$$C^s = (A - C_0{}^s\epsilon^p)/(\epsilon^s - \epsilon^p)$$

where $C^s$ is the concentration of the benzofuran at time $T$, $A$ is the measured absorption at time $T$, $C_0{}^s$ is the initial concentration of the benzofuran, $\epsilon^p$ is the extinction coefficient of the product (determined from $A$ measured after >25 half-lives of reaction), and $\epsilon^s$ is the extinction coefficient of the benzofuran. The calculator program also fit the data points to a first-order rate law equation, determined the correlation coefficient of this fit, and calculated the 95% confidence range (t test) of the rate constant (Table I).

**Determination of Antibacterial Activity of 1 and Nitrofurazone.** Cultures of *Staphylococcus aureus*, *Klebsiella pneumonia*, *Streptococcus faecalis*, *Pseudomonas aeruginosa*, *Proteus mirabilis*, and *Proteus rettgeri* were obtained from the culture file of the Department of Microbiology, UTCHS. Cultures of *E. coli* Br and *E. coli* Br 207 were kindly provided by Dr. D. R. McCalla, McMaster University. *E. coli* B was obtained from ATCC. These cultures were maintained on Penassay agar (Difco) at 25°. An overnight culture of the organism in Penassay broth was adjusted to 60% transmittance and 10 ml of this culture used to inoculate 900 ml of Penassay broth in an Oxford pipettor. The inoculated broth (4.5 ml) was added to a solution of the benzofuran in 10% $Me_2SO$ (0.5 ml) and the tubes were incubated at 37°. All determinations were in stationary culture except *Pseudomonas aeruginosa*. For each assay the turbidity of tubes containing concentrations of benzofuran (dilution series 1, 0.75, 0.5, 0.375, 0.25, 0.187, 0.125, 0.062, and 0.031, four replicates each) and appropriate positive and negative controls was determined after 6 hr (660 nm, Spectronic 20). The minimal inhibitory concentration at 6 hr was defined as the lowest concentration which had a % $T$ of greater than 90%. In order to obtain the data plotted in Figure 1, the cultures were monitored at 20-min intervals for 9 hr after inoculation. The log $A^{660}$ vs. time plots for each concentration were extrapolated to $A = 0.05$. The elapsed time between inoculation and the time of initiation of growth (i.e., when $A = 0.05$) was defined as the lag time for the concentration. The lag times so obtained were then plotted against concentration.

**Comparison of Antibacterial Activity of Various Benzofurans in Minimal Media.** An overnight culture of *E. coli* B in Jensen's minimal media was used to inoculate (10 ml) 90 ml of Jensen media in a Nephelo culture flask. This culture was incubated in a shaker–water bath incubator at 37° until the growth had reached 70% $T$. This culture (1 ml) was added to 8 ml of Jensen's broth (pH 7) and 1 ml of benzofuran in 10% $Me_2SO$. The cuvettes were placed in a J-Y Biophotometer and turbidity changes monitored (with agitation). Five concentrations of benzofuran (dilution series 1, 0.67, 0.33, 0.2, and 0.1) and a control culture containing 1% $Me_2SO$ without benzofuran were monitored in each assay. Plots of log $A$ vs. time were used to calculate percent inhibition as a function of concentration. Log concentration vs. probability plots of these values were linear (Figure 1). Slopes of these lines as well as $IC_{50}$ values obtained from the plots were reproducible to within ±5%.

**Acknowledgment.** The author thanks Frederick Petty of the Charles B. Stout Neuroscience Mass Spectrometry Laboratory for obtaining the mass spectral analyses. The author would also like to thank Ms. Frances Byrd and Ms. JoAnn Rodgers for expert technical assistance and Charles Cruze for developing the program used to calculate the solvolysis rate constants.

## References and Notes

(1) K. Miura and H. K. Reckendorf, *Prog. Med. Chem.*, 5, 320 (1967).
(2) M. C. Dodd and W. B. Stillman, *J. Pharmacol. Exp. Ther.*, 82, 11 (1944).
(3) L. I. Kheml'nitskii, T. S. Novikova, S. S. Novikova, K. E. Ovcharov, and N. N. Borisova, *Fiziol. Akt. Veshchestva Ikh Primen. Rastenievod., Dokl. Nauchn. Konf.*, 215–222 (1963); *Chem. Abstr.*, 66, 45676d (1967).
(4) W. Hoyle, G. P. Roberts, and O. Meth-Cohn, *J. Med. Chem.*, 16, 709 (1973).
(5) L. J. Powers and M. P. Mertes, *J. Med. Chem.*, 13, 1102 (1970).
(6) R. Royer, L. Renē, P. Demerseman, R. Cavier, and J. Cēnac, *Chim. Ther.*, 6, 79 (1971).
(7) R. Cavier, J. Cēnac, R. Royer, and L. Renē, *Chim. Ther.*, 7, 361 (1972).
(8) D. Kaminsky, J. Shavel, and R. I. Meltzer, *Tetrahedron Lett.*, 859 (1967).
(9) A. Mooradian, *Tetrahedron Lett.*, 407 (1967).
(10) A. Mooradian and P. E. Dupont, *J. Heterocycl. Chem.*, 4, 441 (1967).
(11) L. C. King, *J. Am. Chem. Soc.*, 66, 894 (1944).
(12) J. A. Krynitsky and H. W. Carhart, "Organic Syntheses", Collect. Vol. IV, Wiley, New York, N.Y., 1963, p 436.
(13) G. W. Perold, *S. Afr. Ind. Chem.*, 74 (1956).
(14) C. Pene, M. Hubert-Harbart, and R. Royer, *Eur. J. Med. Chem.*, 9, 202 (1974).
(15) R. A. Jensen, *Genetics*, 60, 707 (1968).
(16) D. L. Cramer and M. C. Dodd, *J. Bacteriol.*, 51, 293 (1946).
(17) F. Kavanagh in "Analytical Microbiology", Vol. II, F. Kavanagh, Ed., Academic Press, New York, N.Y., 1972, pp 64–73.
(18) W. F. Harrigan and M. E. McCance, "Laboratory Methods in Microbiology", Academic Press, New York, N.Y., 1966, p 21.
(19) D. R. McCalla, A. Reuvers, and C. Kaiser, *J. Bacteriol.*, 104, 1126 (1970).
(20) H. P. Treffers, *J. Bacteriol.*, 72, 108 (1956).
(21) Reference 17, p 85.
(22) T. C. McIlvaine, *J. Biol. Chem.*, 49, 183 (1921).

# Correlation Analysis of Baker's Studies on Enzyme Inhibition. 1. Guanine Deaminase, Xanthine Oxidase, Dihydrofolate Reductase, and Complement†,[1a]

Carlo Silipo[1b] and Corwin Hansch*

*Department of Chemistry, Pomona College, Claremont, California 91711.* *Received May 5, 1975*

Five correlation equations are presented which relate inhibitory activity of 578 inhibitors of guanine deaminase, xanthine oxidase, dihydrofolate reductase, and complement to their chemical structures. The use of correlation analysis in enzyme studies for drug development is discussed. The importance of indicator variables in such studies is emphasized.

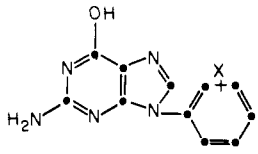
Starting about 1964, the late B. R. Baker and his students published over 100 papers studying the effect of over 1500 inhibitors on various enzymes. It was Baker's view that one should be able to develop more selective and more effective drugs through the selective control of enzymes. Baker demonstrated that one can rather quickly develop extremely potent enzyme inhibitors by more or less sys-

† This paper is dedicated to the memory of Edward Smissman and Bernard R. Baker.

tematic modification of a substrate analog. Along with others,[2] he was able to demonstrate that inhibitors could be designed so that enzyme from two different sources could be inhibited to quite different degrees. It is uncovering this selectivity of inhibitor for a critical isozyme of pathogen without serious inhibition of enzyme from host that offers one of the most promising avenues for future drug research. It behooves us in seeking out these differences to map the enzymes as carefully as possible through activity studies in order to obtain the highest possible inhibitory index. (Inhibitory index = concentration of inhibitor X causing 50% inhibition of host enzyme/concentration of inhibitor X causing 50% inhibition of pathogen enzyme.) To do this properly, even under the best of conditions, will mean the synthesis and testing of very large numbers of increasingly expensive organic compounds. Each molecular probe (inhibitor) must be carefully designed so as to yield the maximum amount of new information. The combined use of regression analysis and cluster analysis[3] offers the best technique for avoiding redundancy in the synthesis of inhibitors. Now that a better understanding of the structure–activity relationship in drug movement[4] and metabolism[5] is beginning to be developed, it will be easier in the future to translate in vitro enzyme studies to in vivo drugs.

It seemed to us that one of the best ways to gain insight into the problems involved in the design of enzyme inhibitors was to undertake a study of Baker's very extensive work. The following quantitative structure–activity relationships (QSAR) were formulated in our initial studies of the results of Baker's group.

1. Inhibition[6] of guanine deaminase by



$$\log 1/C = 0.38(\pi\text{-}4) + 0.69(\text{MR-}3) + 1.10(E_s\text{-}2) + 1.34I + 3.75 \quad (1)$$

| $n$ | $r$ | $s$ |
|---|---|---|
| 32 | 0.928 | 0.363 |

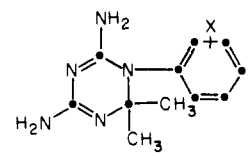
In this and subsequent equations, $C$ is the molar concentration of inhibitor producing 50% inhibition of enzyme and $\pi$, MR, and $E_s$ are the hydrophobic, molar refractivity, and Taft steric constants, respectively, of X. *The numbers attached to these parameters refer to the position of X on the phenyl ring*, $n$ is the number of data points upon which the correlation equation is based, $r$ is the correlation coefficient, and $s$ is the standard deviation. The indicator variable $I$ indicates the presence of a 4-OR function.

2. Inhibition[7] of xanthine oxidase by the same type of inhibitors used with guanine deaminase

$$\log 1/C = 0.20(\text{MR-}3,4) + 1.26(E_s\text{-}2) + 0.43(E_s\text{-}4) + 4.33 \quad (2)$$

| $n$ | $r$ | $s$ |
|---|---|---|
| 30 | 0.924 | 0.228 |

3. Inhibition[8] of dihydrofolate reductase by

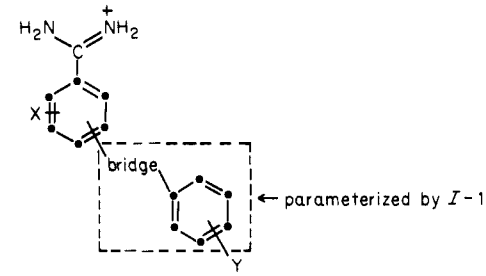


$$\log 1/C = 0.68(\pi\text{-}3) - 0.12(\pi\text{-}3)^2 + 0.23(\text{MR-}4) - 0.024(\text{MR-}4)^2 + 0.24(I\text{-}1) - 2.53(I\text{-}2) - 1.99(I\text{-}3) + 0.88(I\text{-}4) + 0.69(I\text{-}5) + 0.70(I\text{-}6) + 6.49 \quad (3)$$

| $n$ | $r$ | $s$ |
|---|---|---|
| 244 | 0.923 | 0.377 |

$I$-1 in eq 3 is an indicator variable taking the value of 1 for enzyme from Walker tumor and 0 for enzyme from L1210 tumor. $I$-2 accounts for the highly deleterious effect of X in the ortho position, $I$-3 accounts for the bad effect of rigidity of bridges within X to a second benzene ring which is also a part of X, $I$-4 accounts for the activity-enhancing effect of congeners having the highly active leaving group $-SO_2OC_6H_4-$, $I$-5 accounts for the increased activity of flexible bridges between two benzene rings in X, and $I$-6 takes the value of 1 for bridges of the type $CH_2NHCONHC_6H_4-$, $CH_2CH_2CONRC_6H_4-$, and $CH_2CH_2CH_2NRC_6H_4-$ (R = $CH_3$ or H) between two rings in X.

4. Inhibition[9] of complement by

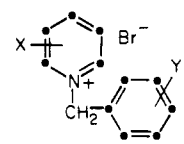



$$\log 1/C = 0.15(\text{MR-X,Y}) + 1.07(I\text{-}1) + 0.52(I\text{-}2) + 0.43(I\text{-}3) + 2.42 \quad (4)$$

| $n$ | $r$ | $s$ |
|---|---|---|
| 108 | 0.935 | 0.258 |

$I$-1 is the indicator variable which parameterizes *both* the bridge and the phenyl ring as shown above, $I$-2 accounts for the presence of a pyridine moiety, and $I$-3 is used to account for the special activating effect of 3-$NHCOXC_6H_5$ in the second ring.

5. Inhibition[10] of complement by



$$\log 1/C = 0.18(\pi\text{-X}) + 0.46(\pi\text{-Y}) + 1.01\sigma_X^+ + 0.72(I\text{-}1) + 2.50 \quad (5)$$

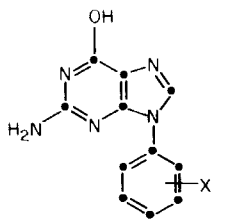
| $n$ | $r$ | $s$ |
|---|---|---|
| 69 | 0.939 | 0.198 |

$\sigma X^+$ in eq 5 refers to substituents on the pyridine ring and $I$-1 parameterizes the activity-enhancing effect of having an $SO_2F$ in the ortho position of the benzyl ring.

The above five equations relate enzymic activity with chemical structure for 483 inhibitors of four different enzymes in an objective fashion which gives one an overview of an enormous amount of work. The purpose of this report is to complete the analysis of Baker's studies with guanine deaminase and xanthine oxidase and draw some general conclusions about the formulation of QSAR from enzyme studies.

**Method.** The dependent variables (log $1/C$) in the equations reported in this paper are all in terms of the molar concentration causing 50% inhibition of enzyme activity. The constants for $\pi$ and MR are taken from our recent compilation[11] or calculated from additivity principles.[12] Many examples illustrating the calculation of $\pi$

Table I. Inhibition Constants and Physicochemical Parameters for the Reversible Inhibition of Guanine Deaminase by 9-(X-Phenyl)guanines




| No. | X | Log 1/C Obsd[a] | Log 1/C Calcd[b] | \|Δ log 1/C\| | π-3[c] | π-4[c] | MR-3[c] | MR-4[c] | $E_s$-2[c] | I-1[c] | I-2[c] | I-3[c] | % irreversible inact.[d] | Enzyme source[e] | Ref[f] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4-$NHCOC_6H_4$-3'-$SO_2F$ | 3.82 | 4.611 | 0.79 | 0.00 | 0.54 | 0.10 | 4.23 | 1.24 | 0 | 1 | 0 | 0 | 1 | 16d |
| 2 | 2-Br | 3.92 | 3.858 | 0.06 | 0.00 | 0.00 | 0.10 | 0.10 | 0.08 | 0 | 0 | 0 | | 2 | 16b |
| 3 | 2-Cl | 4.00 | 4.047 | 0.05 | 0.00 | 0.00 | 0.10 | 0.10 | 0.27 | 0 | 0 | 0 | | 2 | 16b |
| 4 | 4-$NHSO_2C_6H_4$-3'-$SO_2F$ | 4.04 | 4.595 | 0.56 | 0.00 | 0.50 | 0.10 | 4.55 | 1.24 | 0 | 1 | 0 | 38 | 1 | 16d |
| 5 | 4-$NHSO_2C_6H_4$-4'-$SO_2F$ | 4.37 | 4.595 | 0.23 | 0.00 | 0.50 | 0.10 | 4.55 | 1.24 | 0 | 1 | 0 | 0 | 1 | 16d |
| 6 | 4-$N(CH_3)_2$ | 4.47 | 5.084 | 0.61 | 0.00 | 0.18 | 0.10 | 1.56 | 1.24 | 0 | 0 | 0 | | 2 | 16b |
| 7 | 4-$CONH_2$ | 4.64 | 4.411 | 0.23 | 0.00 | −1.49 | 0.10 | 0.98 | 1.24 | 0 | 0 | 0 | | 1 | 16d |
| 8 | 4-$NH_2$ | 4.66 | 4.516 | 0.14 | 0.00 | −1.23 | 0.10 | 0.54 | 1.24 | 0 | 0 | 0 | | 1 | 16d |
| 9 | 2-F | 4.68 | 4.554 | 0.23 | 0.00 | 0.00 | 0.10 | 0.10 | 0.78 | 0 | 0 | 0 | | 2 | 16b |
| 10 | 4-$NHCOC_6H_4$-4'-$SO_2F$ | 4.74 | 4.611 | 0.13 | 0.00 | 0.54 | 0.10 | 4.23 | 1.24 | 0 | 1 | 0 | | 2 | 16c |
| 11 | 3-$NHCOC_6H_4$-4'-$SO_2F$, 4-$OCH_3$ | 4.77 | 5.148 | 0.38 | 0.54 | −0.02 | 4.23 | 0.79 | 1.24 | 1 | 1 | 1 | 0 | 1 | 16d |
| 12 | 3-$NHCONHC_6H_4$-3'-$SO_2F$, 4-$OCH_3$ | 4.82 | 5.167 | 0.35 | 1.84 | −0.02 | 4.62 | 0.79 | 1.24 | 1 | 1 | 1 | 0 | 1 | 16d |
| 13 | 2,3-CH=CHCH=CH | 4.85 | 4.946 | 0.10 | 0.66 | 0.00 | 0.87 | 0.10 | 0.36 | 0 | 0 | 0 | | 2 | 16b |
| 14 | 3-$NHCONHC_6H_4$-4'-$SO_2F$, 4-$OCH_3$ | 4.89 | 5.167 | 0.28 | 1.84 | −0.02 | 4.62 | 0.79 | 1.24 | 1 | 1 | 1 | 0 | 1 | 16d |
| 15 | 3-$NHCOC_6H_4$-3'-$SO_2F$, 4-$OCH_3$ | 4.92 | 5.148 | 0.23 | 0.54 | −0.02 | 4.23 | 0.79 | 1.24 | 1 | 1 | 1 | 0 | 1 | 16d |
| 16 | 4-$NHCOCH_2Br$ | 5.00 | 4.862 | 0.14 | 0.00 | −0.37 | 0.10 | 2.28 | 1.24 | 0 | 0 | 0 | 0 | 1 | 16d |
| 17 | 4-$C(CH_3)_3$ | 5.03 | 5.809 | 0.78 | 0.00 | 1.98 | 0.10 | 1.96 | 1.24 | 0 | 0 | 0 | | 2 | 16b |
| 18 | 3-$NH_2$ | 5.05 | 5.492 | 0.44 | −1.23 | 0.00 | 0.54 | 0.10 | 1.24 | 0 | 0 | 0 | | 1 | 16d |
| 19 | 4-$CH_3$ | 5.09 | 5.237 | 0.15 | 0.00 | 0.56 | 0.10 | 0.57 | 1.24 | 0 | 0 | 0 | | 1 | 16d |
| 20 | 4-$CF_3$ | 5.10 | 5.366 | 0.27 | 0.00 | 0.88 | 0.10 | 0.50 | 1.24 | 0 | 0 | 0 | | 1 | 16d |
| 21 | H | 5.10 | 5.011 | 0.09 | 0.00 | 0.00 | 0.10 | 0.10 | 1.24 | 0 | 0 | 0 | | 1 | 16d |
| 22 | 3,4-$(OCH_3)_2$ | 5.15 | 3.921 | 1.23 | −0.02 | −0.02 | 0.79 | 0.79 | 1.24 | 1 | 0 | 1 | | 2 | 16a |
| 23 | 3-$CF_3$ | 5.35 | 5.450 | 0.10 | 0.88 | 0.00 | 0.50 | 0.10 | 1.24 | 0 | 0 | 0 | | 2 | 16a |
| 24 | 3-$OCH_3$ | 5.42 | 5.738 | 0.32 | −0.02 | 0.00 | 0.79 | 0.10 | 1.24 | 0 | 0 | 0 | | 2 | 16a |
| 25 | 3-$CH_3$ | 5.58 | 5.515 | 0.07 | 0.56 | 0.00 | 0.57 | 0.10 | 1.24 | 0 | 0 | 0 | | 1 | 16d |
| 26[g] | 3-$NHSO_2C_6H_4$-4'-$SO_2F$ | 5.59 | 6.984 | 1.39 | 0.50 | 0.00 | 4.55 | 0.10 | 1.24 | 0 | 1 | 0 | 35 | 1 | 16d |
| 27 | 3-Cl | 5.60 | 5.554 | 0.05 | 0.71 | 0.00 | 0.60 | 0.10 | 1.24 | 0 | 0 | 0 | | 2 | 16a |
| 28 | 4-$C_2H_5$ | 5.64 | 5.422 | 0.22 | 0.00 | 1.02 | 0.10 | 1.03 | 1.24 | 0 | 0 | 0 | | 1 | 16d |
| 29 | 4-Cl | 5.68 | 5.297 | 0.38 | 0.00 | 0.71 | 0.10 | 0.60 | 1.24 | 0 | 0 | 0 | | 1 | 16d |
| 30 | 4-$CH(CH_3)_2$ | 5.92 | 5.628 | 0.29 | 0.00 | 1.53 | 0.10 | 1.50 | 1.24 | 0 | 0 | 0 | | 2 | 16b |
| 31 | 4-OH | 6.00 | 6.349 | 0.35 | 0.00 | −0.67 | 0.10 | 0.29 | 1.24 | 0 | 0 | 1 | | 2 | 16b |
| 32 | 3,4-$Cl_2$ | 6.03 | 5.840 | 0.19 | 0.71 | 0.71 | 0.60 | 0.60 | 1.24 | 0 | 0 | 0 | | 2 | 16a |
| 33 | 4-$O(CH_2)_2NHCOC_6H_4$-3'-$SO_2F$ | 6.15 | 6.372 | 0.22 | 0.00 | 0.92 | 0.10 | 5.37 | 1.24 | 0 | 1 | 1 | 47 | 1 | 16e |
| 34 | 4-$(CH_2)_3CH_3$ | 6.19 | 5.817 | 0.37 | 0.00 | 2.00 | 0.10 | 1.96 | 1.24 | 0 | 0 | 0 | | 2 | 16b |
| 35 | 4-$O(CH_2)_2NHCOC_6H_3$-2'-$CH_3$, 5'-$SO_2F$ | 6.21 | 6.598 | 0.39 | 0.00 | 1.48 | 0.10 | 5.83 | 1.24 | 0 | 1 | 1 | 13 | 1 | 16e |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 36 | 4-$C_6H_5$ | 6.27 | 5.801 | 0.47 | 0.00 | 1.96 | 0.10 | 2.54 | 1.24 | 0 | 0 | 0 | | 1 | 16d |
| 37 | 3-NHCHO | 6.36 | 5.968 | 0.39 | −0.98 | 0.00 | 1.03 | 0.10 | 1.24 | 0 | 0 | 0 | | 2 | 16b |
| 38 | 4-$OCH_3$ | 6.46 | 6.611 | 0.15 | 0.00 | −0.02 | 0.10 | 0.79 | 1.24 | 0 | 0 | 1 | | 1 | 16d |
| 39 | 4-$O(CH_2)_2NHCOC_6H_3$-4'-Cl, 3'-$SO_2F$ | 6.46 | 6.658 | 0.20 | 0.00 | 1.63 | 0.10 | 5.87 | 1.24 | 0 | 1 | 1 | 32 | 1 | 16e |
| 40 | 3-$NHCONHC_6H_2$-2',4'-$(CH_3)_2$, 5'-$SO_2F$ | 6.46 | 6.874 | 0.41 | 2.96 | 0.00 | 5.54 | 0.10 | 1.24 | 0 | 1 | 0 | 0 | 1 | 16d |
| 41 | 3-$NHCONHC_6H_3$-2'-Cl, 5'-$SO_2F$ | 6.46 | 6.951 | 0.49 | 2.55 | 0.00 | 5.12 | 0.10 | 1.24 | 0 | 1 | 0 | 31 | 1 | 16d |
| 42 | 3-$NHSO_2C_6H_4$-3'-$SO_2F$ | 6.47 | 6.984 | 0.51 | 0.50 | 0.00 | 4.55 | 0.10 | 1.24 | 0 | 1 | 0 | 24 | 1 | 16d |
| 43 | 3-$NHCONHC_6H_3$-2'-$CH_3$, 5'-$SO_2F$ | 6.54 | 6.956 | 0.42 | 2.40 | 0.00 | 5.08 | 0.10 | 1.24 | 0 | 1 | 0 | 21 | 1 | 16e |
| 44 | 4-$O(CH_2)_2NHCOC_6H_4$-4'-$SO_2F$ | 6.55 | 6.372 | 0.18 | 0.00 | 0.92 | 0.10 | 5.37 | 1.24 | 0 | 1 | 1 | 46 | 1 | 16e |
| 45 | 3,4-CH=CHCH=CH | 6.55 | 6.087 | 0.46 | 0.66 | 0.66 | 0.87 | 0.87 | 1.24 | 0 | 0 | 0 | | 1 | 16d |
| 46 | 3-$NHCOCH_2OC_6H_4$-4'-$SO_2F$ | 6.60 | 6.972 | 0.37 | 0.54 | 0.00 | 4.92 | 0.10 | 1.24 | 0 | 1 | 0 | 0 | 1 | 16d |
| 47 | 4-$O(CH_2)_2NHCOC_6H_2$-2',4'-$(CH_3)_2$, 5'-$SO_2F$ | 6.62 | 6.823 | 0.20 | 0.00 | 2.04 | 0.10 | 6.29 | 1.24 | 0 | 1 | 1 | 17 | 1 | 16e |
| 48 | 3-$C_6H_5$ | 6.62 | 7.053 | 0.43 | 1.96 | 0.00 | 2.54 | 0.10 | 1.24 | 0 | 0 | 0 | | 2 | 16b |
| 49 | 3-$NHCO(CH_2)_4C_6H_4$-4'-$SO_2F$ | 6.66 | 6.711 | 0.05 | 2.04 | 0.00 | 6.07 | 0.10 | 1.24 | 0 | 1 | 0 | 62 | 1 | 16d |
| 50 | 3-$NHCO(CH_2)_2C_6H_4$-4'-$SO_2F$ | 6.68 | 6.947 | 0.27 | 1.04 | 0.00 | 5.15 | 0.10 | 1.24 | 0 | 1 | 0 | 68 | 1 | 16d |
| 51 | 4-$O(CH_2)_2NHCONHCH_2C_6H_4$-4'-$SO_2F$ | 6.70 | 6.735 | 0.04 | 0.00 | 1.82 | 0.10 | 6.20 | 1.24 | 0 | 1 | 1 | 12 | 1 | 16e |
| 52 | 3-$NHCOCH_2Br$ | 6.70 | 6.909 | 0.21 | −0.37 | 0.00 | 2.28 | 0.10 | 1.24 | 0 | 0 | 0 | 30 | 1 | 16d |
| 53 | 4-$OC_2H_5$ | 6.72 | 6.772 | 0.05 | 0.00 | 0.38 | 0.10 | 1.25 | 1.24 | 0 | 0 | 1 | | 1 | 16d |
| 54 | 3-$NHCONHC_6H_3$-2'-$OCH_3$, 5'-$SO_2F$ | 6.72 | 6.923 | 0.20 | 1.82 | 0.00 | 5.30 | 0.10 | 1.24 | 0 | 1 | 0 | 0 | 1 | 16d |
| 55 | 4-$O(CH_2)_3NHCOC_6H_4$-4'-$SO_2F$ | 6.74 | 6.574 | 0.17 | 0.00 | 1.42 | 0.10 | 5.84 | 1.24 | 0 | 1 | 1 | 0 | 1 | 16d |
| 56 | 3-$NHCOCH_2C_6H_4$-4'-$SO_2F$ | 6.80 | 6.984 | 0.18 | 0.54 | 0.00 | 4.69 | 0.10 | 1.24 | 0 | 1 | 0 | 54 | 1 | 16d |
| 57 | 4-$O(CH_2)_2NHCOC_6H_3$-2'-$OCH_3$, 5'-$SO_2F$ | 6.80 | 6.364 | 0.44 | 0.00 | 0.90 | 0.10 | 6.05 | 1.24 | 0 | 1 | 1 | 0 | 1 | 16e |
| 58 | 4-$O(CH_2)_2NHCOC_6H_3$-4'-$CH_3$, 3'-$SO_2F$ | 6.82 | 6.598 | 0.22 | 0.00 | 1.48 | 0.10 | 5.83 | 1.24 | 0 | 1 | 1 | 0 | 1 | 16d |
| 59 | 4-$O(CH_2)_3NHCOC_6H_4$-3'-$SO_2F$ | 6.85 | 6.574 | 0.28 | 0.00 | 1.42 | 0.10 | 5.84 | 1.24 | 0 | 1 | 1 | 0 | 1 | 16d |
| 60 | 4-$O(CH_2)_2NHCONHC_6H_3$-3'-Cl, 4'-$SO_2F$ | 6.85 | 7.182 | 0.33 | 0.00 | 2.93 | 0.10 | 6.27 | 1.24 | 0 | 1 | 1 | 22 | 1 | 16e |
| 61 | 4-$O(CH_2)_2NHCONHC_6H_3$-2'-Cl, 5'-$SO_2F$ | 6.89 | 7.182 | 0.29 | 0.00 | 2.93 | 0.10 | 6.27 | 1.24 | 0 | 1 | 1 | 18 | 1 | 16d |
| 62 | 4-$O(CH_2)_2NHCONHC_6H_2$-2',4'-$(CH_3)_2$, 5'-$SO_2F$ | 6.89 | 7.347 | 0.46 | 0.00 | 3.34 | 0.10 | 6.67 | 1.24 | 0 | 1 | 1 | 0 | 1 | 16e |
| 63 | 3-$NHCONHC_6H_4$-3'-$SO_2F$ | 6.92 | 6.984 | 0.06 | 1.84 | 0.00 | 4.62 | 0.10 | 1.24 | 0 | 1 | 0 | 45 | 1 | 16d |
| 64 | 3-$NHCOC_6H_4$-3'-$SO_2F$ | 6.96 | 6.966 | 0.01 | 0.54 | 0.00 | 4.23 | 0.10 | 1.24 | 0 | 1 | 0 | 93 | 1 | 16d |
| 65 | 4-$O(CH_2)_2NHCONHC_6H_3$-3'-$CH_3$, 4'-$SO_2F$ | 7.00 | 7.121 | 0.12 | 0.00 | 2.78 | 0.10 | 6.19 | 1.24 | 0 | 1 | 1 | 11 | 1 | 16e |
| 66 | 4-$O(CH_2)_2NHCONHC_6H_3$-4'-Cl, 3'-$SO_2F$ | 7.00 | 7.182 | 0.18 | 0.00 | 2.93 | 0.10 | 6.27 | 1.24 | 0 | 1 | 1 | 40 | 1 | 16e |
| 67 | 4-$O(CH_2)_2NHCONHC_6H_3$-2'-$CH_3$, 5'-$SO_2F$ | 7.05 | 7.121 | 0.07 | 0.00 | 2.78 | 0.10 | 6.19 | 1.24 | 0 | 1 | 1 | 18 | 1 | 16e |
| 68 | 4-$O(CH_2)_2NHCONHC_6H_3$-2'-$OCH_3$, 5'-$SO_2F$ | 7.07 | 6.888 | 0.18 | 0.00 | 2.20 | 0.10 | 6.45 | 1.24 | 0 | 1 | 1 | 0 | 1 | 16d |
| 69 | 3-$NHCOC_6H_3$-4'-Cl, 3'-$SO_2F$ | 7.10 | 6.983 | 0.12 | 1.25 | 0.00 | 4.73 | 0.10 | 1.24 | 0 | 1 | 0 | 100 | 1 | 16d |
| 70 | 3-$NHCOC_6H_5$-3'-Cl, 4'-$SO_2F$ | 7.11 | 6.983 | 0.13 | 1.25 | 0.00 | 4.73 | 0.10 | 1.24 | 0 | 1 | 0 | 100 | 1 | 16e |
| 71 | 3-$NHCONHC_6H_3$-4'-$CH_3$, 3'-$SO_2F$ | 7.14 | 6.956 | 0.18 | 2.40 | 0.00 | 5.08 | 0.10 | 1.24 | 0 | 1 | 0 | 0 | 1 | 16d |
| 72 | 4-$O(CH_2)_2NHCONHC_6H_4$-3'-$SO_2F$ | 7.15 | 6.896 | 0.25 | 0.00 | 2.22 | 0.10 | 5.77 | 1.24 | 0 | 1 | 1 | 43 | 1 | 16d |
| 73 | 4-$O(CH_2)_2NHCONHC_6H_4$-4'-$SO_2F$ | 7.17 | 6.896 | 0.27 | 0.00 | 2.22 | 0.10 | 5.77 | 1.24 | 0 | 1 | 1 | 0 | 1 | 16d |
| 74 | 3-$NHCOC_6H_4$-4'-$SO_2F$ | 7.19 | 6.966 | 0.22 | 0.54 | 0.00 | 4.23 | 0.10 | 1.24 | 0 | 1 | 0 | 94 | 1 | 16d |
| 75 | 3-$NHCONHC_6H_3$-4'-Cl, 3'-$SO_2F$ | 7.21 | 6.951 | 0.26 | 2.55 | 0.00 | 5.12 | 0.10 | 1.24 | 0 | 1 | 0 | 52 | 1 | 16d |

Table I (Continued)

| No. | X | Log 1/C Obsd[a] | Log 1/C Calcd[b] | Δ log 1/C | π-3[c] | π-4[c] | MR-3[c] | MR-4[c] | $E_s$-2[c] | I-1[c] | I-2[c] | I-3[c] | % irreversible inact.[d] | Enzyme source[e] | Ref[f] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 76 | 3-NHCOC$_6$H$_3$-2'-CH$_3$, 5'-SO$_2$F | 7.22 | 6.984 | 0.24 | 1.10 | 0.00 | 4.69 | 0.10 | 1.24 | 0 | 1 | 0 | 59 | 1 | 16e |
| 77 | 3-NHCOC$_6$H$_3$-3'-CH$_3$, 4'-SO$_2$F | 7.25 | 6.984 | 0.27 | 1.10 | 0.00 | 4.69 | 0.10 | 1.24 | 0 | 1 | 0 | 94 | 1 | 16e |
| 78 | 4-O(CH$_2$)$_3$C$_6$H$_5$ | 7.25 | 7.691 | 0.44 | 0.00 | 2.66 | 0.10 | 4.15 | 1.24 | 0 | 0 | 1 | | 2 | 16b |
| 79 | 3-NHCONHC$_6$H$_3$-3'-Cl, 4'-SO$_2$F | 7.28 | 6.951 | 0.33 | 2.55 | 0.00 | 5.12 | 0.10 | 1.24 | 0 | 1 | 0 | 0 | 1 | 16d |
| 80 | 3-NHCOC$_6$H$_3$-2'-OCH$_3$, 5'-SO$_2$F | 7.30 | 6.973 | 0.33 | 0.52 | 0.00 | 4.91 | 0.10 | 1.24 | 0 | 1 | 0 | 72 | 1 | 16e |
| 81 | 3-NHCOC$_6$H$_5$ | 7.30 | 7.435 | 0.14 | 0.49 | 0.00 | 3.46 | 0.10 | 1.24 | 0 | 0 | 0 | | 2 | 16b |
| 82 | 3-NHCONHC$_6$H$_4$-3'-CH$_3$, 4'-SO$_2$F | 7.31 | 6.956 | 0.35 | 2.40 | 0.00 | 5.08 | 0.10 | 1.24 | 0 | 1 | 0 | 12 | 1 | 16e |
| 83 | 3-NHCOC$_6$H$_3$-2',4'-Cl$_2$ | 7.33 | 7.599 | 0.27 | 1.91 | 0.00 | 4.46 | 0.10 | 1.24 | 0 | 0 | 0 | 0 | 1 | 16e |
| 84 | 3-NHCOC$_6$H$_2$-2',4'-Cl$_3$, 5'-SO$_2$F | 7.34 | 6.936 | 0.40 | 1.96 | 0.00 | 5.23 | 0.10 | 1.24 | 0 | 1 | 0 | 100 | 1 | 16e |
| 85 | 4-O(CH$_2$)$_2$NHCONHC$_6$H$_3$-4'-CH$_3$, 3'-SO$_2$F | 7.36 | 7.121 | 0.24 | 0.00 | 2.78 | 0.10 | 6.20 | 1.24 | 0 | 1 | 1 | 0 | 1 | 16d |
| 86 | 3-NHCOC$_6$H$_2$-2',4'-(CH$_3$)$_2$, 5'-SO$_2$F | 7.38 | 6.947 | 0.43 | 1.66 | 0.00 | 5.15 | 0.10 | 1.24 | 0 | 1 | 0 | 0 | 1 | 16e |
| 87 | 3-NHCOC$_6$H$_3$-4'-CH$_3$, 3'-SO$_2$F | 7.42 | 6.984 | 0.44 | 1.10 | 0.00 | 4.69 | 0.10 | 1.24 | 0 | 1 | 0 | 55 | 1 | 16d |
| 88 | 3-NHCOC$_6$H$_4$-3'-Cl | 7.44 | 7.549 | 0.11 | 1.20 | 0.00 | 3.96 | 0.10 | 1.24 | 0 | 0 | 0 | 21 | 1 | 16e |
| 89 | 4-O(CH$_2$)$_3$NHCONHC$_6$H$_4$-4'-SO$_2$F | 7.44 | 7.097 | 0.34 | 0.00 | 2.72 | 0.10 | 6.20 | 1.24 | 0 | 1 | 1 | 0 | 1 | 16d |
| 90 | 3-NHCOC$_6$H$_3$-2'-Cl, 5'-SO$_2$F | 7.46 | 6.983 | 0.48 | 1.25 | 0.00 | 4.73 | 0.10 | 1.24 | 0 | 1 | 0 | 100 | 1 | 16d |
| 91 | 3-NHCONHC$_6$H$_4$-4'-SO$_2$F | 7.47 | 6.984 | 0.49 | 1.84 | 0.00 | 4.62 | 0.10 | 1.24 | 0 | 1 | 0 | 13 | 1 | 16d |
| 92 | 3-NHCOC$_6$H$_4$-2'-Cl | 7.60 | 7.549 | 0.05 | 1.20 | 0.00 | 3.96 | 0.10 | 1.24 | 0 | 0 | 0 | | 1 | 16e |
| 93 | 4-O(CH$_2$)$_3$NHCONHC$_6$H$_4$-3'-SO$_2$F | 8.02 | 7.097 | 0.92 | 0.00 | 2.72 | 0.10 | 6.20 | 1.24 | 0 | 1 | 1 | 0 | 1 | 16d |

[a] See Baker et al.[16] [b] Calculated using eq 12. [c] See section on Method for sources of these constants. [d] % irreversible inactivation from ref 16. [e] Two sources of enzyme were employed: (1) Walker 256 carcinoma rat; (2) rabbit liver. [f] Reference 16 for inhibition data. [g] This point not used in deriving eq 7–12.

for molecules considered in the present paper are to be found in ref 6–10. We have factored substituent effects in all of the work in terms of $\pi$ and MR and explored the factored terms $\pi$-1, $\pi$-2, etc., to seek out position dependence of activity on $\pi$ and MR. The MR values have been scaled by 0.1. In some instances, Taft's $E_s$ parameter[13] has been found to be of value; $E_s$ for NHCOR and NHSO$_2$R has been estimated[7] from similar functions CH$_2$NO$_2$ and CH$_2$SO$_3$ from Talvik and Palm's compilation.[13c] Electronic effects of substituents are not prominent among the inhibitors considered in this report. We have systematically explored the use of $\sigma$, $\sigma^-$, and $\sigma^+$ as well as $\mathcal{F}$ and $\mathcal{R}$ of Swain and Lupton.[11] In appropriate circumstances, $\sigma_{ortho}$ values from Charton's review[14] were also studied. Of course, many indicator variables were considered. The reader unfamiliar with their use should consult ref 15.

The data are from Baker's studies on guanine deaminase[16] and xanthine oxidase.[16a-c,17] The data table on guanine deaminase contains a column indicating the degree of irreversible inhibition; this is only of qualitative importance since the degrees of inhibition for different inhibitors were measured at different times. These figures do show that inhibitors capable of causing irreversible inhibition are as well correlated as the purely reversible inhibitors.

## Results and Discussion

**Guanine Deaminase.** The "best" equation for guanine deaminase inhibitors is eq 12. Table III shows the relative

$$\log 1/C = 1.176\ (\pm 0.25)\ (\text{MR-3}) + 0.403\ (\pm 0.11)\ (\pi\text{-4}) - 3.417\ (\pm 0.44)\ (I\text{-1}) + 1.608\ (\pm 0.29)\ (I\text{-3}) - 0.127\ (\pm 0.05)\ (\text{MR-3})^2 - 0.618\ (\pm 0.25)\ (I\text{-2}) + 0.994\ (\pm 0.43)\ (E_s\text{-2}) + 3.659\ (\pm 0.50) \quad (12)$$

$$n = 92 \qquad r = 0.941 \qquad s = 0.366$$

importance of the various terms in the development of eq 12 and Table II shows the interrelationship of the variables. Except for eq 6, each of the equations has the lowest standard deviation in its class. A slightly better (than eq 6) single-variable equation was found with $I$-2; however, this is only because of very high collinearity between $I$-2 and MR-3,4 (see Table II). The $F$ statistic shows the high significance of the addition of each successive term, except for eq 8 in which $I$-1 and $I$-3 displace $I$-4. ($F_{1,60;\,\alpha\,0.001}$ = 11.97.) Equation 12, which is based on 92 congeners of Table I, correlates 60 more inhibitors than eq 1, all having an SO$_2$F function; nevertheless, the two equations are highly parallel in their points of commonality. Equation 12 contains two new indicator variables, $I$-1 and $I$-2. $I$-1 accounts for five cases in which a second substituent is present with a 4-OCH$_3$. Only one such molecule was present in our first analysis and that poorly fit point was omitted from the study. This point (compound **22**, Table I) is still poorly fit although the other molecules (compounds **11**, **12**, **14**, and **15**) in this class are reasonably well fit. Speculation on this highly deleterious effect was made in an earlier publication.[6] The second new indicator variable, $I$-2, takes the value of 1 for the presence of an SO$_2$F group; strangely, this function lowers reversible inhibitor power. Since Baker's enzyme preparations were rather crude, this lower activity might be caused by reaction with impurity. $I$-3 in eq 12 corresponds to $I$ in eq 1; this indicator variable takes a value of 1 for each congener having a 4-OR group. The coefficient with this term is very close indeed to that of $I$ in eq 1 despite the

Table II. Squared Correlation Matrix Showing Degree of Collinearity ($r^2$) between the Important Variables Used in Guanine Deaminase Correlation Analysis

| | $\pi$-3 | $\pi$-4 | $\pi$-3,4 | MR-3 | MR-4 | MR-3,4 | $E_s$-2 | $I$-1 | $I$-2 | $I$-3 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\pi$-3 | 1.00 | 0.16 | 0.18 | 0.68 | 0.23 | 0.09 | 0.01 | 0.01 | 0.10 | 0.12 |
| $\pi$-4 | | 1.00 | 0.43 | 0.26 | 0.74 | 0.21 | 0.02 | 0.02 | 0.07 | 0.39 |
| $\pi$-3,4 | | | 1.00 | 0.03 | 0.19 | 0.48 | 0.04 | 0.00 | 0.27 | 0.11 |
| MR-3 | | | | 1.00 | 0.36 | 0.11 | 0.02 | 0.04 | 0.16 | 0.16 |
| MR-4 | | | | | 1.00 | 0.30 | 0.02 | 0.01 | 0.16 | 0.57 |
| MR-3,4 | | | | | | 1.00 | 0.12 | 0.00 | 0.80 | 0.22 |
| $E_s$-2 | | | | | | | 1.00 | 0.00 | 0.06 | 0.02 |
| $I$-1 | | | | | | | | 1.00 | 0.01 | 0.11 |
| $I$-2 | | | | | | | | | 1.00 | 0.11 |
| $I$-3 | | | | | | | | | | 1.00 |

Table III. Development of QSAR for Guanine Deaminase from Walker 256 Tumor and Rabbit Liver

| Eq no. | Intercept | MR-3 | $\pi$-4 | $I$-1 | $I$-3 | (MR-3)$^2$ | $I$-2 | $E_s$-2 | $r$ | $s$ | $F_{1,X}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 5.92 | 0.18 | | | | | | | 0.394 | 0.963 | 17 |
| 7 | 5.16 | 0.35 | 0.68 | | | | | | 0.709 | 0.743 | 62 |
| 8 | 5.11 | 0.40 | | −3.38 | 1.70 | | | | 0.842 | 0.571 | |
| 9 | 4.84 | 1.23 | | −3.84 | 1.89 | −0.16 | | | 0.879 | 0.508 | 24 |
| 10 | 4.68 | 1.33 | 0.35 | −3.31 | 1.37 | −0.17 | | | 0.910 | 0.443 | 28 |
| 11 | 4.75 | 1.24 | 0.43 | −3.47 | 1.67 | −0.14 | −0.58 | | 0.926 | 0.406 | 17 |
| 12 | 3.66 | 1.18 | 0.40 | −3.42 | 1.61 | −0.13 | −0.62 | 0.99 | 0.941 | 0.366 | 21 |

fact that only four examples of a 4-OR function were present in the formulation of eq 1. Equation 12 is based on 31 such examples. It is impressive how nicely additive substituent effects often are in enzyme inhibitors.

The coefficients with $E_s$-2 and the intercepts of eq 1 and 12 are, for practical purposes, identical. The coefficient with MR-3 is higher in eq 12 than in eq 1; this is because eq 12 contains an exponential term in MR-3. No substituents with high MR-3 values were present in the set of 32 congeners on which eq 1 rests. Some very large 3-substituents are present in the larger set of Table I and the problem of bulk tolerance is raised.

As pointed out in our earlier publication,[6] the special activating effect of 4-OR is most interesting. It was noted then that this effect prevails with guanine deaminase from two different sources: Walker 256 rat tumor and rabbit liver. Data with enzyme from both sources have been used in the formulation of eq 12. Use of an indicator variable revealed no difference in enzyme from the two different sources. The role of the 4-OR function is quite strong and specific. At present, it is not possible to account for this unusual effect[6] with any assurance; it may be possible to account for this unusual interaction when the x-ray crystallography of inhibitors bound to guanine deaminase is done.

A specific effect for $OCH_3$ has also been found for phenethanolamine inhibitors of *N*-methyltransferase;[18] in this case, however, the weighting factor was negative instead of positive.

One data point which is very poorly fit (compound **26**, Table I) has not been included in the analysis. The reasons for its abnormality are not obvious.

A slight improvement of eq 12 can be made by the addition of a term in $E_s$-4. This yields an equation with $r = 0.947$ and $s = 0.351$; $F_{1,75} = 8.2$.

**Xanthine Oxidase.** Equation 17 has been formulated

$$\log 1/C = 0.267\ (\pm 0.06)\ (\text{MR-3}) - 0.647\ (\pm 0.12)\ (\text{MR-3}\cdot\text{MR-4}) + 1.291\ (\pm 0.39)\ (E_s\text{-2}) + 0.101\ (\pm 0.04)\ (\text{MR-4}) + 0.252\ (\pm 0.11)\ (E_s\text{-4}) + 4.552\ (\pm 0.45) \quad (17)$$

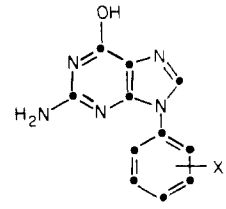
| $n$ | $r$ | $s$ |
|---|---|---|
| 65 | 0.910 | 0.308 |

from the data in Table IV. This equation is based on the 30 data points of eq 2 plus 35 new congeners, all containing the $SO_2F$ function. The stepwise development of eq 17 is shown in Table VI and the interrelationship of the variables in Table V. In comparing eq 17 with eq 2, we see that although no new variables have been added to correlate the much larger data set, the introduction of bulkier substituents has enabled us to make a sharper resolution of the MR terms. While MR-3 of eq 17 has essentially the same coefficient as in eq 2, the interesting, highly significant cross product term (MR-3·MR-4, see Table VI) with its negative coefficient shows that bulky groups in both the 3 and 4 position substantially reduce effectiveness. Substituent space in the region of the 3 and 4 position must be limited. Apparently, the MR-3·MR-4 term carries enough information about MR-4 so that this term is reduced considerably in eq 17. Relatively few examples have been reported where cross products are of value in correlation analysis despite the fact that theory suggests they may play important roles. The problem of systematically screening for meaningful cross product terms is expensive because, even taken only two at a time, the number goes up as $n(n-1)/2$ where $n$ is the number of variables. At present we are only testing those cases where chemical judgment suggests cross products might be significant. The $E_s$-2 terms and the intercepts for eq 2 and 17 are essentially the same; $E_s$-4, however, does not carry as much weight in eq 17. No indicator variable was found to be necessary for congeners containing the $SO_2F$ group.

Four data points in Table IV have not been included in the development of eq 13–17. They are all poorly fit and have so little in common that we cannot make any useful comments about their aberrant activity. However, it does appear that indicator variables formulated for one enzyme may have meaning in other systems. As pointed out above, an indicator variable for 2-$SO_2F$ in pyridinium ions inhibiting complement serves a similar purpose in correlating the same type of inhibitor acting on chymotrypsin. Special interactions of the OR function are present in guanine deaminase inhibitors (eq 1 and 12). Indicator variables bring to light special $OCH_3$ reactions in *N*-methyltransferase. It seems likely that indicator variables will be most helpful in comparative enzymology. The great utility of these correlations will become more apparent as the x-ray crystallography of ligands bound to enzymes begins to develop.

Equations 3–6, 12, and 17 correlate the structure–activity

Table IV. Inhibition Constants and Physicochemical Parameters for the Reversible Inhibition of Xanthine Oxidase by 9-(X-Phenyl)guanines



| No. | X | Log 1/C Obsd[a] | Log 1/C Calcd[b] | \|Δ log 1/C\| | π-3[c] | π-4[c] | MR-3[c] | MR-4[c] | $E_s$-2[c] | $E_s$-4[c] | Ref[d] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2-Cl | 5.09 | 5.245 | 0.16 | 0.00 | 0.00 | 0.10 | 0.10 | 0.27 | 1.24 | 16b |
| 2 | 2-Br | 5.11 | 5.000 | 0.11 | 0.00 | 0.00 | 0.10 | 0.10 | 0.08 | 1.24 | 16b |
| 3 | 3-NHCONHC$_6$H$_4$-3'-SO$_2$F, 4-OCH$_3$ | 5.25 | 5.286 | 0.04 | 1.84 | −0.02 | 4.62 | 0.79 | 1.24 | 0.69 | 17b |
| 4 | 3-NHCOC$_6$H$_4$-3'-SO$_2$F, 4-OCH$_3$ | 5.31 | 5.381 | 0.07 | 0.54 | −0.02 | 4.23 | 0.79 | 1.24 | 0.69 | 17b |
| 5 | 3-NHCONHC$_6$H$_4$-4'-SO$_2$F, 4-OCH$_3$ | 5.35 | 5.286 | 0.06 | 1.84 | −0.02 | 4.62 | 0.79 | 1.24 | 0.69 | 17b |
| 6 | 2,3-CH=CHCH=CH | 5.38 | 5.515 | 0.14 | 0.66 | 0.00 | 0.87 | 0.10 | 0.36 | 1.24 | 16b |
| 7 | 3-NHCOC$_6$H$_4$-4'-SO$_2$F, 4-OCH$_3$ | 5.39 | 5.381 | 0.01 | 0.54 | −0.02 | 4.23 | 0.79 | 1.24 | 0.69 | 17b |
| 8[e] | 4-NH$_2$ | 5.43 | 6.358 | 0.93 | 0.00 | −1.23 | 0.10 | 0.54 | 1.24 | 0.63 | 16a |
| 9 | 4-NHSO$_2$C$_6$H$_4$-4'-SO$_2$F | 5.60 | 6.022 | 0.42 | 0.00 | 0.50 | 0.10 | 4.55 | 1.24 | −1.25 | 17b |
| 10 | 4-N(CH$_3$)$_2$ | 5.68 | 6.116 | 0.44 | 0.00 | 0.18 | 0.10 | 1.56 | 1.24 | −0.47 | 16a |
| 11 | 4-NHCOCH$_2$Br | 5.72 | 5.888 | 0.17 | 0.00 | −0.37 | 0.10 | 2.28 | 1.24 | −1.47 | 16c |
| 12[e] | 3-NHCONHC$_6$H$_4$-3'-SO$_2$F | 5.74 | 7.400 | 1.66 | 1.84 | 0.00 | 4.62 | 0.10 | 1.24 | 1.24 | 16c |
| 13 | 4-Cl | 5.74 | 6.269 | 0.53 | 0.00 | 0.71 | 0.10 | 0.60 | 1.24 | 0.27 | 16a |
| 14 | 4-C(CH$_3$)$_3$ | 5.74 | 5.860 | 0.12 | 0.00 | 1.98 | 0.10 | 1.96 | 1.24 | −1.54 | 16b |
| 15 | 4-CH$_3$ | 5.80 | 6.200 | 0.40 | 0.00 | 0.56 | 0.10 | 0.57 | 1.24 | 0.00 | 16a |
| 16 | 4-CF$_3$ | 5.89 | 5.905 | 0.02 | 0.00 | 0.88 | 0.10 | 0.50 | 1.24 | −1.16 | 16b |
| 17[e] | 3-NHSO$_2$C$_6$H$_4$-3'-SO$_2$F | 5.89 | 7.387 | 1.50 | 0.50 | 0.00 | 4.55 | 0.10 | 1.24 | 1.24 | 17b |
| 18 | 3,4-Cl$_2$ | 5.96 | 6.208 | 0.25 | 0.71 | 0.71 | 0.60 | 0.60 | 1.24 | 0.27 | 16a |
| 19 | 4-O(CH$_2$)$_3$NHCOC$_6$H$_4$-4'-SO$_2$F | 6.00 | 6.556 | 0.56 | 0.00 | 1.42 | 0.10 | 5.84 | 1.24 | 0.69 | 17c |
| 20 | 4-NHSO$_2$C$_6$H$_4$-3'-SO$_2$F | 6.02 | 6.022 | 0.00 | 0.00 | 0.50 | 0.10 | 4.55 | 1.24 | −1.25 | 17b |
| 21[e] | 3-NHSO$_2$C$_6$H$_4$-4'-SO$_2$F | 6.14 | 7.387 | 1.25 | 0.50 | 0.00 | 4.55 | 0.10 | 1.24 | 1.24 | 17b |
| 22 | 3,4-(OCH$_3$)$_2$ | 6.14 | 6.216 | 0.08 | −0.02 | −0.02 | 0.79 | 0.79 | 1.24 | 0.69 | 16a |
| 23 | 4-NHCOC$_6$H$_4$-4'-SO$_2$F | 6.15 | 5.956 | 0.19 | 0.00 | 0.54 | 0.10 | 4.23 | 1.24 | −1.47 | 16c |
| 24 | 4-O(CH$_2$)$_2$NHCOC$_6$H$_4$-4'-SO$_2$F | 6.16 | 6.540 | 0.38 | 0.00 | 0.92 | 0.10 | 5.37 | 1.24 | 0.69 | 17c |
| 25 | 4-O(CH$_2$)$_3$NHCONHC$_6$H$_4$-4'-SO$_2$F | 6.16 | 6.570 | 0.41 | 0.00 | 2.72 | 0.10 | 6.25 | 1.24 | 0.69 | 17c |
| 26 | 4-C$_2$H$_5$ | 6.17 | 6.198 | 0.03 | 0.00 | 1.02 | 0.10 | 1.03 | 1.24 | −0.07 | 16a |
| 27 | 4-O(CH$_2$)$_3$NHCOC$_6$H$_4$-3'-SO$_2$F | 6.20 | 6.556 | 0.36 | 0.00 | 1.42 | 0.10 | 5.84 | 1.24 | 0.69 | 17c |
| 28 | 2-F | 6.21 | 5.903 | 0.31 | 0.00 | 0.00 | 0.10 | 0.10 | 0.78 | 1.24 | 16b |
| 29 | 4-(CH$_2$)$_3$CH$_3$ | 6.21 | 6.150 | 0.06 | 0.00 | 2.00 | 0.10 | 1.96 | 1.24 | −0.39 | 16b |
| 30 | 3-NH$_2$ | 6.22 | 6.585 | 0.37 | −1.23 | 0.00 | 0.54 | 0.10 | 1.24 | 1.24 | 16a |
| 31 | 4-O(CH$_2$)$_2$NHCOC$_6$H$_4$-3'-SO$_2$F | 6.28 | 6.540 | 0.26 | 0.00 | 0.92 | 0.10 | 5.37 | 1.24 | 0.69 | 17c |
| 32 | 4-OCH$_3$ | 6.30 | 6.382 | 0.08 | 0.00 | −0.02 | 0.10 | 0.79 | 1.24 | 0.69 | 16a |
| 33 | 4-O(CH$_2$)$_2$NHCOC$_6$H$_3$-4'-CH$_3$, 3'-SO$_2$F | 6.31 | 6.556 | 0.25 | 0.00 | 1.48 | 0.10 | 5.83 | 1.24 | 0.69 | 17c |
| 34 | 4-CONH$_2$ | 6.38 | 6.285 | 0.10 | 0.00 | −1.49 | 0.10 | 0.98 | 1.24 | 0.28 | 16a |
| 35 | 3,4-CH=CHCH=CH | 6.39 | 6.072 | 0.32 | 0.66 | 0.66 | 0.87 | 0.87 | 1.24 | 0.36 | 16a |
| 36 | H | 6.39 | 6.497 | 0.11 | 0.00 | 0.00 | 0.10 | 0.10 | 1.24 | 1.24 | 16a |
| 37 | 4-O(CH$_2$)$_3$NHCONHC$_6$H$_4$-3'-SO$_2$F | 6.40 | 6.570 | 0.17 | 0.00 | 2.72 | 0.10 | 6.25 | 1.24 | 0.69 | 17c |
| 38 | 4-O(CH$_2$)$_2$NHCONHC$_6$H$_4$-4'-SO$_2$F | 6.48 | 6.554 | 0.07 | 0.00 | 2.22 | 0.10 | 5.77 | 1.24 | 0.69 | 17c |
| 39 | 4-NHCOC$_6$H$_4$-3'-SO$_2$F | 6.55 | 5.956 | 0.59 | 0.00 | 0.54 | 0.10 | 4.23 | 1.24 | −1.47 | 17b |
| 40 | 3-Cl | 6.57 | 6.597 | 0.03 | 0.71 | 0.00 | 0.60 | 0.10 | 1.24 | 1.24 | 16a |
| 41 | 4-CH(CH$_3$)$_2$ | 6.60 | 6.114 | 0.49 | 0.00 | 1.53 | 0.10 | 1.50 | 1.24 | −0.47 | 16b |
| 42 | 4-C$_6$H$_5$ | 6.60 | 6.339 | 0.26 | 0.00 | 1.96 | 0.10 | 2.54 | 1.24 | 0.28 | 16b |
| 43 | 3-CH$_3$ | 6.62 | 6.590 | 0.03 | 0.56 | 0.00 | 0.57 | 0.10 | 1.24 | 1.24 | 16a |
| 44 | 3-NHCHO | 6.64 | 6.682 | 0.04 | −0.98 | 0.00 | 1.03 | 0.10 | 1.24 | 1.24 | 16b |
| 45 | 3-OCH$_3$ | 6.66 | 6.634 | 0.03 | −0.02 | 0.00 | 0.79 | 0.10 | 1.24 | 1.24 | 16a |
| 46 | 4-OH | 6.68 | 6.364 | 0.32 | 0.00 | −0.67 | 0.10 | 0.29 | 1.24 | 0.69 | 16b |
| 47 | 4-O(CH$_2$)$_2$NHCONHC$_6$H$_4$-3'-SO$_2$F | 6.74 | 6.554 | 0.19 | 0.00 | 2.22 | 0.10 | 5.77 | 1.24 | 0.69 | 17c |
| 48 | 3-CF$_3$ | 6.82 | 6.577 | 0.24 | 0.88 | 0.00 | 0.50 | 0.10 | 1.24 | 1.24 | 16a |
| 49 | 4-O(CH$_2$)$_2$NHCONHC$_6$H$_3$-4'-CH$_3$, 3'-SO$_2$F | 6.92 | 6.568 | 0.35 | 0.00 | 2.78 | 0.10 | 6.19 | 1.24 | 0.69 | 17c |
| 50 | 3-NHCOC$_6$H$_4$-3'-SO$_2$F | 6.96 | 7.322 | 0.36 | 0.54 | 0.00 | 4.23 | 0.10 | 1.24 | 1.24 | 16c |
| 51 | 4-OC$_2$H$_5$ | 6.96 | 6.398 | 0.56 | 0.00 | 0.38 | 0.10 | 1.25 | 1.24 | 0.69 | 16a |

Table IV (*Continued*)

| No. | X | Log 1/C Obsd[a] | Log 1/C Calcd[b] | \|Δ log 1/C\| | π-3[c] | π-4[c] | MR-3[c] | MR-4[c] | $E_s$-2[c] | $E_s$-4[c] | Ref[d] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 52 | 3-NHCOCH$_2$OC$_6$H$_4$-4'-SO$_2$F | 7.00 | 7.461 | 0.46 | 0.54 | 0.00 | 4.92 | 0.10 | 1.24 | 1.24 | 17b |
| 53 | 4-O(CH$_2$)$_2$NHCONHC$_6$H$_3$-2'-Cl, 5'-SO$_2$F | 7.04 | 6.571 | 0.47 | 0.00 | 2.93 | 0.10 | 6.27 | 1.24 | 0.69 | 17c |
| 54 | 3-NHCOC$_6$H$_3$-4'-CH$_3$, 3'-SO$_2$F | 7.04 | 7.415 | 0.38 | 1.00 | 0.00 | 4.69 | 0.10 | 1.24 | 1.24 | 17b |
| 55 | 4-O(CH$_2$)$_3$C$_6$H$_5$ | 7.08 | 6.498 | 0.58 | 0.00 | 2.66 | 0.10 | 4.15 | 1.24 | 0.69 | 16b |
| 56 | 3-C$_6$H$_5$ | 7.09 | 6.984 | 0.11 | 1.96 | 0.00 | 2.54 | 0.10 | 1.24 | 1.24 | 16b |
| 57 | 3-NHCOC$_6$H$_5$ | 7.14 | 7.170 | 0.03 | 0.49 | 0.00 | 3.46 | 0.10 | 1.24 | 1.24 | 16b |
| 58 | 3-NHCOCH$_2$Br | 7.15 | 6.932 | 0.22 | −0.37 | 0.00 | 2.28 | 0.10 | 1.24 | 1.24 | 17a |
| 59 | 3-NHCOC$_6$H$_3$-2'-Cl, 5'-SO$_2$F | 7.15 | 7.422 | 0.27 | 1.25 | 0.00 | 4.73 | 0.10 | 1.24 | 1.24 | 17b |
| 60 | 4-O(CH$_2$)$_2$NHCONHC$_6$H$_3$-2'-OCH$_3$, 5'-SO$_2$F | 7.16 | 6.576 | 0.58 | 0.00 | 2.20 | 0.10 | 6.42 | 1.24 | 0.69 | 17c |
| 61 | 3-NHCONHC$_6$H$_3$-2'-Cl, 5'-SO$_2$F | 7.28 | 7.500 | 0.22 | 2.55 | 0.00 | 5.12 | 0.10 | 1.24 | 1.24 | 17b |
| 62 | 3-NHCOC$_6$H$_4$-4'-SO$_2$F | 7.29 | 7.322 | 0.03 | 0.54 | 0.00 | 4.23 | 0.10 | 1.24 | 1.24 | 16c |
| 63 | 3-NHCONHC$_6$H$_3$-3'-Cl, 4'-SO$_2$F | 7.48 | 7.500 | 0.02 | 2.55 | 0.00 | 5.12 | 0.10 | 1.24 | 1.24 | 17b |
| 64 | 4-NHCO(CH$_2$)$_2$C$_6$H$_4$-4'-SO$_2$F | 7.58 | 7.507 | 0.07 | 1.04 | 0.00 | 5.15 | 0.10 | 1.24 | 1.24 | 17b |
| 65 | 3-NHCONHC$_6$H$_4$-4'-SO$_2$F | 7.62 | 7.400 | 0.22 | 1.84 | 0.00 | 4.62 | 0.10 | 1.24 | 1.24 | 17b |
| 66 | 3-NHCONHC$_6$H$_3$-4'-CH$_3$, 3'-SO$_2$F | 7.74 | 7.493 | 0.25 | 2.40 | 0.00 | 5.08 | 0.10 | 1.24 | 1.24 | 16c |
| 67 | 3-NHCONHC$_6$H$_3$-2'-OCH$_3$, 5'-SO$_2$F | 7.80 | 7.537 | 0.26 | 1.82 | 0.00 | 5.30 | 0.10 | 1.24 | 1.24 | 17b |
| 68 | 3-NHCOCH$_2$C$_6$H$_4$-4'-SO$_2$F | 7.82 | 7.415 | 0.41 | 0.54 | 0.00 | 4.69 | 0.10 | 1.24 | 1.24 | 17b |
| 69 | 3-NHCO(CH$_2$)$_4$C$_6$H$_4$-4'-SO$_2$F | 8.00 | 7.692 | 0.31 | 2.04 | 0.00 | 6.07 | 0.10 | 1.24 | 1.24 | 17b |

[a] From ref 16 and 17. Irreversible inactivation is also shown only by compounds 11, 17, 20, 21, 23, and 39. [b] Calculated using eq 17. [c] See section on Method for sources of these constants. [d] See ref 16 and 17. [e] These points not used in deriving eq 13–17.

Table V. Squared Correlation Matrix Showing Degree of Collinearity ($r^2$) between the Important Variables Used in Xanthine Oxidase Correlation Analysis

| | π-3 | π-4 | MR-3 | MR-4 | $E_s$-2 | $E_s$-4 |
|---|---|---|---|---|---|---|
| π-3 | 1.00 | 0.10 | 0.61 | 0.14 | 0.01 | 0.11 |
| π-4 | | 1.00 | 0.19 | 0.61 | 0.02 | 0.06 |
| MR-3 | | | 1.00 | 0.25 | 0.02 | 0.21 |
| MR-4 | | | | 1.00 | 0.03 | 0.12 |
| $E_s$-2 | | | | | 1.00 | 0.04 |
| $E_s$-4 | | | | | | 1.00 |

data on 578 inhibitors acting on four different enzymic systems. This vast effort of Baker's, much of which was completed in about 6 years, clearly demonstrates that very potent inhibitors, yielding data which can be treated quantitatively, can be found quickly by making what he called nonclassical antimetabolites; by this he meant making gross changes in a normal substrate. Most of the so-called classical antimetabolites have been made by making a relatively small change, say an N or S for a CH=, in a substrate to produce an inhibitor. While "classical" antimetabolites may be ideal to block an enzyme and can yield fine drugs as in the case of allopurinol for gout, one often faces the more difficult problem in chemotherapy of having to selectively block a pathogen enzyme without seriously hindering the corresponding enzyme in the host. Baker reasoned that, in general, one could expect to find differences between isoenzymes more readily in the region outside the active site. Current evidence suggests that in the evolution of enzymes the structure of the active site is strongly conserved while many changes may occur outside the active site. Hence, Baker's propensity was to use large substituents which would reach into space surrounding the active site.

Baker's modus operandi for an enzymic approach to drug design consists of four steps.[19]

1. A suitable enzyme must be selected and a reversible inhibitor found. Modification of the normal substrate's more active groups will lead to reversible inhibitors. Binding points on the reversible inhibitor that complex with the enzyme should be determined.

2. Areas on the inhibitor should be determined in which bulky groups can be placed. This uncovers two types of positions: large, flexible, hydrophobic areas (Baker termed these "bulk tolerance areas" when they are within the enzyme), and noncontact areas between inhibitors and enzyme.

3. Once the noncontact area is determined, a group that can form a covalent bond with common enzymic functions should be placed in this area. The length of the side chain by which the function is attached to the parent inhibitor must be varied so that the active function can react irreversibly with a group on the enzyme outside the active site.

4. After finding the ideal length and flexibility of the side chain which is to act irreversibly, the variations in the active function itself should be investigated in order to find the function with the ideal stereoelectronic specificity; that is, a function must be found which will react covalently

Table VI. Development of QSAR for Xanthine Oxidase from Bovine Milk

| Eq no. | Intercept | MR-3 | MR-3·MR-4 | $E_s$-2 | MR-4 | $E_s$-4 | $r$ | $s$ | $F_{1,X}$ |
|---|---|---|---|---|---|---|---|---|---|
| 13 | 6.22 | 0.17 | | | | | 0.486 | 0.629 | 19 |
| 14 | 6.38 | 0.26 | −0.56 | | | | 0.743 | 0.485 | 44 |
| 15 | 4.88 | 0.25 | −0.60 | 1.30 | | | 0.841 | 0.395 | 32 |
| 16 | 4.92 | 0.32 | −0.69 | 1.05 | 0.10 | | 0.877 | 0.353 | 16 |
| 17 | 4.55 | 0.27 | −0.65 | 1.29 | 0.10 | 0.25 | 0.910 | 0.308 | 20 |

at the active site but not react to a significant degree in the random walk process through the whole animal[4] when employed in the in vivo work,

It is important in the context of chemotherapy and from the point of view of current techniques of correlation analysis to reexamine Baker's four steps for making highly effective in vitro enzyme inhibitors. The first part of the first step is of course the most difficult operation. There has been little systematic analysis of how one should go about modifying a substrate to make an effective inhibitor.

Pauling[20] made a classic observation in 1948 that, "Enzymes are molecules that are complementary in structure to the activated complexes of the reaction that they catalyze, this is, to the molecular configuration that is intermediate between the reacting substances and the products of reaction for these catalyzed processes. The attraction of the enzyme molecule for the activated complex would thus lead to a decrease in its energy, and hence to a decrease in the energy of activation of the reaction and to an increase in the rate of reaction". For a classical antimetabolite, one should design an inhibitor whose stereoelectronic configuration resembles the substrate transition complex as much as possible. This approach has uncovered some very effective inhibitors.[21] However, for inhibition with selectivity for isozymes, this is probably not the best approach since the character of the transition state is likely to have been conserved in enzyme evolution.

Baker's work shows that the activity of the functional group[22] can be increased by as much as 1000-fold by the attachment of bulky groups or groups which indirectly enhance the interaction of the active site. Following Baker's ideal, a systematic study of substituents on the parent inhibitor can serve to characterize enzymic space around the inhibitor. At this point in inhibitor development, correlation analysis becomes most helpful. A substituent at a given point on the inhibitor may not interact directly with the enzyme; it may interact hydrophobically, or it may find itself in the polar space of the enzyme. If enzyme space is highly flexible and the system is under thermodynamic control, there may be no detectable steric hindrance to the substituent–enzyme interaction. If enzymic space is inflexible or the system is under kinetic control, steric hindrance may be observed. Of course, steric effects may be intra- as well as intermolecular. Finally, the electronic effect of the substituent on the parent structure may greatly influence binding whether or not the substituent actually makes contact with the enzyme. All of these possibilities have been encountered and characterized using correlation analysis.[23]

In exploring substituent space using correlation analysis, it is of the utmost importance to practice good experimental design. Because of the multidimensional character of substituent space it is not an easy matter to select sets of substituents constituting orthogonal vectors. However, progress is being made in systematic selection procedures for obtaining maximum information from a set of molecular probes.[3]

The difficulty with $\pi$ and MR can be seen in Tables II and V. There is general agreement now that two types of space, hydrophobic and polar, are to be expected in enzymes. While there is considerable evidence to show that $\pi$ well characterizes ligand interactions in hydrophobic space, it is not yet entirely clear how to parameterize nonspecific interactions in polar space. We have been exploring[24] the possibility that molar refractivity might be a suitable parameter to assess such interactions; its use in biological correlations has been of interest for some time.[25,26] However, the collinearity between $\pi$ and MR of most sets of congeners designed without consideration for this problem is so high that one cannot make a firm statement at this time that $\pi$ and MR do correlate two quite different classes of enzymic space. It is recognized of course that any large section of enzymic space will not be strictly homogeneous; when discussing hydrophobic and polar space, we mean the predominant character. $\pi$ and MR are highly collinear for apolar substituents; it is only by selecting a mixture of polar and apolar functions that this collinearity can be broken. Even though there is rather high collinearity between $\pi$ and MR for many of the equations considered in this report, we believe the preponderance of the evidence supports the use of these two parameters in characterizing enzymic space. *In general* (but not always), when MR is scaled by 0.1 so that for apolar functions $\pi$ and MR are more or less equiscalar, it is found in the correlation equations that the coefficients with MR fall in the range 0.1–0.4 while $\pi$ is more often[27] in the range 0.4–1.2. This leads one to suspect that correlation equations linear in $\pi$ and having a low coefficient with $\pi$ are not assessing "true" hydrophobic interactions which are presumed to be primarily the desolvation of apolar functions. The large substituents correlated by MR terms with low coefficients may be inhibiting enzymic action in two ways. They may do so by simple occupation of the active site; in this circumstance, MR may model the dispersion forces between ligand substituent and enzyme.[26] Alternatively, bulky substituents in or on the enzyme may hinder the substrate from achieving its normal transition state. This might or might not involve conformational changes in the enzyme produced by the ligand.

One of the most useful techniques in correlation analysis brought out by this study of Baker's results is the importance of indicator variables. The results show that large groups, really outside the range of what one usually thinks of as substituents, make a surprisingly additive and independent contribution to inhibitory activity. While it may not be possible at present to rationalize indicator variables in terms of the physicochemical properties of the substituents, they do allow one to include very complex groups in the formulation of QSAR. This enormously increases one's scope in structuring large amounts of data in the complex type of structure–activity study necessary for modern drug development.

In summary, the use of correlation analysis is essential in steps 1 and 2 of Baker's modus operandi. It provides the only systematic way to minimize the number of molecular probes that must be synthesized and to maximize the informational content of their perturbations of the enzyme.

The third step of Baker's scheme was to incorporate groups into the reversible inhibitors which would bind covalently in enzymic space around the active site. He recommended a variety of functional groups[2a] such as $NHCOCH_2Br$, $SO_2OC_6H_4X$, $COCH_2Br$, and $SO_2F$. In particular, Baker vigorously explored the use of the last function, making many hundreds of inhibitors containing this group. It is instructive to study the set of about 260 triazines inhibiting dihydrofolate reductase.[8] Here it is seen that many inhibitors containing the $SO_2F$ function are simple reversible inhibitors, while others with different geometry may be extremely potent irreversible inhibitors. Again, with benzylpyridinium ions inhibiting complement (eq 5), proper positioning greatly increases the activity of the $SO_2F$ function; nevertheless, out of Baker's quest for antitumor drugs of about 260 triazines, many with log $1/C$ values of 7–9, the congener which eventually reached the clinic does not contain an $SO_2F$ function. This function

ppears to undergo metabolism in whole animals in the andom walk from site of injection to site of action. We oelieve that correlation analysis can play a major role in "fine tuning" a reactive function such as $SO_2F$. This might be accomplished electronically or sterically; for example, a recent high $\rho$ value of 2.79 has been reported[28] for the alkaline hydrolysis of para-substituted benzenesulfonyl fluorides. By placing proper substituents on the inhibitor with the $SO_2F$ function, one could develop maximum stability with respect to metabolism compatible with reasonable irreversible enzyme inhibition. At each of Baker's four steps in inhibitor design, correlation analysis, coupled with cluster analysis for substituent selection,[3] can play a crucial role in drug development at the enzymic level.

The role of substituents in metabolism[5] and the random walk process must also be considered in making the transition from in vitro work to in vivo whole animal studies. It is clear from Baker's triazine study[8] that gaining more inhibitory in vitro activity by increasing MR is not likely to be valuable in whole animal studies. MR seems to model the most nonspecific kind of interaction between enzyme and ligand. Baker did not distinguish[2a] clearly between hydrophobic and polar areas; indeed, this is difficult to do even using regression analysis.

One must make maximum use of the directional nature[23] of hydrophobic binding to ensure maximum interaction between ligand and pathogen enzyme and, if possible, minimize this type of interaction with the host enzyme. When maximum hydrophobicity has been attained in one part of the inhibitor, one must attach polar groups which will fall in the polar space of the enzyme or project into the aqueous phase to counterbalance the overall hydrophobicity of the potential drug. There are few examples where log $P_0$ for a set of drugs exceeds 4 (in vivo).

### References and Notes


(1) (a) This investigation was supported by Public Health Service Research Grant No. CA-11110 from the National Cancer Institute. (b) NATO Fellow and Visiting Professor of Chemistry from the Institute of Pharmaceutical and Toxicological Chemistry, University of Naples, Italy.
(2) (a) B. R. Baker, "Design of Active-Site-Directed Irreversible Enzyme Inhibitors", Wiley, New York, N.Y., 1967. (b) J. Burchall, *J. Infect. Dis.*, 128, S437 (1973).
(3) C. Hansch, S. H. Unger, and A. B. Forsythe, *J. Med. Chem.*, 16, 1217 (1973).
(4) C. Hansch and J. M. Clayton, *J. Pharm. Sci.*, **62**, 1 (1973).
(5) C. Hansch, *Drug Metab. Rev.*, 1, 1 (1972).
(6) C. Silipo and C. Hansch, *Mol. Pharmacol.*, 10, 954 (1974).
(7) C. Silipo and C. Hansch, *Il Farmaco, Ed. Sci.*, 30, 35 (1975).
(8) C. Silipo and C. Hansch, *J. Am. Chem. Soc.*, **97**, 6849 (1975).
(9) C. Hansch and M. Yoshimoto, *J. Med. Chem.*, 17, 1160 (1974).
(10) M. Yoshimoto, C. Hansch, and P. Y. C. Jow, *Chem. Pharm. Bull.*, **23**, 437 (1975).
(11) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, 16, 1207 (1973).
(12) (a) A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, 71, 525 (1971); (b) C. Hansch, A. Leo, and D. Nikaitani, *J. Org. Chem.*, **37**, 3090 (1972).
(13) (a) R. W. Taft, Jr., in "Steric Effects in Organic Chemistry", M. S. Newman, Ed., Wiley, New York, N.Y., 1956, pp 556–675; (b) E. Kutter and C. Hansch, *J. Med. Chem.*, 12, 647 (1969). (c) I. V. Talvik and V. A. Palm, *Org. React.*, **8**, 445 (1971).
(14) M. Charton, *Prog. Phys. Org. Chem.*, 8, 231 (1971).
(15) C. Daniel and F. S. Wood, "Fitting Equations to Data", Wiley-Interscience, New York, N.Y., 1971, pp 55, 169, 203.
(16) (a) B. R. Baker and W. F. Wood, *J. Med. Chem.*, **10**, 1101 (1967); (b) *ibid.*, 11, 644 (1968); (c) *ibid.*, 11, 650 (1968); (d) *ibid.*, **12**, 216 (1969); (e) B. R. Baker and H. V. Siebeneick, *ibid.*, **14**, 802 (1971).
(17) (a) B. R. Baker and W. F. Wood, *J. Med. Chem.*, 10, 1106 (1967); (b) *ibid.*, **12**, 211 (1969); (c) *ibid.*, **12**, 214 (1969).
(18) C. Hansch and W. R. Glave, *J. Med. Chem.*, **15**, 112 (1972).
(19) Reference 1, p 156.
(20) L. Pauling, *Nature (London)*, **161**, 707 (1948).
(21) (a) R. Wolfenden, *Acc. Chem. Res.*, **5**, 10 (1972); (b) G. E. Lienhard, *Annu. Rep. Med. Chem.*, **7**, 249 (1972).
(22) In this instance, by functional group we mean that minimum constellations of inhibitor atoms are necessary to produce the specific stereoelectronic interaction at the active site; this is not necessarily the same active site at which substrates react.
(23) C. Hansch, *Adv. Pharmacol. Chemother.*, 13, 45 (1975).
(24) C. Hansch and E. Coats, *J. Pharm. Sci.*, **59**, 731 (1970).
(25) L. Pauling and D. Pressman, *J. Am. Chem. Soc.*, **67**, 1003 (1945).
(26) D. Agin, L. Hersh, and D. Holtzman, *Proc. Natl. Acad. Sci. U.S.A.*, **53**, 953 (1965).
(27) C. Hansch and W. J. Dunn III, *J. Pharm. Sci.*, 61, 1 (1972).
(28) E. Ciuffarin and L. Senatore, *Tetrahedron Lett.*, 1635 (1974).


# Correlation Analysis of Baker's Studies on Enzyme Inhibition. 2. Chymotrypsin, Trypsin, Thymidine Phosphorylase, Uridine Phosphorylase, Thymidylate Synthetase, Cytosine Nucleoside Deaminase, Dihydrofolate Reductase, Malate Dehydrogenase, Glutamate Dehydrogenase, Lactate Dehydrogenase, and Glyceraldehyde-phosphate Dehydrogenase†,[1a]


Masafumi Yoshimoto[1b] and Corwin Hansch*

*Department of Chemistry, Pomona College, Claremont, California 91711.*

*Received May 5, 1975*


The inhibitory activity of 1058 inhibitors of the title enzymes has been formulated in 13 equations correlating chemical structure with inhibitory potency. Two types of regions in enzymes have been defined by means of $\pi$ and molar refractivity constants. The use of indicator variables has been extensively developed to suggest special enzyme–ligand interactions. Several examples are given of the use of correlation equations in comparing structural features of different systems.


In the first paper in this series,[2] five correlation equations were presented which relate chemical structure and inhibitory activity of 578 reversible inhibitors of guanine deaminase, xanthine oxidase, dihydrofolate reductase, and complement. In this paper, most of the rest of Baker's studies during the period 1964–1972 are correlated by 13 equations describing the QSAR for 1053

† This paper is dedicated to the memory of Edward Smissman and Bernard R. Baker.