724	31.0	961.0	0.54	0.29	-0.46	0.21	0.43	-0.14	-0.57	-0.04	-0.47
(10-17)	1.114	23.2	538.2	0.28	0.08	-0.46	0.21	0.62	1.59	0.98	1.47	0.85
(10-18)	1.427	27.3	745.3	1.56	2.43	-0.14	0.02	0.69	0.94	0.25	1.00	0.31
(10-19)	1.749	31.3	979.7	1.56	2.43	-0.46	0.21	0.93	0.28	-0.65	0.42	-0.51
(10-20)	1.674	30.4	924.2	0.94	0.88	-0.46	0.21	0.97	0.11	-0.86	0.22	-0.75
(10-21)	1.487	28.1	789.6	1.56	2.43	0.45	0.20	2.00			1.11	-0.89
(10-22)	0.984	21.3	453.7	0.99	0.98	-0.46	0.21	2.40	2.67	0.27	2.48	0.08
(10-23)	1.512	28.4	806.6	-0.67	0.45	1.95	3.80	2.44			2.34	-0.10
(10-24)	0.939	20.7	428.5	0.99	0.98	-0.46	0.21	2.89	2.93	0.04	2.71	-0.18
(10-25)	0.939	20.7	428.5	0.99	0.98	-0.46	0.21	2.99	2.93	-0.06	2.71	-0.28
(10-26)	0.952	20.9	436.8	0.99	0.98	-0.46	0.21	3.46	2.84	-0.62	2.63	-0.82
(10-27)	0.939	20.7	428.5	0.99	0.98	1.18	1.39	3.71	2.93	-0.78	3.82	0.11

were considered because the crystal conformation of Baker's soluble antifol has $\theta \simeq 70°$. No regression equation having $R > 0.835$ could be constructed for $\theta = 70°$ as the "active" conformer state!
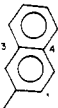
3. Testing the Correlation Equations.

A good way to test a correlation equation is to see how well it predicts the activity of some known compounds not included in the data base used to construct the equation. This has been done by using the seven

Table VI. Prediction of the Activity of Some Known Compounds Using Schemes I and II, Equations B

| | | | | | obsd | calcd activities | | | |
| | | | | | | Scheme I, eq B | | Scheme II, eq B | |
compd	no.	S_0, A^2	$\Sigma\pi$	D_4, A	log (1/C)	log (1/C)	obsd − pred	log (1/C)	obsd − pred
2-CH_3	28	21.6	0.46	0	4.00	4.41	−0.41	4.51	−0.51
3-SO_2-NH_2	29	29.7	−2.94	0	5.32[a]	5.75	−0.43	5.40	−0.08
3-$CONH_2$	30	30.1	−1.74[b]	0	5.70[a]	6.26	−0.56	5.97	−0.27
3-OH	31	28.4	−0.21	0	6.38[a]	6.56	−0.28	6.40	−0.02
3-Fl	32	28.0	0.60	0	7.45[a]	6.78	0.67	6.68	0.77
3C(CH_3)$_3$	33	29.7	2.59	0	7.50[a]	7.85	−0.35	7.80	−0.30
3-CN	34	29.6	−0.11	0	7.69[a]	6.81	0.98	6.62	1.07

[a] Taken from Table III of ref 19. The actual log (1/C) were made from bovine liver DHFR. It was found for the six compounds common to Table III of ref 19 and Table III of this work that the log (1/C) differed by an average of 0.63 ± 0.10 with the Table III activities in this work being larger. Hence obsd log (1/C) reported in this table are those givien in Table III of ref 19 + 0.63. The 3-OCH_3 is a noninculded outlyer in that it is observed to be 0.37 log (1/C) more active in bovine liver DHFR than in the Baker assay!   [b] The $OCNH_2$ is assigned a value of −1.94 as given by Rekker.[28]

Table VII. Compounds Predicted to be of High Activity Using Schemes I and II, Equations B

| | | | | log (1/C) predicted | |
compd	no.	S_0, A^2	$\Sigma\pi$	D_4, A	Scheme I, eq B	Scheme II, eq B
3,4-Et_2	35	32.4	3.10	−0.22	8.26	8.31
3-$CH(CH_3)_2$, 4-Cl	36	33.6	3.13	0.00	8.27	8.30
3-Cl, 4-$CH(CH_3)_2$	37	34.0	3.13	−0.22	8.26	8.28
	38	32.2	1.26	−0.37	7.61	7.72
	39	31.9	1.52	−0.27	7.64	7.62

compounds listed in Table VI. The range of observed activity is 3.69 − log (1/C) units. Unfortunately, no 4-position substituted compounds were available for the test. Also, all compounds, except (no. 28)-[2-CH_3], were taken from a data base where the log (1/C) measurements were made on bovine liver DHFR, and not Walker 256 tumor or L1210 leukemia cells. The method of adjusting the activity scales is described in footnote (a) of Table VI. Table VI also contains observed and predicted activities. The predicted activities generally fall within the standard deviations of the fits (±0.40 to ±0.50). Compound (no. 34)-[3-CN] is an exception. It is predicted to be about ten times less active than observed.

4. Prediction of New Compounds. A major reason for generating structure–activity correlation equations is to suggest new compounds, thus eliminating the costly synthesis of marginal compounds. Table VII contains five suggested compounds which are predicted to have activities greater than 7.6 on the log (1/C) scale. Four of the compounds, (no. 35) to (no. 38), are predicted to have log (1/C) > 8.2. The actual activities of these compounds may be higher than predicted because the correlation equations tend to underestimate activity on the high end of the activity scale.

The first three compounds reported in Table VII (no. 35, 36, and 37) are simple congeneric derivatives to (I) (see Figure 2) which should have reasonable activity. The last two compounds (no. 38 and 39) represent *noncongeneric* analogues to (I) which were deduced on the basis of shape similarity as measured by S_0. Compound (no. 38) is an especially attractive candidate since it represents a new structure on which considerable chemistry can be done on the fused ring. Compound (no. 39) has the disadvantage of having three asymmetric substitution carbon centers.

5. General Discussion. The desire to assign a physicochemical interpretation to Schemes I–IV must be tempered by their empirical nature.

It does seem clear that nonpolar substituents at the 3 and 4 positions enhance activity. This suggests, in turn, that these substituents may be involved in stabilizing van der Waals interactions with nonpolar groups of the DHFR at the receptor site. Silipo and Hansch[18] find that the MR values of the 4 substituents enhance their correlation equation. They assign MR to be a measure of polarity which is at odds with our model. It is of interest in this regard to note that the Baker compounds, [3Cl,4OCH$_2$C$_6$H$_4$-4′SO$_2$F] and [3Cl,4(CH$_2$)$_2$C$_6$H$_4$-4′SO$_2$F], are very likely to be about identical in shape based upon compounds (no. 2) and (no. 8) of Table III. The compound with the more nonpolar 4 substituent, having the (−CH$_2$−)$_2$ group, has a $\Sigma\pi$ about 0.96 units larger than the other compound. It is also more active by 0.27 log (1/C) units than the (−OCH$_2$−) containing compound. Scheme I would predict the more nonpolar compound to be more active by 0.42 log (1/C) units. Both Silipo and Hansch and this work conclude that nonpolar substituents at the 3 position enhance activity.

The function involving D_4, devised to incorporate the decreased activity of the 4-CN and 4-C$_6$H$_5$ substituted Baker triazines, is based upon a destabilizing steric interaction with some group(s) at the binding site of DHFR. The ΔD_4 (D_4) term may be an equivalent representation of the "semirigid area" between the 3 and 4 positions of the substituted benzyl ring postulated in the Silipo and Hansch structure–activity analysis of the Baker triazines.

Lastly with regard to the Silipo and Hansch study, we have constructed the observed vs. predicted activity correlation equation, using their reported activity predictions.[18] The equation is

$$\log (1/C)_{obsd} = 0.976 \log (1/C)_{pred} + 0.231 \qquad (14)$$

$$N = 24 \qquad R = 0.876 \qquad S = 0.73$$

Compounds (no. 10), (no. 15), and (no. 16) of Table III were not considered by Silipo and Hansch. Hence, they were not included in generating eq 14. Clearly, the Hansch and Silipo structure–activity equation must be judged to be inferior to Schemes I–IV. It should be noted, however, that Silipo and Hansch did not include certain "outlayer" compounds in generating their QSAR. These compounds include (no. 1), (no. 14), and (no. 18) of Table III, among others listed in the original data base.[8]

The S_0 terms in Schemes I–IV must be viewed very carefully. The optimum values of S_0 and, correspondingly, V_0, have no physical meaning. The S_0 represent *relative* numerical scales which reflect the need to have an analogue adopt the $\theta = 310°$ conformer state such that the 3 and/or 4 substituents possess size and/or conformational freedom so that specific spaces are occupied.

From this single shape analysis study it is not at all clear how to select the particular reference compound from the data base, and/or shape descriptor, which will optimize the significance of a correlation. Probably all pairwise shape descriptors and reference compounds will need to be tested as was done here. However, since one is usually interested in optimizing activity, the reference

compound and/or pairwise shape descriptor which best fits the high end of the activity data base should probably be chosen. Also, those reference compounds having diverse multiple group substitutions, e.g., $X_3 \neq X_4$ and both $\neq$ H for the triazines, might be preferable. Such compounds cover greater structural diversity and might extend the utility of a correlation with respect to the possible range of group substitutions.

Lastly we conclude this report by pointing out the significance of the work reported here:

Three-dimensional molecular shape descriptors have been defined for a set of Baker triazines which are DHFR inhibitors. A maximum of two shape descriptors, and one physicochemical feature, are needed to explain the variance in DHFR inhibition activity. No compounds in the data base had to be eliminated in the QSAR analysis as outlyers. The QSAR constructed with the use of molecular shape descriptors is superior to that of Silipo and Hansch[18] which is based solely upon physicochemical and substructural features. Hence, our study demonstrates a clear instance where the systematic consideration of three-dimensional molecular geometry is essential to describing biological drug potency.

Note Added in Proof: On January 15, 1980 part of this work was presented for the first time in a seminar at Warner-Lambert Company, Ann Arbor, MI. This author reported the predicted activity of compound 38 (log $(1/C)$ = 7.61 and 7.72, Table VII). Dr. M. L. Black, of Warner-Lambert, was in attendance and noted that compound 38 had been synthesized by Warner-Lambert but never tested for DHFR activity. On February 6, 1980 a sample of compound 38 was sent to this author by Dr. Black. I immediately forwarded the sample to Professor C. Hansch at Pomona College for testing. The observed activity according to Professor Hansch and colleagues, based upon the correction for enzyme species as noted in Table VI, is 7.36. A second QSAR study of bovine and rat liver DHFR inhibition by the 3-position substituted 2,4-diaminotriazines, reported in ref 19, is complete and the results are to be published in *Arch. Biochem. Biophys.* The predicted activity for bovine liver DHFR is 6.58 while the observed value is 6.73 for compound no. 38. These successful predictions give additional support for the reliability of the QSAR reported in this paper and suggest that the QSAR might be used to develop new lead compounds.

Photoassisted Water–Gas Shift Reaction over Platinized TiO2 Catalysts

S. Sato and J. M. White*

*Contribution from the Department of Chemistry, University of Texas, Austin, Texas 78712.
Received April 24, 1980*

Abstract: The reaction of gas-phase water with CO (water–gas shift reaction) over platinized, powdered TiO_2 is found to be catalytic under UV illumination. The reaction kinetics has been studied at temperatures from 0 to 60 °C. At 25 °C, the reaction is zero order both in CO and H_2O when $p_{CO} \geq 0.3$ torr and $p_{H_2O} \geq 5$ torr and the activation energy is 7.5 kcal/mol. The wavelength dependence of the reaction rate shows a cut-off near, but slightly below, the band gap of TiO_2. The quantum efficiency of the reaction is about 0.5% at 25 °C. A mechanism is proposed which involves the photodecomposition of H_2O over platinized TiO_2.

1. Introduction

Heterogeneous photocatalysis is a relatively new branch of catalysis and in almost all cases semiconductors are used because band-gap photon absorption leads to spatially separated electrons and holes which can be used in oxidation–reduction reactions.[1] Metallized semiconductors have recently appeared which are more effective in some cases than semiconductors alone. For example, Bulatov and Khidekel[2] have reported the photodecomposition of acidified water by using platinized TiO_2 (Pt/TiO_2). Kraeutler and Bard[3] have reported the selective decomposition of liquid acetic acid to methane over illuminated Pt/TiO_2. In the gas phase, Hemminger et al.[4] have claimed the photocatalytic production of methane from CO_2 and gaseous H_2O using a $SrTiO_3$ crystal contacted with a Pt foil.

We have recently begun to investigate the photocatalytic properties of TiO_2 and Pt/TiO_2 and have found that UV-illuminated Pt/TiO_2, but not TiO_2 alone, catalytically decomposes liquid H_2O.[5] Platinized TiO_2 also photocatalyzes the reactions of gaseous H_2O with hydrocarbons,[6] active carbon,[6] and lignite[7] to produce H_2 and CO_2, all of which are thermodynamically unfavorable. We report here another photoassisted reaction over Pt/TiO_2, the reaction of gaseous H_2O with CO (water–gas shift reaction). Although this reaction is not thermodynamically uphill

(1) See e.g. M. Formenti and S. J. Teichner, *Catalysis (London)* 2, 87 (1979).

(2) A. V. Bulatov and M. L. Khidekel, *Izv. Akad. Nauk SSSR, Ser. Khim.*, 1902 (1976).

(3) B. Kraeutler and A. J. Bard, *J. Am. Chem. Soc.*, 100, 2239 (1978).

(4) J. C. Heminger, R. Carr, and G. A. Somorjai, *Chem. Phys. Lett.*, 57, 100 (1978).

(5) S. Sato and J. M. White, *Chem. Phys. Lett.*, in press.

(6) S. Sato and J. M. White, *Chem. Phys. Lett.*, 70, 131 (1980).

(7) S. Sato and J. M. White, *Ind. Eng. Chem. Prod. Res. Dev.*, submitted for publication.