MOLECULAR PHARMACOLOGY, 21:187–195

Ames Test and Antitumor Activity of 1-(X-Phenyl)-3,3-dialkyltriazenes

Quantitative Structure-Activity Studies Based upon Molecular Shape Analysis

A. J. HOPFINGER[1] AND R. POTENZONE, JR.[2]

Department of Macromolecular Science, Case Institute of Technology, Case Western Reserve University, Cleveland, Ohio 44106, and Management Information and Data Systems Division, United States Environmental Protection Agency, Washington, D.C. 20460

Received April 14, 1981; Accepted September 17, 1981

SUMMARY

A quantitative structure-activity relationship (QSAR) based upon molecular shape analysis has been developed for a set of 18 1-(X-phenyl)-3,3-dialkyltriazenes for which mutagenic potency is reported. An "active" molecular shape is proposed which suggests, in turn, a self-consistent mechanism for the microsomal hydroxylation of the nitrogen methyl group. A second QSAR has been constructed for an alternate set of 24 1-(X-phenyl)-3,3-dialkyltriazenes for which antitumor potency is reported. The QSAR suggests that the "active shape for antitumor potency is the same as that hypothesized for Ames mutagenic potency. A QSAR for a therapeutic index (TI) has been constructed to allow optimization of the difference between antitumor and mutagenic potencies. Large TIs require substituents on position 3 of the phenyl ring which are hydrophilic. The antitumor QSAR has been applied to three known compounds to test its reliability. A compound is predicted that is expected to have high antitumor activity and low mutagenicity. An overview of the methodology of molecular shape analysis, and its limitations, is included as part of this report on the development of QSARs.

INTRODUCTION

Recently Venger *et al.* (1) have reported a QSAR[3] of the mutagenicity of a set of 1-(X-phenyl)-3,3-dialkyltriazenes (Structure I) reported in Table 1.

$$\text{Log}(1/C) = 1.09[\log P] - 1.63[\sigma^+] + 5.58 \qquad (1)$$

$$N = 17, \qquad R = 0.974, \qquad S = 0.315$$

$$\text{Average error in prediction} = 7.5\%$$

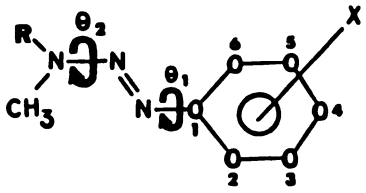
where P is the water/octanol partition coefficient and σ^+ is a measure of the resonance electron-withdrawing power of the X-substituent. Mutagenic activity was measured in terms of the molar concentration, C, of triazene producing 30 mutations/10^8 bacteria above background based upon the Ames test. *Salmonella typhimurium* strain TA92 bacteria were used in the assays.

This work was supported by the National Cancer Institute (Contract N01:CP-76927), the National Science Foundation (Grant ENV 77-74-61), and the National Institutes of Health (Contract 217041).

[1] Case Western Reserve University. Alternate address, Drug Design Group, Searle Research and Development, G. D. Searle and Company, Chicago, Ill. 60680.

[2] United States Environmental Protection Agency.

[3] The abbreviations used are: QSAR, quantitative structure-activity relationship; MSA, molecular shape analysis; TI, therapeutic index; CSP, common spatial pattern; CRSS, common reference shape structure; BM, bond moment.



STRUCTURE I

The phenyltriazenes also exhibit anticancer activity. In particular, Hatheway *et al.* (2) have reported a QSAR study of 61 1-(X-phenyl)-3,3-dialkyltriazenes based upon antitumor potency against L1210 leukemia in mice. Their correlation equation is

$$\log(1/C_{\text{tum}}) = 0.100[\log P] - 0.42[\log P]^2 - 0.312[\sigma^+] - 0.178[MR\text{-}2,6] + 0.391[E_s - R] + 4.124 \qquad (2)$$

$$N = 61, \qquad R = 0.836, \qquad S = 0.191$$

$$\log(P_0) = 1.18 \text{ (activity maximized)}$$

$$\text{Average error in prediction} = 13.5\%$$

C_{tum} is the moles per kilogram producing a $t/c = 140$ (t = life-span of test animal inoculated with leukemia and treated with drug, c = life-span of control animal inoculated intraperitoneally, but receiving no drug). MR-2,6 refers to the sum of the molar refractivities of the substituents flanking the triazene side chain, E_s is the Taft

0026-895X/82/010187-09$02.00/0

steric parameter for R, and σ^+ and P are the same as defined in Eq. 1.

In a subsequent paper (3) these investigators (Hansch and co-workers) concluded that there was no means for the synthesis of potent antitumor 1-(X-phenyl)-3,3-dialkyltriazenes, which are low in toxicity because a QSAR for toxicity in mice is close in correspondence to Eq. 2. However, Hansch and co-workers (2) reversed this position on the basis of Eq. 1 developed to explain the Ames mutagenicity of the 1-(X-phenyl)-3,3-dialkyltriazenes. They qualitatively illustrated how mutagenicity (and presumably carcinogenicity) can be minimized with little loss in antitumor potency. However, a quantitative scheme to achieve this end was not given. This may in part have been due to the fact that the mutagenicity and antitumor QSARs, respectively, contain different physicochemical descriptors and, consequently, there is a limited common ground for comparison.

Equation 1 is based upon excluding Compound 1 (X = 4-$CONH_2$, R = t-Bu) of Table 1 from the analysis. The inclusion of this compound into the QSAR reduces the correlation coefficient R to 0.875 and raises the standard deviation S to 0.611. This is because Eq. 1 predicts Compound 1 to be 2.43 $\log(1/C)$ units more active than observed. Venger *et al.* (2) have suggested that the inactivity of Compound 1 is due to "steric problems." In addition, Eq. 2 is not a particularly reliable relationship. Only 70% of the variance in the activity is explained, and three compounds in the original data base had to be neglected in order to achieve even this limited fit. These observations, coupled with the success of MSA in developing QSARs for four sets of dihydrofolate reductase inhibitors (4–7), prompted us to apply MSA to the triazenes. This is also the first application of MSA to data sets in which both the mechanism and the site of action are not known. Thus, the deduction of an "active" shape, or conformer, as has proven possible in the previous applications, could have particular value in clarifying mechanisms of action. An attempt has also been made to develop a quantitative TI to optimize the difference between antitumor potency and Ames mutagenicity.

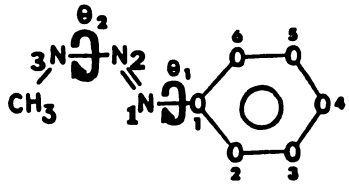
METHODS AND RESULTS

Ames mutagenicity QSAR. This is now our fifth QSAR study based upon MSA. Some generalizations concerning the methodology have emerged and are reported as part of the details of the triazene mutagenicity analysis.

In MSA only free-space intramolecular conformational analyses are considered. Conformational energetics are computed using molecular mechanics potentials according to the parameterization of Hopfinger (8, 9). The calculations are carried out using the CHEMLAB system (10, 11). CHEMLAB is the new name for an expanded and more flexible version of CAMSEQ-II (10). The methods and parameters used in a conformational analysis employing CHEMLAB are identical with those in CAMSEQ-II.

The fixed-valence geometry approximation is employed in the conformational analysis so that molecular shape and conformation are utlimately a function of only torsional bond rotations. The geometry of the phenyltriazene unit, O—N = N—N, was taken from the crystal coordinates of C_6H_5N = NNC_6H_3(2′,4′-Br_2) (12). The remaining valence bond geometries of the triazenes (R, X, CH_3) were generated using "standard" bondlengths and angles (8) via the structure editor in CHEMLAB.

The first step in MSA is to identify the parent, or common, structural unit in the compounds under consideration. For the 1-(X-phenyl)-3,3-dialkyltriazenes, this common structural component is represented in Structure II:



STRUCTURE II

It is next necessary to determine which compounds in the data set will be included in the QSAR analysis. In this study all compounds in Table 1 here and in ref. 1 are considered. However, in both the Baker triazine (4) and the quinazoline (6) MSA-QSAR studies it was necessary to select subsets of each of the compound data bases. There are simply too many compounds in each data base for performing practical conformational analyses. The selection of a subset for QSAR analysis is arrived at by dividing the total data base into classes based upon unique sets of torsional bond rotations (degrees of conformational freedom). Members of each conformational class, spanning the largest and most uniform range in activity for each class, are selected for membership in the QSAR data base. The reasonableness of this approach is discussed and tested in ref. 4 and 6.

The third step in a MSA-QSAR study is to identify the conformational degrees of freedom, torsional rotations about bonds, inherent to the parent structure. The principal bond rotations are θ_1 and θ_2 (see Structures I and II). Rotation of the methyl group is frozen so that one proton is *trans* to N_2. The —N_1N_2— bond is *trans* planar as in the crystal structure (12). This bond also has a large double-bond character so that rotation about it is energetically unlikely. However, a *cis* configuration about —N_1N_2— is also a minimum energy state which needs to be considered.

The next step is the identification of the "active" conformation of the parent structure. This is critical to the practical implementation of MSA. Implicit to this step is the acceptance of the concept that a particular conformation (shape) is necessary for high activity. Conversely, the less stable this "active" conformation, the less active is the compound. To facilitate the methodology, we make the major assumption that one of the free-space intramolecular energy minima is either the "active" conformer of the parent structure, or very close to it. Thus, only minimum energy conformers (with a resolution of ±5° in each of the torsional angles) are considered. A constraint, employed whenever possible, is to select compounds having major differences in activity, but all physicochemical features, except potentially shape, nearly equal. This enhances the chances of unambiguously identifying the active conformer of the parent. In the present case we first carried out a conformational

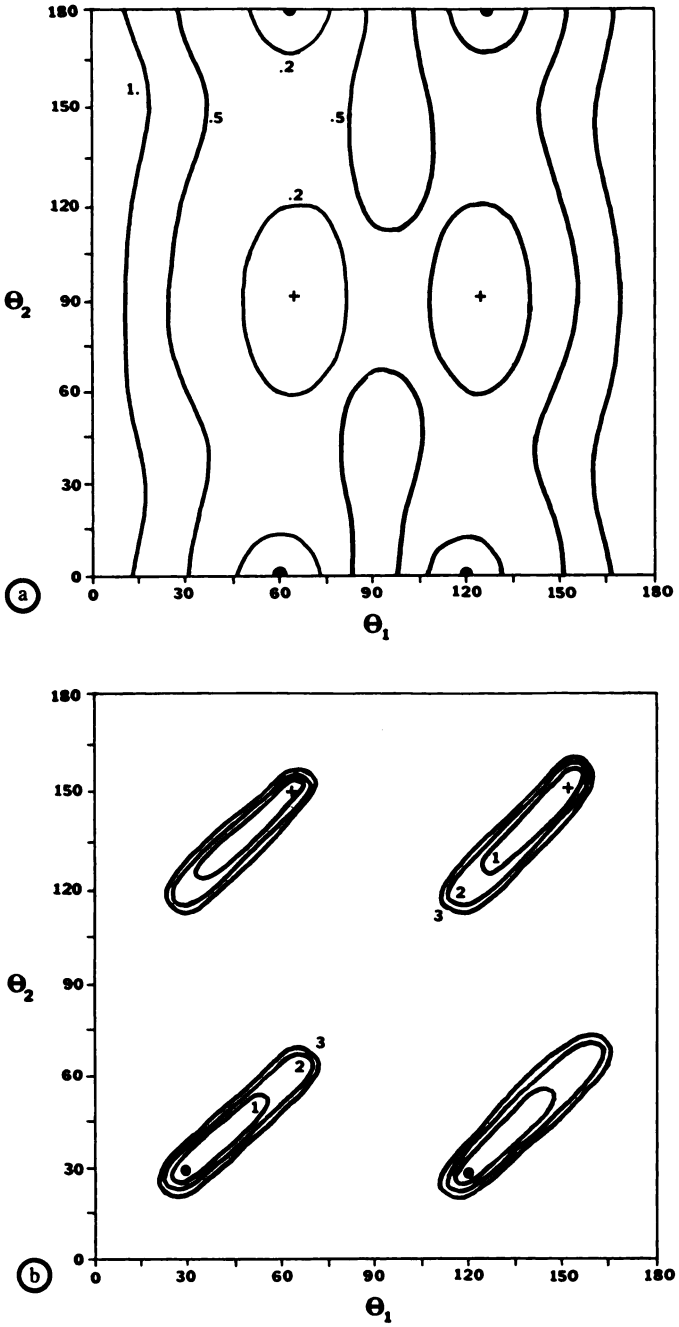


FIG. 1. *Conformational energy maps for Compound 17 of Table 1 (R = CH_3, X = 4-CH_3)*

a, *Trans* C_1-N_1 = N_2-N_3; b, *cis* C_1-N_1 = N_2-N_3. Energy contours are in units of kilocalories per mole relative to the global energy minimum denoted by +. Secondary minima are denoted by ●.

analysis of Compound 17 (R = CH_3, X = 4—CH_3) in Table 1. Conformational flexibility is mainly a function of θ_1 and θ_2, and the corresponding energy maps for (a) the *trans*-N_1N_2-isomer, and (b) the *cis*-N_1N_2-isomer are shown in Fig. 1. Both of the rotations are symmetrical with respect to 180°, and, therefore, only quarter-energy maps are reported. The energy was minimized for rotation of the 4-CH_3 at each θ_1 and θ_2, even though the *methyl* group has virtually no influence on θ_1 or θ_2. The next compound considered was Compound 1 (R = *t*-but, X = 4-$CONH_2$) in Table 1. This compound has a water/octanol partition coefficient, P, and an electron-withdrawing parameters, σ^+, respectively, very nearly the same as Compound 17; see Table 1. The values of log P and σ^+ determined and used by Venger *et al.* (1) were also employed in our work reported here. Venger *et al.* (1) demonstrated by Eq. 1 that these two physiochemical properties correlate highly with mutagenic activity. However, Compound 17 is observed to be 3.17 log(1/C) units more active than Compound 1. The θ_1 versus θ_2 conformational energy maps for Compound 1 with (a) the *trans*-N_1N_2-isomer, and (b) the *cis*-N_1N_2-isomer are shown in Fig. 2. In each case the energy for each (θ_1, θ_2) conformer has been minimized with respect to rotation for the *t*-butyl and 4-$CONH_2$ groups.

A comparison of the *cis*-N_1N_2-energy maps in Figs. 1 and 2 for Compounds 17 and 1, respectively, indicate no differences in conformational preference for these two molecules. However, a comparison of the respective *trans*-N_1N_2-isomers of these compounds indicates that the shallow degenerate conformer energy minima for Compound 17 at θ_1 = 60° (120°) and θ_2 = 180° are destroyed in Compound 1. All that remains is the structurally degenerate global energy minimum, common to both compounds, located at θ_1 = 60° (120°) and θ_2 = 90°. We therefore postulate that the θ_1 = 60°, θ_2 = 180° or θ_1 = 120°, θ_2 = 180° conformer is, or is close to, the "active" conformer. For a frame of reference, θ_1 = 0° corresponds to C_2 and N_2 being *cis* planar, and θ_2 = 180° corresponds to the *t*-butyl being *cis* to N_1.

We must be quite cautious as to the active conformer state of the parent structure because θ_1 is nearly free to rotate in both compounds. The major energy destabilization comes from θ_2. We must also retain both θ_1 = 60° and 120° as possible active conformers because we have not yet considered asymmetrical X substituents in positions 2, 3, 5, and 6.

In order to resolve the θ_1 dilemma, we treat it as equivalent to a substituent degree of freedom. The general procedure we have adopted to handle flexible substituents is as follows: Freeze the parent structure into the active conformer. In this case θ_2 is set to 180° and we let θ_1 go from 0° to 180° at 30° increments. For each value of θ_1 the relative energy minima due to torsional rotations in flexible R and X substituents are determined. For those compounds possessing substituents with amide bonds, both the *trans* and *cis* isomers have been considered. All of these conformers of each compound are individually considered as representative shape states in the MSA.

The final step prior to constructing the QSAR is carrying out the MSA. This currently consists of computing the common overlap steric volume, V_0, between *all pairs* of conformers of all compounds in the MSA data base. This overlap volume computation is made unique by spatially superimposing a set of atoms (the CSP) common to all compounds being analyzed, and orientationally independent of all substituent bond rotations. For the triazenes this CSP is the phenyl ring. Van der Waals spheres, centered at atomic positions, constitute the basis for a molecular steric volume. The mathematics associated with the common overlap volume calculation is given in ref. 4. Reference 4 also contains a color, space-

TABLE 1

Parameters used in the formulation of the QSARs for the mutagenic activity of X—C_6H_4N=$NN(CH_3)R$

Compound no.	X	R	Observed	Log 1/C				S_0[a]	BM[b]	Log P[c]	σ^+[c]
				Calculated (Eq. 3)	Difference (Eq. 3)	Calculated (Eq. 4)	Difference (Eq. 4)				
1	4-$CONH_2$	*t*-Bu	3.83	3.85	0.02	3.92	0.09	39.01	−0.0115	2.61	−0.30
2	3,5-CN	CH_3	3.46	3.50	0.04	3.19	−0.27	44.20	−0.0289	2.18	1.12
3	4-SO_2NH_2	CH_3	3.49	3.35	−0.14	3.26	−0.23	45.82	−0.0242	0.98	0.57
4	3-$CONH_2$	CH_3	3.51	3.69	0.18	4.01	0.50	47.69	−0.0188	1.21	0.28
5	4-$CONH_2$	CH_3	4.04	3.75	−0.29	4.24	0.20	43.44	−0.0199	1.20	0.36
6	4-$CONH_2$	allyl	4.16	4.80	0.64	4.66	0.50	44.73	−0.0225	2.09	0.36
7	3-$NHCONH_2$	CH_3	4.19	4.05	−0.14	3.82	−0.37	48.35	−0.0181	1.29	−0.03
8	4-CN	CH_3	4.43	4.46	0.03	4.92	0.49	43.85	−0.0219	2.39	0.66
9	4-$COCH_3$	CH_3	4.47	4.63	0.16	5.40	0.93	43.92	−0.0195	2.27	0.50
10	H	CH_3	5.32	5.58	0.26	5.33	0.01	42.95	−0.0195	2.59	0.00
11	4-$CONH_2$	*n*-Bu	5.41	5.12	−0.29	6.07	0.66	44.28	−0.0176	2.46	0.36
12	4-$NHCONH_2$	CH_3	5.59	5.84	0.25	5.08	−0.51	43.23	−0.0161	1.25	−0.84
13	4-$NHCONH_2$	CH_3	5.83	5.70	−0.13	5.25	−0.58	43.23	−0.0165	1.54	−0.60
14	4-CF_3	CH_3	5.99	5.86	−0.13	5.96	−0.03	44.14	−0.0226	3.70	0.61
15	3-CH_3	CH_3	6.44	6.31	−0.13	5.81	−0.63	45.18	−0.0204	2.85	−0.07
16	4-Cl	CH_3	6.48	6.32	−0.16	5.89	−0.59	43.80	−0.0212	3.33	0.11
17	4-CH_3	CH_3	7.00	6.58	−0.42	6.72	−0.28	43.40	−0.0156	2.93	−0.31
18	4-C_6H_5	CH_3	7.67	7.91	0.24	7.76	0.09	44.10	−0.0174	4.40	−0.18

[a] CRSS = Compound 7, *trans* amide bond with $\theta_1 = 0°$; see Structure I in text.
[b] BM = $Q(C_1)Q(N_1)$ where Q(X) is the residual charge density on atom X computed by CNDO/2; ref. 13.
[c] Taken from Table 1 of ref. 1.

filling computer-graphics example of how overlap volumes are established and defined. Additional comments on overlap volume shape descriptors are given in ref. 6.

In order to compare shape similarity for more than two structures, all structures under analysis must be compared with a CRSS. The determination of the CRSS is made by considering all structures in the data base as possible CRSSs. That compound conformer which maximizes the statistical significance of the MSA-dependent QSAR is selected as the CRSS.

Multidimensional linear regression analysis is used to establish the QSAR. V_0 and two of its dimensional powers

$$S_0 = V_0^{2/3} \quad \text{and} \quad L_0 = V_0^{1/3}$$

are used as molecular shape descriptors. There is no biochemical justification to relate measures of biological activity to V_0, $S_0(V_0^{2/3})$, or $L_0(V_0^{1/3})$. In our first MSA study (4) the best correlation equations were empirically found for V_0 powers in the range of 0.3–1.0. This also turned out to be true in the studies reported in refs. 5–7. V_0, S_0, and L_0 represent a uniform range of values over this preferred V_0 power interval. The dimensions of these shapes descriptors conveniently correspond to volume, area, and length, respectively. However, the dimensions of these descriptors have no physical significance.

In the first part of this study a molecular shape term was added to the log *P* and σ^+ of Table 1 in order to construct a QSAR. The optimal correlation equation using *all* 18 compounds of Table 1 is:

$$\log(1/C) = 6.198[S_0] - 0.0686[S_0]^2 + 0.966[\log P] - 1.753[\sigma^+] - 136.66 \quad (3)$$

$$N = 18, \quad R = 0.981, \quad S = 0.245$$

S_0(optimum) = 45.17 (activity is maximized)

CRSS = Compound 7, 3-$NHCONH_2$ with the amide bond *trans* and $\theta_1 = 0°$. It should be noted that Eq. 3 and Eqs. 4 and 5 which follow each contain a high number of terms relative to the number of observations (compounds) used in the respective derivation. There is the possibility that such relationships may not be statistically valid. However, these relationships make physical sense, so such a possibility is likely remote. The average error in prediction of Eq. 3, which we view as the most straightforward measure of the significance of a QSAR, is 6.4%. The QSAR drops in significance as θ_1 goes from 0° to 90° and then again increases in quality from 90° to 180°. This, taken with the observations concerning θ_2 and activity, leads us to hypothesize that the conformer of the phenyl-dialkyltriazenes shown in Fig. 3a is the active state. The conformer shown in Fig. 3b is speculated to be the least active state based upon the regression analyses. The conformer of each compound which is least active has $\theta_1 = 90°$, $\theta_2 = 90°$ and is close in both intramolecular energy and shape ($\theta_1 = 60°$, 120° and $\theta_2 = 90°$) to the global energy minimum of all compounds in Table 1. The active conformer is characterized by the O—N_1 = N_2—N_3 being *trans* coplanar such that a methyl group must be *cis* planar to N_1.

Since σ^+ is, in most cases, a measure of change in electron density, it occurred to us that calculated residual charge densities might also be activity correlates. The CNDO/2 charge densities (13) of each of the nitrogen and ring carbon atoms were individually considered as replacements for σ^+ in the correlation analyses. No regression equation could be found having a correlation coefficient greater than 0.88. Next, the total net charge density on the phenyl ring carbons was computed for each compound. However, the regression equation resulting from using this feature in place of σ^+ has a correlation coefficient of only 0.875. Last, we considered the product of charges $Q(C_1)Q(N_1)$ and $Q(N_1)Q(N_2)$ as correlation descriptors. Since the bond lengths C_1N_1 and

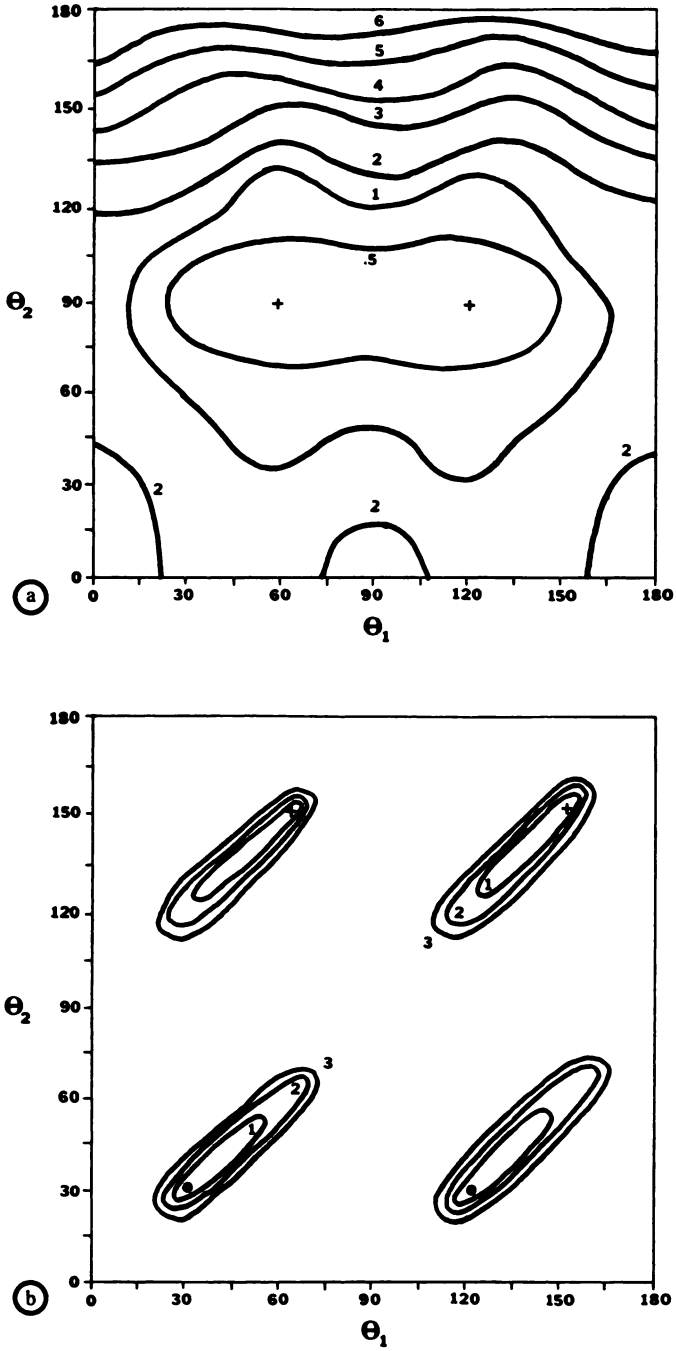


FIG. 2. *Same as Fig. 1 but for compound 1 of Table 1 (R = t-butyl, X = 4-CONH₂)*

N_1N_2 are held constant in the analyses, the corresponding charge products are relative measures of the BM of these two bonds. A significant QSAR was found when $Q(C_1)Q(N_1)$ replaced σ^+. The QSAR is

$$\log(1/C) = 9.649[S_0] - 0.1072[S_0]^2 + 0.8654[\log P] + 232.56[Q(C_1)Q(N_1)] - 209.15 \quad (4)$$

$N = 18, \quad R = 0.932, \quad S = 0.459$

S_0(optimum) = 45.00 (activity maximized)

CRSS = Compound 7, 3-$NHCONH_2$ with the amide bond *trans* and $\theta_1 = 0°$

Average error in prediction = 11.6%

It is important to note that the CNDO/2 method may not provide accurate charge densities in some applications. Thus the selected charge densities may not be good correlation descriptors because they are wrong. However, the fact that a particular product of charge densities is a strong correlating feature tends to negate this possibility.

Equation 4 is certainly not as good as Eq. 3, but the charge density can be determined by calculation as opposed to σ^+, which is experimentally measured in most cases.

The Antitumor QSAR. The values for σ^+ and log *P* given in Table 2 are those reported by Hatheway *et al.* (2). Those compounds in Table 2 which were also studied for mutagenicity (see Table 1) have their mutagenic activities reported. The predicted mutagenic activities are based upon Eq. 3.

The all-planar conformer, having a *trans*-$N_1 = N_2$-geometry and a *cis* $N_1 = N_2—N_3—CH_3$ wad deduced to be the "active" conformer state. This is the same active conformer as found in the Ames mutagenicity QSAR investigation and is shown in Fig. 1 for X = 3-CH_3 and *R* = CH_3. Although the data base in Table 2 contains substituents in positions 2 and/or 6, it was not possible to deduce whether the active conformer requires $\theta_1 = 0°$ (C_2-C_1-N = N, *cis*) or $\theta_1 = 180°$ ($C_2—C_1—N = N$, *trans*).

All 24 compounds and their intramolecular minimum energy conformers were used as CRSSs in the MSA. In addition, the compound found to be the optimal CRSS in the formulation of Eq. 3, *R* = CH_3, X = 3-*trans*$NHCONH_2$, was considered as a possible CRSS.

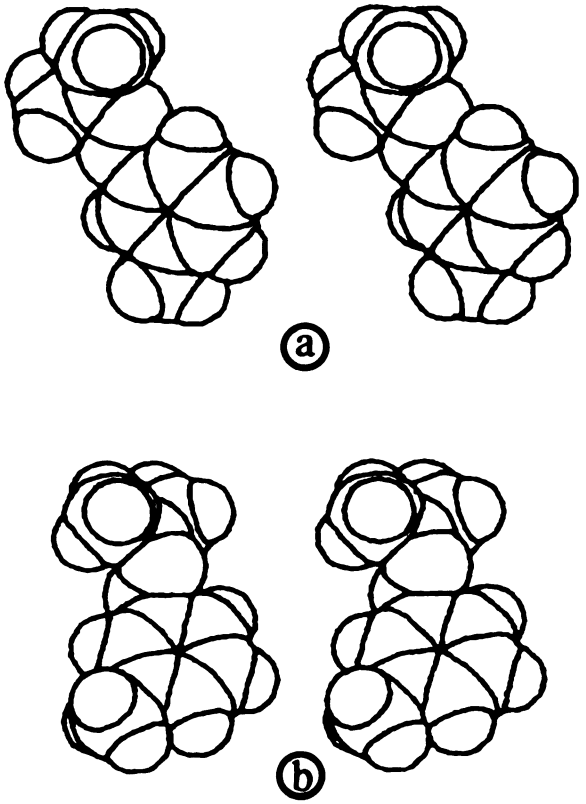


FIG. 3. *Space-filling stereographic representation of Compound 10 of Table 1 (R = CH_3, X = 3-CH_3)*

a, The postulated active conformer, $\theta_1 = 0°$, $\theta_2 = 180°$; b, the postulated least active conformer, $\theta_1 = 90°$, $\theta_2 = 90°$.

TABLE 2

Parameters Used in the Formulation of the QSAR for the Antitumor Activity of $X(n)C_6H_{5-n}N{=}NN(CH_3)R$

Ames mutagenic activity from ref. 1 is also reported.

Compound no.	Compound		Observed		Predicted		Predicted	Observed	S_0[c]	σ^{+}[c]	Log P[c]
	X	R	Antitumor[b]	Ames Mutag.	Antitumor (Eq. 5)	Ames Mut (Eq. 3)	Antitumor	Ames mutation			
1	3,5-CN	CH_3	2.78	3.46	2.93	3.50	0.16	0.04	44.20	1.12	2.18
2	4-$CO_2C_2H_5$	i-C_4H_9	2.80	—	2.79	—	−0.01	—	38.66	0.48	4.34
3	3-$CONH_2$-2,5Cl	CH_3	3.04	—	3.18	—	0.14	—	48.85	0.79	2.42
4	2-$CONH_2$,4,5Cl	CH_3	3.14	—	3.09	—	−0.06	—	41.06	0.59	2.85
5	4-$CONH_2$	i-C_3H_9	3.17	—	3.22	—	0.05	—	38.66	0.36	2.00
6	4-CO_2CH_3	CH_3	3.20	—	3.23	—	0.03	—	43.41	0.49	2.77
7	4-COO^-	CH_3	3.22	—	3.53	—	0.31	—	42.41	−0.02	−1.77
8	2-Cl	CH_3	3.26	—	3.43	—	0.17	—	43.37	0.11	2.97
9	2-$CONH_2$	CH_3	3.31	—	3.42	—	0.11	—	43.39	0.36	1.73
10	4-$CONH_2$	$CH_2CH{=}CH_2$	3.38	4.16	3.42	4.80	0.04	0.64	44.73	0.36	2.09
11	3-CH_3	CH_3	3.40	6.44	3.60	6.31	0.20	−0.13	45.18	−0.07	2.85
12	3-CO_2CH_3	CH_3	3.42	—	3.39	—	−0.03	—	47.53	0.37	2.72
13	4-$CONH_2$	n-C_4H_9	3.47	5.41	3.37	5.12	−0.10	−0.29	44.28	0.36	2.76
14	4-$CONH_2$	CH_3	3.51	4.04	3.45	3.75	−0.06	−0.29	43.44	0.36	1.20
15	H	H	3.60	—	3.50	—	−0.10	—	40.10	0.00	1.94
16	4-SO_2NH_2	CH_3	3.60	3.49	3.38	3.35	−0.22	−0.14	45.82	0.57	0.98
17	2,6-F	CH_3	3.63	—	3.62	—	−0.01	—	43.84	−0.15	2.87
18[a]	2-COO^-	CH_3	3.64	—	3.44	—	−0.20	—	44.05	−0.02	−2.66
19	4-CH_3	CH_3	3.76	7.00	3.69	6.58	−0.07	−0.42	43.40	−0.31	2.93
20	3-$CONH_2$	CH_3	3.80	3.51	3.58	3.69	−0.22	0.18	47.69	0.28	1.21
21	H	CH_3	3.85	5.32	3.54	5.58	−0.31	0.26	42.95	0.00	2.59
22	3-$CONH_2$,6-OCH_3	CH_3	3.95	—	4.07	—	0.12	—	48.03	−0.50	0.44
23	4-$NHCONH_2$	CH_3	3.97	5.59	4.07	5.84	0.10	0.25	43.23	−0.69	1.25
24	4-$NHCOCH_3$	CH_3	4.04	5.83	4.00	5.70	−0.04	−0.13	43.23	−0.60	1.54

[a] An outlier with respect to the QSAR of ref. 2.
[b] From ref. 2.
[c] CRSS; R=CH_3; X=3-*trans* $NHCONH_2$. Taken from Table 1 of ref. 1.

The optimal CRSS is Compound 12, R = CH_3, X = 3-*trans*CO_2CH_3. The corresponding correlation equation has a correlation coefficient of 0.909 and average error in prediction of 11.3%. However, the second-best CRSS is R = CH_3, X = 3-*trans*$NHCONH_2$. We decided to adopt the correlation equation developed for this CRSS because it can be directly compared with Eq. 3; that is, there is a basis for formulating a therapeutic index relationship. The QSAR is

$$\log(1/C_{tum}) = 0.2062[S_0] - 0.0021[S_0]^2 + 0.0243[\log P] - 0.281[\log P]^2 - 0.603[\sigma^+] - 1.39 \quad (5)$$

$N = 24, \quad R = 0.901, \quad S = 0.147$

S_0(optimum) = 49.09 (activity maximized)

Log P(optimum) = 0.43 (activity maximized)

CRSS = [R = CH_3, X = 3-$NHCONH_2$] with the amide bond *trans* and $\theta_1 = 0°$

Average error in prediction = 11.7%

The calculated values of S_0, and predicted activities using Eq. 5, are given as part of Table 2. The absolute values between predicted and observed activities are also listed. Equation 5 represents an improvement over Eq. 2 with respect to statistical fit, inclusion of a previous outlier, and general similarity of form to Eq. 1. Nevertheless, a correlation coefficient of only 0.901 is not ideal. However, this may be a statistical consequence of the narrow range in activity for the data base (2.78–4.04 log(1/C) units). A higher precision is required of S_0, σ^+, and log P to explain the smaller differences in activity in developing Eq. 5 as compared with Eq. 3. Unfortunately, the required added precision cannot be achieved. This may explain why an attempt to use residual charge densities, or bond moments in place of σ^+ in constructing a QSAR, which was successfully done on the mutagenicity data base failed for the antitumor activities. The uncertainty in calculated charge densities (bond moments) is probably sufficiently large so as to diminish significantly any structure-activity fit. This is also the case for the Ames mutagenicity data base in which the alternate QSAR for Eq. 3, based on a bond moment in place of σ^+, has a correlation coefficient of 0.937 as compared with 0.982 for Eq. 3. However, the situation is much worse for the data base in Table 2. No correlation equation involving charge densities or bond moments could be constructed having a correlation coefficient greater than 0.845.

A therapeutic index to maximize the difference between antitumor potency and mutagenicity can be constructed by subtracting Eq. 3 from Eq. 5

$$TI = -5.992[S_0] + 0.0665[S_0]^2 - 0.942[\log P] - 0.028[\log P]^2 + 1.150[\sigma^+] + 135.27 \quad (6)$$

A term-by-term analysis of Eqs. 3, 5, and 6 provides insight into the quantitative design of compounds having large TIs. First, it is noted that antitumor activity is maximized for $S_0 = 49.09$, whereas mutagenicity is maximized for $S_0 = 45.17$. Since the CRSS has a 3-substituent, one is led to the conclusion that antitumor potency is intrinsically increased, and mutagenicity decreased, for a given substituent, by placing it at position 3, as opposed to position 4. This is qualitatively illustrated by comparing Compounds 14 (X = 3-$CONH_2$) and 20 (X = 4-$CONH_2$) in Table 2. The values of σ^+ and log P are nearly the same for these two compounds, yet the X = 3-$CONH_2$ compound has twice the antitumor potency and only one-third the mutagenic potency of X = 4-$CONH_2$. Space-filling stereo models of these two compounds are shown in Fig. 4 to illustrate the shape differences. Thus a first criterion to maximize TI is to consider position-3 substituents having molecular volumes similar to an $NHCONH_2$ group.

If we assume that we select such a position-3 substituent so that $S_0 = 49.09$ and log P is at the value to optimize antitumor activity (i.e., 0.43), then Eqs. 3, 5, and 6, respectively, reduce to

$$\text{Log}(1/C_{\text{mut}}) = 2.71 - 1.753[\sigma^+] \quad (7a)$$

$$\text{Log}(1/C_{\text{tum}}) = 3.67 - 0.603[\sigma^+] \quad (7b)$$

$$\text{TI} = 0.96 - 1.150[\sigma^+] \quad (7c)$$

If we again assume X is chosen so that $S_0 = 49.09$ and $\sigma^+ = 0$, then Eqs. 3, 5, and 6, respectively, become

$$\text{Log}(1/C_{\text{mut}}) = 2.29 + 0.966[\log P] \quad (8a)$$

$$\text{Log}(1/C_{\text{tum}}) = 3.67 + 0.0243[\log P] - 0.0281[\log P]^2 \quad (8b)$$

$$\text{TI} = 1.380 - 0.942[\log P] - 0.281[\log P]^2 \quad (8c)$$

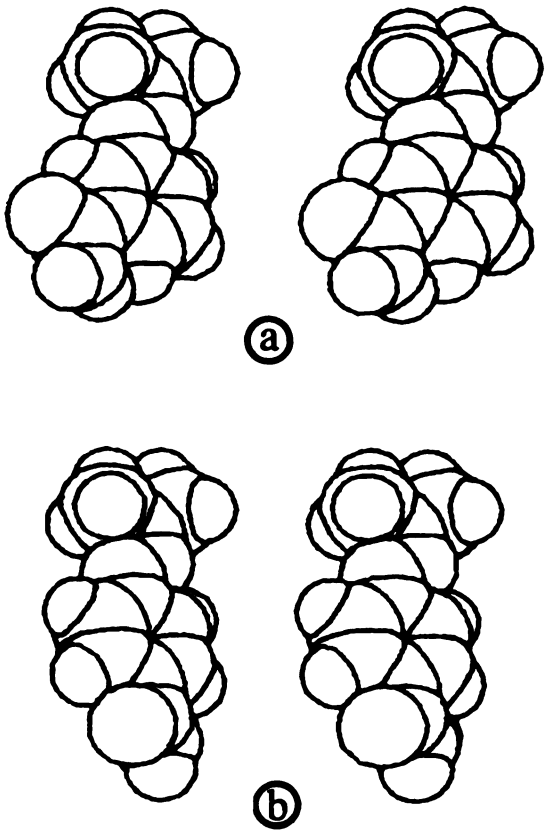


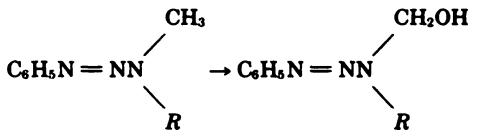
FIG. 4. *Stereographic space-filling models of the active shapes* a, X = 3-$CONH_2$, $R = CH_3$; b, X = 4-$CONH_2$, $R = CH_3$.

Equations 7a–c indicate that mutagenic potency increases nearly 3 times as rapidly as antitumor activity when σ^+ decreases. In contrast, mutagenic activity can be greatly reduced with little loss in antitumor activity by making log P increasingly negative. This is clear from Eqs. 8a–c. Molecular design seems to come down to achieving a lower value in acceptable antitumor activity through σ^+, and tuning in a desired TI through log P. For example, if log $P = -1.5$, $\sigma^+ = -0.5$ and $S_0 = 49.09$, then $\log(1/C_{\text{tum}}) = 3.87$, $\log(1/C_{\text{mut}}) = 1.71$, and TI = 2.16.

Last, the antitumor tumor activities of three compounds in the original data base (2), one of which is an outlier with respect to Eq. 2 taken from ref. 2, were predicted using Eq. 5. The results are given in Table 3. The predicted activities of Compounds 26 and 27 are within the standard deviation limits of Eq. 5. Compound 25, which is an outlier in the QSAR of Hatheway *et al.* (2), is predicted to be about one-half $\log(1/C_{\text{tum}})$ units more active than observed. This is a marginal prediction, but considerably better than that of Eq. 2. The loss in the activity of Compound 25 is predicted by our QSAR to be a result of not being able to assume the active shape with respect to θ_2. The CH_3 group of the N_3 nitrogen atom cannot be *trans* planar to N_1. If N_3 and N_1 are *trans* planar, then there is either an electrostatic repulsion between the oxygen atom of $COCH_3$ and N_1, or an unfavorable steric interaction between the CH_3 of $COCH_3$ and N_1. The choice of destabilization depends upon rotation about the N_3—$COCH_3$ bond.

DISCUSSION

Equations 3 and 4 suggest a mechanism of mutagenic action consistent with the proposed first step of metabolic activation of the 1-(X-phenyl)-3,3-dialkyltriazenes by microsomes (1). This is the hydroxylation of the nitrogen methyl:



Our QSARs suggest that the microsomal substrate might sterically require a *trans* planar $C_6H_5N = NN(CH_3)R$ conformer with a methyl group *trans* planar to the nitrogen bonded to the ring (see Fig. 3a). If the intramolecular energy of this state is high (unstable), the activity is low. The $Q(C_1)Q(N_1)$ term in Eq. 4 might also be related to the methyl hydroxylation. $Q(C_1)Q(N_1)$ is negative because $Q(N_1) < 0$ (electron-rich) and $Q(C_1) > 0$ (electron-poor). Mutagenicity increases as BM increases, that is, as the absolute value of BM goes to zero. Neither $Q(C_1)$

TABLE 3

Predicted antitumor activities of three known compounds using Eq. 5

Compound		Compound no.	Log $(1/C_{tum})$		Predicted −observed	S_0	σ^+	Log P
X	*R*		Observed[a]	Predicted				
4-CH_3	$COCH_3$	25	2.78	3.27	0.49	36.19	−0.31	3.13
2-$CONH_2$	$CH_2CH{=}CH_2$	26	3.43	3.30	−0.13	44.91	0.36	2.62
4-NHCOH	CH_3	27	3.85	3.93	0.08	43.72	−0.60	1.53

[a] Taken from ref. 2.

nor $Q(N_1)$ individually correlates well with activity. Thus one concludes that mutagenic activity is dependent upon the actual dipolar character of the C_1N_1 bond. When the methyl group is *trans* planar to N_1 it is at its maximal distance from the C_1N_1 bond. Equation 4 might be telling us that the hydroxylation reaction is retarded as the C_1N_1 dipole increase; i.e., $Q(N_1)$ becomes more negative and/or $Q(C_1)$ becomes more positive. In other words, hydroxylation is stabilized as the C_1N_1 bond moment approaches zero.

This mechanistic interpretation of the QSAR is completely hypothetical. However, it does explain how mutagenic potency can vary as a function of the X and *R* substituents on the basis of a known step in the metabolic activation of the phenyl-dialkyltriazenes. It also identifies the methyl hydroxylation as the rate-limiting step in mutagenic activation.

The shape term in Eq. 3 is not only responsible for fitting Compound 1 of Table 1 into the QSAR but also somewhat enhances the QSAR for the other 17 compounds. This can be realized from the fact that the log *P* and σ^+ coefficients in Eqs. 1 and 3 are, respectively, about the same. However, the average error in prediction of Eq. 1 is 7.5% whereas that for Eq. 3 is 6.4%. This suggests that the shape term is also responsible for the increased predictive accuracy.

The "active" conformation, deduced by the correlation analyses and shown in Fig. 3a, is not a relative intramolecular energy minimum with respect to θ_1 for any compound in Table 1. This is not the first application of MSA in which the "active" conformer is not found to be an intramolecular minimum energy conformer. However, this might be due to the fact that the conformers of any phenyltriazene in Table 1, corresponding to $\theta_2 = 180°$ and θ_1 variable, differ in intramolecular energy by less than 0.80 kcal/mole, which is probably less than the accuracy of the method of calculation. That is, for $\theta_2 = 180°$, there is nearly free rotation about θ_1. Thus receptor binding interactions could supply the activation energy needed to go from $\theta_1 = 60°$ and/or 120° to 0°.

The QSARs developed here suggest that both mutagenic and antitumor activity require the same "active" shape of the phenyltriazenes with respect to θ_1 and θ_2 (see Structure I). This requires that the methyl group bonded to N_3 be *trans* planar to N_1 and the substituted triazene fragment be coplanar to the phenyl ring. We hypothesize that this active shape is required for the microsomal hydroxylation of the nitrogen methyl group. If we further postulate that mutagenicity and antitumor potency require the same "active" shape with respect to the X substituents, then we have sufficient data to make one generalization regarding conformation. The conformations of flexible X-substituents which yield the most significant QSARs correspond to intramolecular energy minima that are most nearly coplanar to the phenyl ring. In other words, an over-all conclusion regarding the active molecular shape of the phenyltriazenes, for both Ames mutagenicity and antitumor potency, is that the entire compound should be in a coplanar conformation.

Insofar as Ames mutagenicity is a measure of carcinogenicity, we have a quantitative recipe, through Eqs. 6–8, to make "safe," yet potent, antitumor agents. This recipe calls for 3-X substituents which have a σ^+ value small enough to achieve an acceptable level of antitumor activity and a negative log *P* value selected to achieve a particular TI. We have estimated some σ^+ and log *P* and suggest that the compound ($R = CH_3$, X = 3-$NHCONHCH_2COO^-$, 4-CH_3) will have $S_0 = 48.35$, $\sigma^+ = -0.21$, log $P = -2.06$. If these are accurate estimations, then $\log(1/C_{mut}) = 1.02$, $\log(1/C_{tum}) = 3.53$ and TI = 2.51. This compound is suggested as a candidate for synthesis on the basis of its predicted properties.

Equations 3 and 5 demonstrate that the inclusion of 3-dimensional molecular shape features in a QSAR investigation can lead to both improvement in quality and general applicability of a correlation equation as compared with neglecting shape descriptors. Moreover, it is also possible to separate spatial contributions to activity from thermodynamic and/or electronic effects. This is particularly evident in the development of the TI in this paper, where the inclusion of S_0 makes it possible to analyze how σ^+ and log *P* can be selected to optimize antitumor activity for a minimal increase in mutagenic potency. Moreover, MSA-based QSARs have the fringe benefit of making it possible to hypothesize an "active" shape required of a homologous series of compounds.

However, it is also important to point out a limitation of the shape descriptor V_0 and its empirical variants S_0 and L_0. V_0 cannot be used to predict compounds more active than the most active compound in the data base being investigated. The prediction of more potent compounds must come from the non-shape descriptors in the QSAR. This follows from two properties of the common overlap steric volume descriptor. First, V_0 for any CRSS cannot be larger than the actual volume of the CRSS. Second, V_0 cannot be less than the volume of the CSP for any choice of the CRSS. One does not wish to maximize activity by minimizing V_0, which corresponds to using the CSP as the CRSS. This chemically requires deleting all substituents. Rather, one wishes to increase activity by increasing V_0, which requires making key substitutions at key sites. However, there are only two

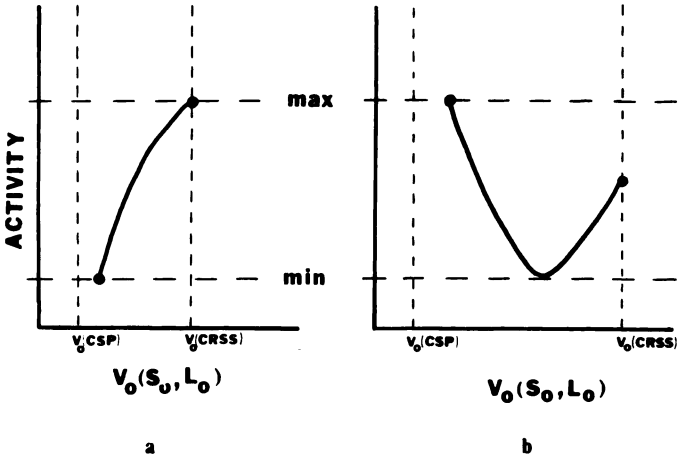


FIG. 5. *The two possible relationships between overlap volume shape descriptor, V_0, and activity for which activity can increase as V_0 increases*

a, The most active compound in the data base is the CRSS. A linear dependence is also possible; b, CRSS has a volume larger than the most active compound and is more active than the least active compound.

ways in which activity can increase directly with V_0 through a linear or parabolic dependence. These are shown in Fig. 5. In Fig. 5a the CRSS is the most active compound. In this case the maximum in activity must correspond to the maximum in V_0. In the second case the CRSS is any compound in the data base that has a volume larger than the most active compound, and an activity greater than the least active compound. Nevertheless, in neither case can we extend the activity versus V_0 curve irrespective of what compound we choose. V_0 in each case is already at its maximum possible value, namely the steric volume of the CRSS. We are forced to stay within the range of descriptor space of V_0 as dictated by the data base and our choice of the CRSS.

A qualitative attempt can be made to enhance activity through shape by studying space-filling stereographic models (like those in Figs. 3 and 4) of very active and very inactive compounds. Differences in spatial occupancy of substituents may be discernible between active and inactive analogs. This in turn could suggest substituents homologous to those of active compounds but slightly larger in size. If the relationship between the shape descriptor and activity is not either of those shown in Fig. 5, one is most likely at an impasse regarding enhancing activity through shape. This is the case for both Eqs. 3 and 5 in this paper. Log$(1/C)$ is maximized at $S_0 = 45.17$ and $S_0 = 45.19$ for Eqs. 3 and 5, respectively. This is within the data base of S_0 values (39.01–48.35). At best, Eqs. 3 and 5 suggest that large substituents in position 3 suppress mutagenic potency.

A final point to reiterate about the shape descriptors is that there is no physicochemical basis for believing that activity should be related to V_0 through integer power relationships. Thus it should not be surprising that S_0 works better than V_0 in some correlation analyses. This is simply an empirical consequence of the mathematical forms of both the shape descriptors and regression equations.

REFERENCES

1. Venger, B. H., C. Hansch, G. J. Hatheway, and Y. V. Amrein. Ames test of 1-(X-phenyl)-3,3-dialkyltriazenes: a quantitative structure activity study. *J. Med. Chem.* **22**:473–477 (1979).
2. Hatheway, G. J., C. Hansch, K. H. Kim, S.. R. Milstein, C. L. Schmidt, R. N. Smith, and F. R. Quinn. Antitumor 1-(X-Aryl)-3,3-dialkytriazenes. 1. Quantitative structure-activity relationships vs. L1210 leukemia in mice. *J. Med. Chem.* **21**:563–572 (1978).
3. Hansch, C., G. J. Hatheway, F. R. Quinn, and N. Greenberg. Antitumor 1-(X-aryl)-3,3-dialkyltriazenes. 2. On the role of correlation analysis in decision making drug modification: toxicity quantitative structure-activity relationships of 1-(X-phenyl)-3-3-dialkyltriazenes in mice. *J. Med. Chem.* **21**:574–578 (1978).
4. Hopfinger, A. J. A QSAR investigation of dihydrofolate reductose inhibition by Baker triazines based upon molecular shape analysis. *J. Am. Chem. Soc.* **120**:7196–7206 (1980).
5. Hopfinger, A. J. A general QSAR for dihydrofolate reductase inhibition by 2,4-diaminotriazines based upon molecular shape analysis. *Arch. Biochem. Biophys.* **206**:153–163 (1981).
6. Battershell, C., D. Malhotra, and A. J. Hopfinger. Inhibition of dihydrofolate reductase: structure activity correlations of 2,4-diaminoquinazolines based upon molecular shape analysis. *J. Med. Chem.* **24**:812–818 (1981).
7. Hopfinger, A. J. Inhibition of dihydrofolate reductase: structure activity correlations of 2,4-diamino 5-benzyl pyrmidines based upon molecular shape analysis. *J. Med. Chem.* **24**:818–822 (1981).
8. Hopfinger, A. J. *Conformational Properties of Macromolecules.* Academic Press, New York (1973).
9. Hopfinger, A. J., and S. K. Tripathy. Molecular structure calculations applied to solid state polymer physics, part I, in *Critical Reviews in Solid State Physics and Materials Sciences* (R. Hoffman and D. Schuele, eds.). CRC Press, West Palm Beach, Fl., 285–371 (1980).
10. Potenzone, R., Jr., E. Cavicchi, H. J. R. Weintraub, and A. J. Hopfinger. Molecular mechanics and the CAMSEQ processor. *Computers Chem.* **1**:187–194 (1977).
11. Hopfinger, A. J. Description and application of the molecular structure calculator CHEMLAB. *Proc. NRCC Workshop on Computer Simulation of Organic and Biological Molecules*, in press.
12. Kondrashev, Y. D. *Zstkai* **15**:517 (Cambridge X-ray Crystal:13113-75-2.01) (1974).
13. Pople, J. A., and D. C. Beveridge. *Approximate Molecular Orbital Theory.* McGraw-Hill, New York (1970).

Send reprint requests to: Dr. A. J. Hopfinger, Drug Design Group, Searle Research and Development, G. D. Searle and Company, Box 5110, Chicago, Ill. 60680.