Molecular Graphics and QSAR in the Study of Enzyme–Ligand Interactions. On the Definition of Bioreceptors

CORWIN HANSCH*† and TERI E. KLEIN‡

Department of Chemistry, Pomona College, Claremont, California 91711, and Computer Graphics Laboratory, Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Received March 24, 1986 (Revised Manuscript Received July 15, 1986)

Introduction

The use of model systems to extend our understanding of structure–activity relationships (SAR) has a long history originating with Meyer and Overton in the last century. Independently, they suggested the use of oil–water partition coefficients to model the partitioning of simple neutral organic compounds into nerve tissues and thus provided a mechanism for the study of narcotic action. A great advance was made by Hammett[1] about 1935 when he formulated σ, from the ionization constants of benzoic acids, to account for the electronic effect of substituents on organic reaction rates and equilibrium constants. Following this line of thinking, Taft[2] defined E_s for the steric effects of substituents, and Pauling and Pressman[3] proposed molar refractivity (MR) for the correlation of dispersion forces of substituents in biochemical reactions. Leffler and Grunwald have provided a scholarly generalization of this approach to the study of organic reaction mechanisms.[4]

In the early sixties the octanol–water hydrophobic parameters were formulated,[5] and this provided the missing link for attacking biomedical SAR in quantitative terms (QSAR).[6] Fortuitously, large computers were becoming generally available, and the stage was set to use the linear combination of numerical parameters in a statistical assault on biological SAR problems that had remained largely intractable.

Another less abstract line of model building which might be said to have come into its own with Pauling[7] and the Pauling–Corey models, and achieved a spectacular success with Watson and Crick,[8] is the building of solid 3-dimensional models as an aid in understanding structure–function relationships. The major advance for model building of macromolecules is real-time interactive stereo computer graphics.[9] The recent addition of color and dot surface representation[10,11] opens the possibility for combining two different model systems to serve as independent checks in the study of the SAR of biochemical reactions when the X-ray crystallographic structure of the bioreceptor is known. The statistical correlation equations are based on enzymic kinetic parameters from reactions occurring in solution, while the molecular graphics models are based on information obtained from the macromolecule in the crystalline state. Nevertheless, in the past few years, we have found good qualitative agreement between models derived from the two vastly different systems.

Since its initiation[6] QSAR has grown rapidly so that thousands of equations have been published (see ref 12 for examples), and it has been widely used in the study and design of bioactive compounds.[13] Although a large variety of parameters are now employed in QSAR studies,[14] this account will be limited to a discussion of σ, MR, and the hydrophobic parameters log P and π from the octanol–water system, in the study of enzyme–ligand reactions.

The Hammett σ constant is defined as: $\sigma = \log K_x - \log K_H$, where K_H is the ionization constant of benzoic acid in water and K_x is that of a substituted benzoic acid. A positive sign with σ indicates electron withdrawal by the substituent. The hydrophobic parameter π is defined analogously using the octanol–water partition coefficient P: $\pi = \log P_x - \log P_H$, where P_H is the partition coefficient for benzene and P_x is that for a substituted benzene. Molar refractivity is defined as $MR = ((n^2 - 1)/(n^2 + 1)(MW/d)$ where n is the index

Corwin Hansch received his undergraduate education in chemistry at the University of Illinois and his Ph.D. in Organic Chemistry from New York University in 1944. After working with the Du Pont Company, first on the Manhattan project and then in Wilmington, DE, he joined the Pomona College faculty in 1946. The present review is the outgrowth of the study of the chemical–biological interactions of plant growth regulators commenced at that time with Professor R. M. Muir, then at Pomona, now at the University of Iowa.

Teri Klein graduated from the University of California at Santa Cruz and is now a member of the Medical Information Sciences Department and a doctoral candidate in the Computer Graphics Laboratory at the University of California in San Francisco.

† Pomona College
‡ University of California.

(1) Hammett, L. P. *Physical Organic Chemistry*, 2nd ed.; McGraw-Hill: New York, 1970.
(2) Taft, R. W. *J. Am. Chem. Soc.* **1953**, *75*, 4538.
(3) Pauling, L.; Pressman, D. *J. Am. Chem. Soc.* **1945**, *67*, 1003.
(4) Leffler, J. E.; Grunwald, E. *Rates and Equilibria of Organic Reactions*; Wiley: New York, 1963.
(5) Fujita, T.; Iwasa, J.; Hansch, C. *J. Am. Chem. Soc.* **1964**, *86*, 5175.
(6) (a) Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, M. M. *Nature (London)* **1962**, *194*, 178. (b) Hansch, C. *Acc. Chem. Res.* **1969**, *2*, 232.
(7) Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, NY, 1960.
(8) Watson, J.; Crick, F. H. *Nature (London)* **1953**, *171*, 964.
(9) Langridge, R. *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **1974**, *33*, 2332.
(10) Langridge, R.; Ferrin, T. E.; Kuntz, I. D.; Connolly, M. L. *Science* **1981**, *211*, 661.
(11) Bash, P. A.; Pattabiram, N.; Huang, C.; Ferrin, T. E.; Langridge, R. *Science* **1983**, *222*, 1325.
(12) (a) Hansch, C. In *Drug Design*; Arëins, E. J., Ed.; Academic: New York, 1971; Vol. 1, p 271. (b) *Quantitative Structure–Activity Relationships of Drugs*; Topliss, J. G., Ed.; Academic: New York, 1983. (c) Dunn, W. J., III; Hansch, C. *Chem.-Biol. Interact.* **1974**, *9*, 75. (d) Hansch, C.; Clayton, J. M. *J. Pharm. Sci.* **1973**, *62*, 1. (e) *QSAR and Strategies in the Design of Bioactive Compounds*; Seydel, J. K., Ed.; VCH Verlagsgesellshaft: Weinheim, Germany, 1985. (f) *Biological Activity and Chemical Structures*; Buisman, J. A. K., Ed.; Elsevier: Amsterdam, 1977. (g) *Quantitative Structure–Activity Relationships*; Tichy', M. Ed.; Birkhauser Verlang: Basel, Switzerland, 1976. (h) *Quantitative Approaches to Drug Design*; Dearden, J. C., Ed.; Elsevier: Amsterdam, 1983. (i) *Structure–Activity Correlation as a Predictive Tool in Toxicology*; Goldberg, L., Ed.; Hemisphere: Washington, DC, 1983. (j) *QSAR in the Design of Bioactive Compounds*; Kuchar, M., Ed.; J. R. Prous International: Barcelona, Spain, 1984. (k) The Journal, *Quantitative Structure–Activity Relationships* published by Verlag Chemie since 1982 is devoted to the subject. In addition it publishes abstracts of QSAR from all sources.
(13) Hopfinger, A. J. *J. Med. Chem.* **1985**, *28*, 133.
(14) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley-Interscience: New York, 1979.

0001-4842/86/0119-0392$01.50/0

of refraction, MW is molecular weight and d is density. Since n does not vary much for most organic compounds MR is primarily a measure of volume.

In the phenomenological statistical approach of QSAR, one needs to categorize an enzyme surface as simply as possible. The greater the number of categories the greater the number of probe compounds (substrates or inhibitors) one needs to establish their significance. In correlation analysis it is generally accepted that one needs five well spread data points, as a *minimum*, for each term in an equation. Thus one is caught in the dilemma of either attempting to use too few parameters and getting an oversimplified picture or using so many parameters that it becomes impractical to synthesize and test compounds to justify them.

To data we have found that hydrophobic parameters from the MR and octanol–water system give a rough characterization of the polar and hydrophobic patches of a protein surface.[15] In our first efforts to correlate QSAR-defined surface with X-ray crystallographically defined surface via graphics we colored hydrophobic space red (for carbon) and hydrophilic space blue for nitrogen and oxygen surfaces. We have now found that an intermediate type of surface, coded yellow in this report, appears to correlate with MR. The intermediate surfaces C_α, C_2 in histidine, C in C=O, and C located within 3 Å of the ϵ-NH_2 of lysine or C in the guanidine group of arginine. From the partitioning of model compounds it is concluded that these carbon atoms are not typically hydrophobic.[15j,k] Attaching carbon to electron withdrawing groups reduces its hydrophobic character.[16]

There still remains much confusion about the role of van der Waals forces (including steric as well as dispersion attraction effects) and hydrophobic forces in enzymic interactions with organic compounds. Recently, experimental work by Pashley et al.,[17] has shown that hydrophobic forces are 10 to 100 times stronger than the van der Waals forces that would operate in the absence of any surface-induced structure in water. The magnitude of the attractive interaction is a function of the degree of hydrophobicity of the surface. Thus it seems reasonable that log P (or π) and MR parameters might serve to diagnose pure hydrophobic surfaces and polar surfaces; how useful they will be in the discrimination of surfaces of intermediate character remains to be seen.

Chart I

View I. Wire model in blue with 4-hexyloxycarbonylsulfanilamide bound to carbonic anhydrase. Dotted line shows potential for hydrogen bonding between carbonyl of ester and NH of Gly-92. Note slightly concave, uniform red (hydrophobic) surface where alkyl group is approximately 50% desolvated.

View II. 4-Hexylpyrazole (with green van der Waals surface) binding in deep pocket in alcohol dehydrogenase. Blue wire model in background is NAD. Note very close fit of hexyl chain to surface.

View III. Side view of alcohol dehydrogenase with 4-hexylpyrazole showing tight fit of alkyl chain to surface. Front side of the pocket is essentially all red and has been deleted. A water molecule in purple on the same scale shows that desolvation would have to be complete for the hexyl group to fit into the cavity.

View IV. Shows green wire model of V with X = Cl. Yellow dotted line shows hydrogen bonding between ester carbonyl and OH of Ser-195. Note that the Cl falls on yellow space defined by the disulfide bridge between Cys-42 and Cys-48.

Naturally, not just any solvent will serve to define hydrophobicity,[18] but as eq 1 shows, octanol does reasonably well in correlating the protein binding of organic substances.[19] In this expression, C is the molar concentration of organic compound producing a 1:1 molar complex with bovine serum albumin. Equation 1 is based on 42(n) neutral compounds including 20 phenols, 6 anilines, bulky compounds such as neopentyl alcohol, camphorquinone, and hydroxyadamantane, and compounds with little hydrogen bonding ability: nitrobenzene, naphthalene. The correlation coefficient is represented by r, and s is the standard deviation.

$$\log 1/C = 0.75 \log P + 2.30 \quad (1)$$

$$n = 42 \qquad r = 0.960 \qquad s = 0.159$$

In our approach, electronic effects of substituents are those modeled by σ where σ and π (or log P) are independent variables (orthogonal), and when MR is also orthogonal, it represents steric bulk and van der Waals type forces. Hence, in using this approach care must be taken to see that collinearity among variables is minimized.

Carbonic Anhydrides Inhibition by Sulfonamides

Carbonic anhydrase catalyzes the hydration of CO_2 to H_2CO_3 and is an important enzyme in medicinal chemistry as well as biochemistry.[20] Equation 2 has been derived from the binding constants (K) of sulfonamides I (X-C_6H_4-SO_2NH_2) with human carbonic anhydrase.[15i] The indicator variable I_1 takes the value of 1 for meta substituents (0 for all others) and I_2 is assigned 1 for ortho substituents. All ortho and meta substituents were of the type COOR where R is limited to normal alkyl groups. From the negative coefficients of I_1 and I_2 it was assumed that they accounted for

$$\log K = 1.55\sigma + 0.64 \log P - 2.07I_1 - 3.28I_2 + 6.94 \quad (2)$$

$$n = 29 \qquad r = 0.991 \qquad s = 0.204$$

(15) (a) Smith, R. N.; Hansch, C.; Kim, K. H.; Omiya, B.; Fukumura, G.; Selassie, C. D.; Jow, P. Y. C.; Blaney, J.; Langridge, R. *Arch. Biochem. Biophys.* **1982**, *215*, 319. (b) Li, R. L.; Hansch, C.; Matthews, D.; Blaney, J. M.; Langridge, R.; Delcamp, T. J.; Susten, S. S.; Freisheim, J. H. *Quant. Struct.-Act. Relat.* **1982**, *1*, 1. (c) Hansch, C.; Li, R. L.; Blaney, J. M.; Langridge, R. *J. Med. Chem.* **1982**, *25*, 777. (d) Hansch, C.; Hathaway, B. A.; Guo, Z. R.; Selassie, C. D.; Dietrich, S. W.; Blaney, J. M.; Langridge, R.; Volz, K. W.; Kaufman, B. T. *J. Med. Chem.* **1984**, *27*, 129. (e) Carotti, A.; Smith, R. N.; Wong, S.; Hansch, C.; Blaney, M., Langridge, R. *Arch. Biochem. Biophys.* **1984**, *229*, 112. (f) Hansch, C.; Blaney, J. M. In *Drug Design: Fact or Fantasy?* Jolles, G., Wooldridge, K. R. H., Eds.; Academic: New York, 1985; p 185. (g) Carotti, A.; Hansch, C.; Mueller, M. M.; Blaney, J. M. *J. Med. Chem.* **1984**, *27*, 1401; **1985**, *28*, 261. (h) Blaney, J. M.; Hansch, C.; Silipo, C.; Vittoria, A. *Chem. Rev.* **1984**, *84*, 333. (i) Hansch, C.; McClarin, J.; Klein, T.; Langridge, R. *Mol. Pharamacol.* **1985**, *27*, 493. (j) Recanatini, M.; Klein, T.; Yang, C. Z.; McClarin, J.; Langridge, R.; Hansch, C. *Mol. Pharamacol.* **1986**, *29*, 436. (k) Recanatini, M.; Klein, T.; Langridge, R.; Hansch, C., unpublished results. (l) Hansch, C.; Klein, T.; McClarin, J.; Langridge, R.; Cornell, N. W. *J. Med. Chem.* **1986**, *29*, 615. (m) Hansch, C.; Smith, R. N.; Rockoff, D.; Calef, D. F.; Jow, P. Y. C.; Fukunaga, J. Y. *Biochem. Biophys. Acta* **1977**, 183. (n) Carotti, A.; Ragueso, C.; Hansch, C. *Quant. Struct.-Act. Relat.* **1985**, *4*, 145. (o) Unpublished results.

(16) (a) Reference 14, p 39. (b) Hansch, C.; Leo, A.; Nikaitani, C. *J. Org. Chem.* **1972**, *37*, 3090.

(17) Pashley, R. M.; McGuiggan, P. M.; Ninham, B. W.; Evans, D. F. *Science* **1985**, 229, 1088.

(18) Smith, R. N.; Hansch, C.; Ames, M. M. *J. Pharm. Sci.* **1975**, *64*, 599.

(19) Helmer, F.; Kiehs, K.; Hansch, C. *Biochemistry* **1968**, *7*, 2858.

(20) *Ann. N. Y. Acad. Sci.* **1984**, *429*, 1–640.

unfavorable contacts of the carboxylate group after which the rest of the R group shows the normal hydrophobic interaction characterized by 0.64 log P. The positive coefficient with σ shows that electron withdrawal by X favors binding. This makes good sense since there is evidence that it is the ionized form of SO_2NH_2 that binds to the positively charged Zn atom in carbonic anhydrase; electron withdrawal would favor ionization.

After the derivation of eq 2, construction of the computer graphics model of 4-CH_3-$(CH_2)_5OCOC_6H_4SO_2NH_2$ (blue wire model in View I, Figure 1) bound to carbonic anhydrase confirmed our expectations. Most striking was the large red surface (without any blue or yellow dots to signify nonhydrophobic area) upon which the alkyl chain falls. This model, as well as all of the others in this account, was constructed from the X-ray coordinates of a similar inhibitor (in this case a sulfonamide) bound to the enzyme which fixes the position of I except for X. Hydrogen atoms are not shown in the protein but their own van der Waals radii are taken into account in constructing the surface. The surface is the "solvent-accessible" surface and is defined by causing a spherical probe of 1.4 Å to move over the atoms of the protein and place a dot at each contact point of the sphere with the van der Waals surface of the protein. A smoothed out surface results in which the interstices too small to accept the probe are eliminated.

To actually see the beautiful red hydrophobic region in view I demanded by eq 2, after so many years of wondering what real meaning our QSAR parameters had at the molecular level, was to experience an enormous sigh of relief. Those of us who work in the impossibly complex area of biological SAR know all too well what Einstein meant in his jingle:

A thought that sometimes makes me hazy:
Am I—or are the others crazy?

View I shows the ligand in the form of a blue wire model where deep in the pocket the SO_2NH_2 binds to Zn. The yellow dotted line shows the possibility for hydrogen bonding between Gln-92 (NE2) and the carbonyl of an ester or an amide in the para position of the sulfonamide; the fact that such carbonyl moieties are well fit without any correction for hydrogen bonding indicates it not to be a significant factor.

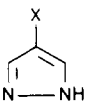
Assuming the hydrophobic effect to be equivalent to the free energy of desolvation of one side of the alkyl chain and of the equivalent amount of protein surface one might expect this to be roughly equivalent to the amount of water involved in the hydrophobic hydration of the alkyl chain. This would be about the amount of water contact removed from the alkyl group in partitioning into octanol. Hence, if octanol were the "perfect" model one might expect a coefficient of 1 with log P in eq 2. In fact, it is about half that figure. We have found coefficients with log P or π near 0.5 in other instances where a more or less flat hydrophobic protein surface is involved, which would imply that the 0.5 coefficient may be a marker of such interactions.[15e,g]

Study of the graphics model shows that the meta carboxylate groups suffers a poor contact with Leu-198, and the ortho carboxylate moiety is sterically hindered by Pro-201 which nicely accounts for I_1 and I_2 in eq 2. Despite these initial bad contacts the R groups bind to a hydrophobic surface with the same free energy change/CH_2 as para substituents.[15i]

Inhibition of Alcohol Dehydrogenase

Alcohol dehydrogenase (ADH) is an NAD dependent system that mediates the interconversion of alcohols and aldehydes. QSAR for inhibitors have been obtained in terms of their inhibition constants (K_i) as follows:[15l]

Pyrazole (II) Inhibition of Rat Liver ADH



II

$$\log 1/K_i = 1.22 \log P - 1.80\sigma_{meta} + 4.87 \quad (3)$$

$n = 14 \quad r = 0.985 \quad s = 0.316$

Pyrazole Inhibition of Human Liver ADH

$$\log 1/K_i = 0.87 \log P - 2.06\sigma_{meta} + 4.60 \quad (4)$$

$n = 13 \quad r = 0.977 \quad s = 0.303$

Pyrazole Inhibition of Horse Liver ADH

$$\log 1/K_i = 0.96 \log P + 5.70 \quad (5)$$

$n = 5 \quad r = 0.990 \quad s = 0.207$

Inhibition of Horse Liver ADH by Amides III $XC_6H_4CH_2CONH_2$ (III)

$$\log 1/K_i = 0.89 \log P + 3.56 \quad (6)$$

$n = 11 \quad r = 0.960 \quad s = 0.197$

Inhibition of Horse Liver ADH by Amides IV XCH_2CONH_2 (IV)

$$\log 1/K_i = 0.98 \log P - 0.83\sigma^* + 3.69 \quad (7)$$

$n = 14 \quad r = 0.937 \quad s = 0.280$

The common feature of the above equations is that the coefficient with log P is near 1. On the basis of eq 2 we might be led to expect that complete desolvation of X in hydrophobic space is occurring on binding to ADH, assuming, as usual, that binding as modeled by log P is straight forward with no subtle compensation processes occurring.[21] Equation 5 is the best guide in this respect since it is based on examples were X represents only normal alkyl groups (hence no electronic term is needed). In a number of instances it was found that branched groups are not well fitted depending on their distance from the functional group.[15l]

In eq 3, 4, and 7 the negative coefficients with the electronic terms show that electron release to the attaching functional group increases binding strength. The Hammett constants in eq 3 and 4 are for the meta position. This is in line with expectations for the electron pair on pyrazole nitrogen orthogonal to the pi electron system to be involved in the attachment to Zn. No electronic term occurs in eq 6, in part because of the insulating effect of the CH_2 moiety, but also because a rather small range in σ constants was studied. In eq 7, Taft's σ^* parameter for aliphatic substituents has been employed. Since it appears to be the lone-pair electrons on N in II and the electrons of carbonyl oxygens in III and IV which play a crucial role in the binding to Zn, it is found, as expected, that electron

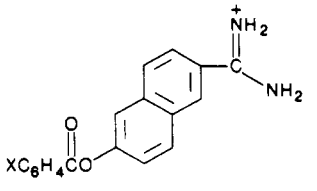
(21) Lumary, R.; Rajender, S. *Biopolymers* **1970**, *9*, 1125.

release by X increases the affinity of the inhibitors. Since the data for the above five equations comes from three different laboratories, on three different types of compounds, interacting with ADH from three different sources, the results offer convincing evidence that we can build reliable mathematical SAR models via substituent constants and regression analysis.

Stereoviews II and III, constructed from the X-ray crystallographic coordinates of ternary ADH–pyrazole–NAD complex, show the wire model of 4-hexylpyrazole. The end on view II reveals how tightly circumscribed the fit of the hexyl moiety is in the hydrophobic hole. (Note the close fit of the green van der Waals surface of the ligand to the red hybrophobic surface of the enzyme). In the side view III on the right of the picture the pyrazole N is shown binding to Zn and in blue is the NAD cofactor. The side of the hydrophobic cave nearest to the viewer has been removed for clarity so that red dots visible are only those on the back side of the pocket. The fit is tight for the length of the cavity although at some points some branching can be tolerated.[15l] On the top in view III is a water molecule in purple with its van der Waals surface on the same scale as the active site. Clearly there is no room for such a large substance in the active site along with the pyrazole so that desolvation of the ligand on entering the cavity must be complete.

Inhibition of Trypsin by Amidines[15j]

A recent study of amidine inhibition of the hydrolase, trypsin, yielded eq 8 for inhibitors V.



V

$$\log 1/K_i = -1.40\sigma + 0.47MR_4 + 2.59 \quad (8)$$

$$n = 21 \quad r = 0.915 \quad s = 0.322$$

The negative coefficient with σ shows that electron release by X favors inhibition (binding). No dependency of inhibition on the hydrophobic properties of X could be found, but MR_4 is a significant parameter which accounts for about 20% of the variance in log $1/K_i$. Correlation with MR rather than π suggests that binding is occurring on a polar- or intermediate-type enzyme surface. Meta substituents are not parameterized by either MR or π; their only effect is electronic, correlated via σ. Therefore it was assumed that meta substituents do not contact the enzyme.

View IV, shows the wire model of V with X = Cl. A dotted yellow line from the carbonyl group of the ester linkage to the OH of Ser-195 indicates the possibility for hydrogen bonding between these two moieties. This would account for the negative coefficient with the σ term in eq 8. That is, the higher the electron density on the carbonyl oxygen, the better it would bind to the H of the serine OH. The distance between the two as they are seen in view IV is 2.7 Å. The phenyl portion of the ester is twisted 44° out of coplanarity with the naphthyl ring, and this allows the phenyl group to make better contact with the hydrophobic surface behind it and provides a better hydrogen bonding distance (2.7 Å). In view IV it is clear that there is not space for meta substituents to fit between the ring and the enzyme surface. Rotation of the phenyl ring 180° can place meta substituents in two, possibly different, types of space. In the present example, meta substituents must be held away from the enzyme surface; this explains the need to set MR for meta substituents to 0.1, the value of H.

A most interesting feature of view IV is that small substituents are seen to fall onto the yellow surface of the disulfide bridge between cysteines 42 and 58. Thus it appears that binding to the intermediate type of enzyme surface (–S–S–) may correlate better with MR than with π. The forces involved would be dispersion with the possibility of some weak hydrophobic contribution.

Hydrolysis of Ester VI and VII by Papain

The cysteine hydrolase, papain, is an extensively studied enzyme whose X-ray crystallographic structure has been well defined (1.65 Å resolution). For the hydrolysis of esters VI and VII the QSAR 9 and 10 have been derived in terms of their Michaelis constants (K_m).

Hydrolysis of $XC_6H_4OCOCH_2NHCOC_6H_5$ (VI) by Papain[15a]

$$\log 1/K_m = 1.03\pi'_3 + 0.57\sigma + 0.61MR_4 + 3.80 \quad (9)$$

$$n = 25 \quad r = 0.907 \quad s = 0.208$$

Hydrolysis of $XC_6H_4OCOCH_2SO_2CH_3$ (VII) by Papain[15e]

$$\log K_m = 0.61\pi'_3 + 0.55\sigma + 0.46MR_4 + 2.00 \quad (10)$$

$$n = 32 \quad r = 0.945 \quad s = 0.178$$

In eq 9 and 10, π'_3 refers only to the more hydrophobic of the two meta substituents. For example, in the derivation of these equations, if X = 3-OH (π = –0.67), then the π value used is that for H (0). For the 3,5-dichloro congener the π value for only one Cl is used. Empirically it was found that only by this procedure could a good equation be formulated. The thought that evolved from this result was that only one meta substituent at a time can contact hydrophobic space on the enzyme; the other meta substituent is, preforce, placed in aqueous space. The MR_4 term applies only to the 4-substituents, and it cannot be replaced by π; hence it was assumed that 4-substituents contact polar space. The positive coefficient with MR suggests that the larger the X-4 the better the binding in the ES complex. The positive coefficient with σ shows that electron withdrawal by X makes for better binding, no doubt facilitating the tetrahedral complex formation with SH of Cys-25.

Since congeners VI are not very soluble, which limits the use of large hydrophobic X, congeners VII were studied to develop eq 9. This QSAR was a surprise to us since we had expected to find a coefficient of about 1 with π'_3 as in eq 9. View V (Figure 2) shows how this problem can be rationalized. The phenoxy end of VI is shown with two meta chloro substituents with their attendant van der Waals surface. The purple chloro group is placed in a shallow hydrophobic pocket, better visualized in an earlier model,[15a] where the small substituents upon which Eq 9 is based are essentially completely desolvated. Rotation of the green phenoxy ring directs the chloro group toward a large hydrophobic

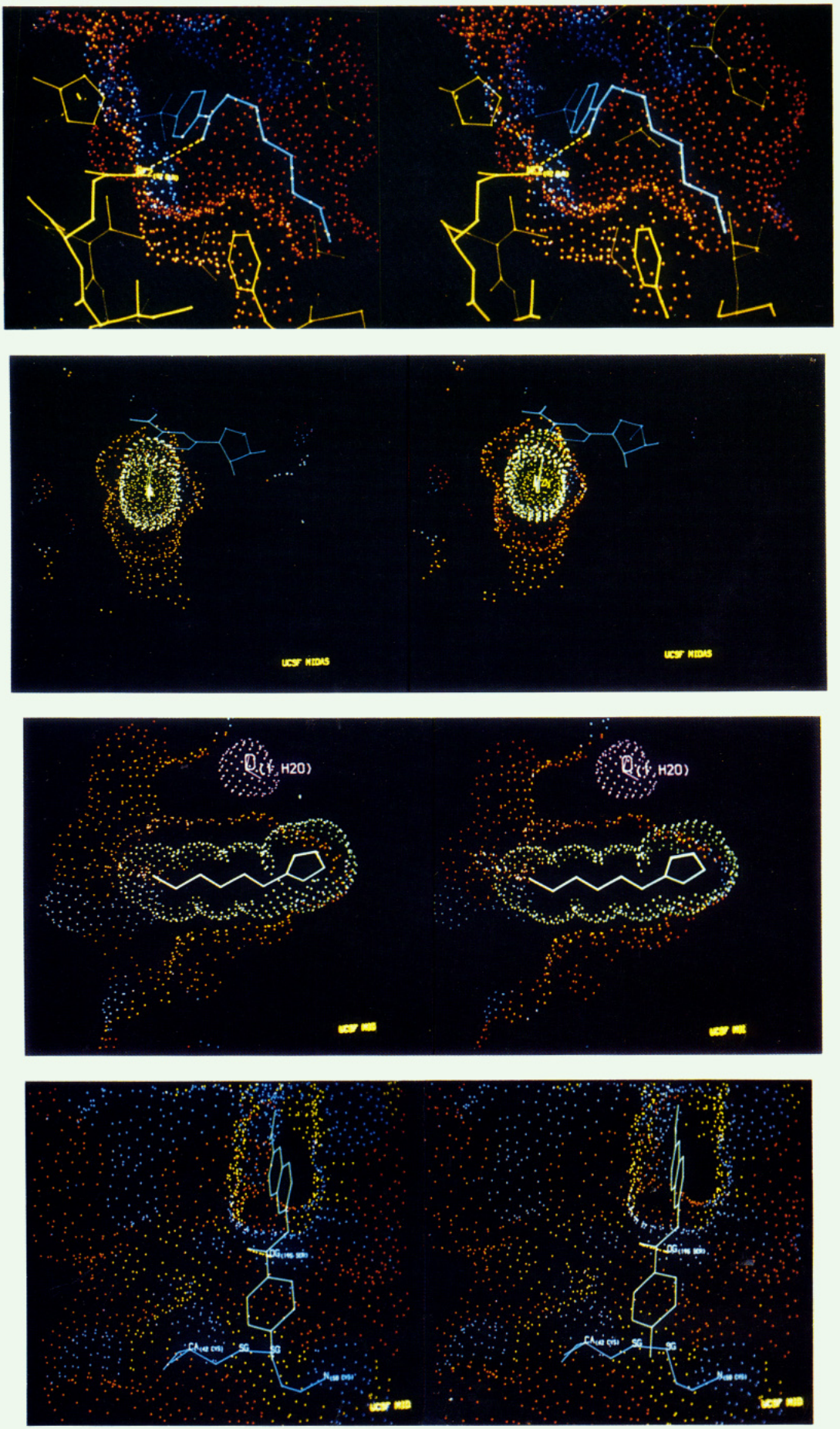


Figure 1. Stereoviews I–IV. See Chart I for stereoview description.

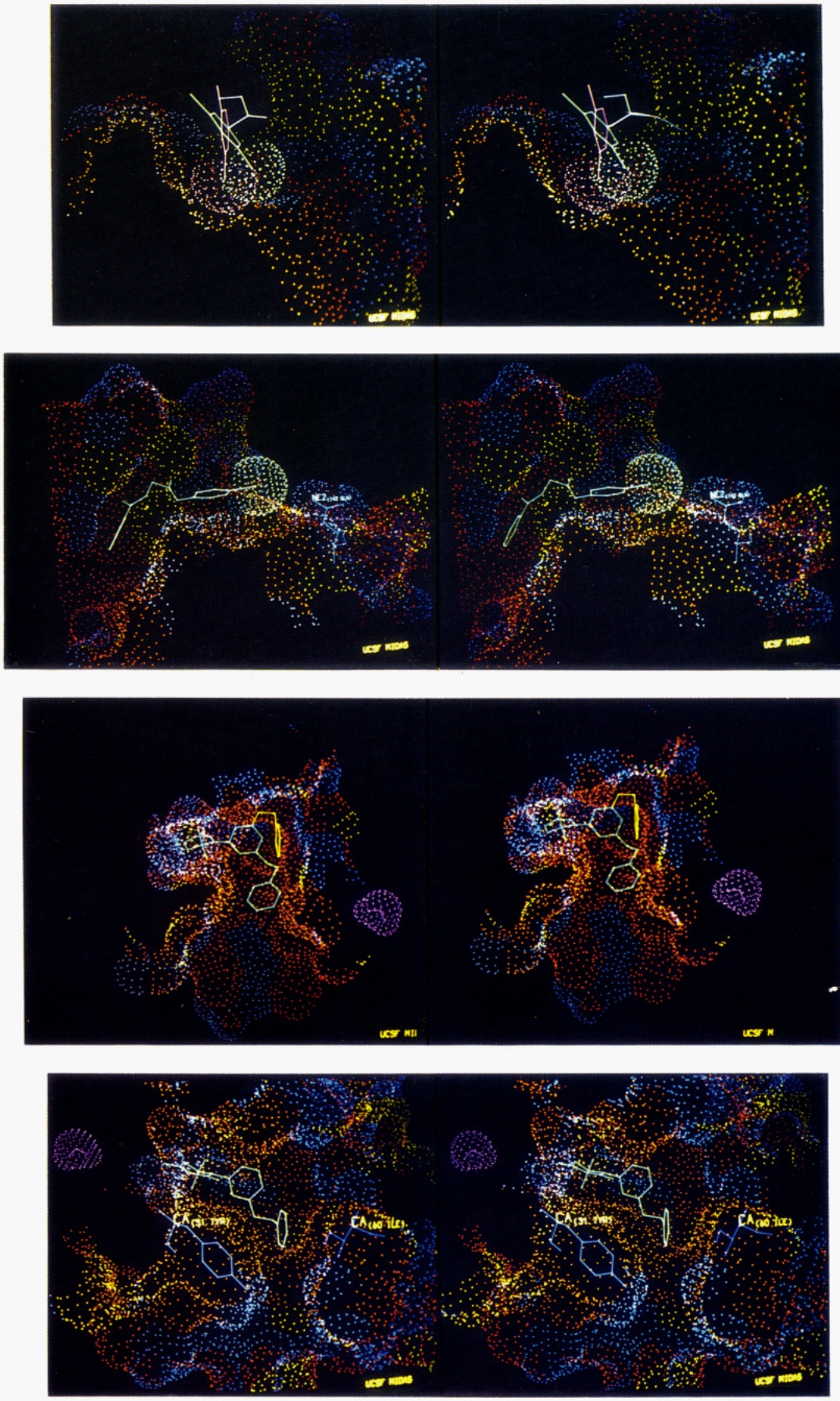


Figure 2. Stereoviews V–VIII. See Chart II for stereoview description.

plane. A different view of this site has previously been published.[15e] In each of these positions the second chloro group projects away from the enzyme surface which accounts for π_3' being set to zero for small hy-

Chart II

View V. Two end-on views of 2-$ClC_6H_4OCOCH_2NHCOC_6H_5$ bound to papain V. The Cl with the purple surface is rotated so as to bind in a shallow pocket (see ref 15a for alternative view). The Cl with the green surface is rotated to bind on the edge of large hydrophobic plane to the right. (See ref 15e for view of 3-oxymethylene-2′-naphthyl binding to this surface). The glutamine side-chain in the foreground is cut away.

View VI. Side view of 4-$ClC_6H_4OCOCH_2NHCOC_6H_4$ bound to papain as in View V. Green surface of the amide of the Gln-142 side chain. This is presumed to account for the MR_4 term in eq 9.

View VII. *L. casei* DHFR with triazine VIII where X, the *m*-$CH_2OC_5H_5$ group, is shown in two possible binding modes. On the right side in purple is a water molecule with a van der Waals surface on the same scale. Note that substituents in meta or para positions of yellow substituent would not contact the enzyme.

View VIII. Meta-substituted phenyltriazine binding in the active site of chicken DHFR. The phenyl in the substituent falls between the two hydrophobic residues Tyr-31 and Ile-60. Substituents in the meta and para positions of this phenyl group do not contact the enzyme surface.

drophilic groups. The large hydrophobic plane on the right side of the picture can easily accommodate substituents as large as 3-$OCH_2C_6H_4$-4′-Cl or 3-oxymethylene-2′-naphthyl.[15e] However, on the hydrophobic plane only one side of the substituent would be desolvated which would account for the lower coefficient with π'_3 in eq 10.

The question is, why does one type of binding occur in eq 9 and another in eq 10? An explanation must lie in the nature of the two different amide moieties $NHCOC_6H_5$ (π = 0.49) and $NHSO_2CH_3$ (π = −1.18). The hydrophobic cleft into which these fit must somewhat repel one and attract the other. Thus the $NHSO_2CH_3$ may not bind so deeply and for some reason this may favor rotation about the oxygen–phenyl bond to place substituents on the hydrophobic plane.

View VI is a profile of the 4-chloro substituent on the phenoxy moiety showing how it is directed toward the surface of the amide group of the Gln-142 residue. It is this flexible surface unit (−$CH_2CH_2CONH_2$) which we believe accounts for the MR_4 terms in eq 8 and 9 and similar terms in the QSAR of actinidin,[15g] ficin, bromelain B and bromelain D.[22] In the case of actinidin the amino unit of Lys-145 corresponds to the amide of Gln-142 in papain. In view VI the two surfaces do not make direct contact; however, we believe that since the amide group (or ϵ-NH_2 of Lys in the case of actinidin) would be strongly hydrogen bonded to water, its effective size would be larger than that shown in view VI. The positive steric effect of Gln-142 (modeled by MR_4) would appear to be the result of a buttressing action on the substrate which helps to hold it in place in the ES complex. Evidence is accumulating to show that small changes in the placement of reactive groups or factors which increase their time of contact can have huge effects on the rate of organic reactions.[23]

In our earlier graphics models of papain,[15a,e] the carbon atom of the ester group of the substrate was shown in the sp^2 state which may be unlikely for the ES complex. Although it makes only a slight difference in how the phenoxy substituents contact the enzyme, in the models in this report the ester carbon is in the tetrahedral state bonded to S of Cys-25.

Table I

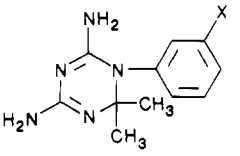
ρ^a	enzyme	congener	ref
0.56	papain	VI	15a
0.55	papain	VII	15e
0.57	ficin	VI	24
0.62	ficin	VII	22
0.74	actinidin	VI	15g
0.70	bromelain B	VI	22
0.68	bromelain D	VI	22
0.50	subtilisin	VI	15n
0.42	chymotrypsin	VI	15o

[a] 0.59 av.

The coefficients with σ (ρ) in eq 9 and 10 are in close agreement after correction has been made for the MR and π effect of the substituents which are the result of quite different sets of substituents. The ρ value for the uncatalyzed hydrolysis of VI in buffer at the same pH (6) is much larger, 1.91, showing that the enzyme takes over much of the job from the substituents of cleaving the phenoxy moiety. This statement can be made since k_{cat} for congener VI and VII is essentially constant.[15e] The following values for ρ have been obtained from our QSAR for the action of various hydrolases on VI and VII.

Inhibition of Dihydrofolate Reductase (DHFR) by Triazines VIII

An enormous amount of study has been devoted to the inhibition of dihydrofolate reducase (DHFR) since the enzyme is essential for the synthesis of DNA and



VIII

since two inhibitors, methotrexate and trimethoprim, are valuable drugs in cancer and antibacterial chemotherapy respectively.[15h]

Equation 11 correlates[15d] the inhibition of DHFR from *L. casei* cells using the bilinear model[25] for π. In

$$\log 1/K_i = 0.83\pi'_3 - 0.91(\beta \times 10^{\pi'_3} + 1) + 0.71I + 4.60 \quad (11)$$

$$n = 38 \qquad r = 0.961 \qquad s = 0.244 \qquad \pi_0 = 2.69$$

this model inhibitor potency increases linearly with slope of 0.83 π' until the point of π_0 where a second linear relationship sets in with slope of 0.83 − 0.91 = −0.08. Thus the right side of the bilinear curve has a slope of essentially 0 showing that there is no significant hydrophobic effect beyond certain size substituents (π_0). For all of the usual substituents the normal π value is used, but for substituents of the type 3-$CH_2ZC_6H_4Y$ (Z = O, NH, S, Se) Y had essentially no effect on K_i, which suggested that Y did not contact the enzyme. Therefore π for Y was set equal to zero.

In eq 11, I is an indicator variable which is given the value of 1 for all substituents containing the −$CH_2ZC_6H_4$− moiety and 0 for all other substituents. This structural feature increases activity beyond that projected by π alone by a factor of about 5. Why this

(22) Carotti, A.; Raguseo, C.; Hansch, C. *Chem.-Biol. Interact.* **1985**, *52*, 279.
(23) Menger, F. M. *Acc. Chem. Res.* **1985**, *18*, 128.
(24) Carotti, A.; Casinii, G.; Hansch, C. *J. Med. Chem.* **1984**, *27*, 1427.
(25) Kubinyi, H.; Kehrhahan, O. H. *Arzneim.-Forsch.* **1978**, *28*, 598.

effect occurs is not clear. The disposable parameter β is estimated by an iterative procedure since the least squares method cannot be used for nonlinear equations.[25]

In view VII is displayed the active site of *L. casei* DHFR containing triazines VIII with two possible binding modes for a meta $CH_2OC_6H_5$ group. We favor binding as shown by the yellow group since neither hydrophilic nor hydrophobic meta or para substituents in the second ring have much affect on log $1/K_i$. The coefficient with π of near 1 (0.83) is indicative of complete desolvation of X, and indeed it is difficult to see how a water molecule could do more than contact the edge of the outer yellow phenyl group. Uniformly hydrophobic substituents of $\pi > \pi_0$ have essentially no effect on log $1/K_i$ as shown by −0.08 slope of the right side of the bilinear model. However this rough parameterization of lengthy groups will not work for substituents of the type $CH_2Z(CH_2)_nCH_3$ where Z is large and hydrophilic and n is large enough so that the substituent projects beyond the enzyme surface. Nor will it do for substituents of the type $-(CH_2)_nZ$ where n is large enough so that a very hydrophilic Z projects beyond the enzyme. Ideally what is needed are atomic hydrophobic parameters. While these are available for simple substituents such as −O−, −NH−, Cl, etc., it is not obvious how complex substituents such as $-CONH_2$, SO_2NH_2, $NHCONH_2$, etc. can be broken into atomic contributions.

Inhibition of Enzymes in Vivo

One of the important unanswered questions in biochemistry is do enzymes in vivo behave like purified enzymes in dilute buffer solution? Beece et al.[26] have shown that the viscosity of the solution affects the structure of enzymes and the way ligands react with them. This being so, one would expect enzymes in vivo to behave differently from those in buffer solution.

Equation 12 correlates the action (C = molar concentration of triazine producing 50% inhibition of growth) of triazines on wild type *L. casei* cells.[27]

$$\log 1/C = 0.83\pi'_3 - 1.06 \log((\beta \times 10\pi'_3) + 1) - 0.94MR_y + 0.80I = 4.37 \quad (12)$$

$$n = 34 \quad r = 0.929 \quad s = 0.371 \quad \pi_0 = 2.94$$

With the exception of the standard deviation and the term in MR_y the parameters of eq 11 and 12 are very similar indicating similarity of action in vivo and in vitro. The higher standard deviation of the in vivo equation suggests small differences, probably steric, which could be the result of a more tightly constrained enzyme in vivo which could inhibit access of some inhibitors relative to others.[28] The MR_y term is the molar refractivity of Y (scaled by 0.1) and is primarily a measure of its bulk. No parameterization was found for Y in the in vitro equation nor was it expected on the basis of the molecular modeling. Hence it was a great surprise to find the negative MR term which is hard to interpret other than as a steric effect. This effect may be from within the DHFR molecule and result from a different conformation of part of the enzyme in vivo, or as seems more likely, it results from contact with a nearby macromolecule. To our knowledge this is the first instance where a quite specific difference is found for the way an enzyme reacts in vitro and in vivo.

Equation 13 correalates[27] the effect of triazines on *L. casei* cells which are 10^8 times more resistant to the antifolate methotrexate than the sensitive cells of eq 12. This is a much different QSAR than that for the

$$\log 1/C = 0.42\pi'_3 + 1.09I - 0.48MR_y + 3.37 \quad (13)$$

$$n = 38 \quad r = 0.965 \quad s = 0.259$$

sensitive cells or the isolated DHFR. It was not possible to put a firm value on π_0 but it is estimated to be near 6. That is, a linear increase in potency follows a linear increase in π over 1000 times the range of that of the sensitive cells. The coefficient with the π term is about half that of eq 11 and π for the whole substituents X yields a slightly better correlation than π'.

What appears to account for the difference in the hydrophobic terms in eq 13 compared with eq 12 is that the normal active transport system in sensitive cells is severely impaired in the resistant cells.[28] Hence, the triazines must enter via passive diffusion. Cells which do not have an active transport system turn out to contain a 0.5 π term and those with an active transport system have a coefficient with π near 1.[28,29]

Those cells with a transport protein might be expected to have a more complex QSAR since it would be determined by reaction with DHFR and reaction with transport protein. The similarity of eq 11 for isolated DHFR and eq 12 for sensitive cells suggests that the receptor of the transport protein must look much like the DHFR receptor or else it makes very little structural demand on the antifolates it accepts. In fact, one wonders if a slightly modified DHFR molecule could be employed as the transport protein.

One thought that occurs from eq 12 is that the MR_y term might be the result of some steric effect in the transport protein, but since this same term occurs in eq 13 where the transport system is greatly impaired this seems unlikely.

Comparison of the activity of triazines VIII with cancer DHFR, (L1210 leukemia) L1210 cells sensitive and resistant to methotrexate yields the following three equations.[28]

Inhibition of Purified DHFR from L1210 Leukemia by Triazines

$$\log 1/K_i = 0.98\pi'_3 - 1.14 \log(\beta \times 10\pi'_3 + 1) + 0.79\sigma + 6.12 \quad (14)$$

$$n = 58 \quad r = 0.900 \quad s = 0.264 \quad \pi_0 = 1.76$$

Inhibition of L1210 Cells Sensitive to Methotrexate by Triazines

$$\log 1/C = 1.13\pi'_3 - 1.20 \log(\beta \times 10\pi'_3 + 1) + 0.94\sigma + 0.66I_R - 0.32I_{0R} + 6.72 \quad (15)$$

$$n = 61 \quad r = 0.890 \quad s = 0.241 \quad \pi_0 = 1.45$$

Inhibition of L1210 Cells Resistant to Methotrexate by Triazines

(26) Beece, D.: Eisenstein, L.; Frauenfelder, H.; Good, G.; Marden, M.; Reinish, L.; Reynolds, A. H.; Sorensen, L. B.; Yue, K. T. *Biochemistry* **1980**, *19*, 5147.

(27) Coats, E. A.; Genther, C. S.; Dietrich, S. W.; Guo, Z. R.; Hansch, C. *J. Med. Chem.* **1981**, *24*, 1422.

(28) Selassie, D.; Strong, C. D.; Hansch, C.; Delcamp, T. H.; Freisheim, T. J.; Khwaja, T. A. *Cancer Res.* **1986**, *46*, 744.

(29) Coats, E. A.; Genther, C. S.; Selassie, C. D.; Strong, C. D.; Hansch, C. *J. Med. Chem.* **1985**, *28*, 1910.

$$\log 1/C = 0.42\pi'_3 - 0.15\text{MR} + 4.83 \quad (16)$$

$$n = 62 \quad r = 0.941 \quad s = 0.220 \quad \pi_0 \sim 6$$

Assuming that inhibition of growth by the triazines is due simply to inhibition of DHFR as in the case of *L. casei*, one can compare eq 14 and 15. The bilinear role of the hydrophobic effect is very much the same as is the electronic role for σ in vitro and in vivo. The new indicator parameter I_R is assigned the value of 1 for 3-alkyl groups which were all assigned a π value of 0 in eq 14. While there are great similarities between in vitro and in vivo actives of the triazines, there are also distinct differences. However, eq 16 is radically different from eq 14 and 15. The parameter π_0 could only be estimated as about 6, similar to that for eq 13. The small MR term refers to all of X including Y and hence is not comparable to MR_Y of eq 12 and 13. We suspect that the major reason for the differences between eq 15 and 16 is that active transport is important for the sensitive cells but not for the resistant cells.

Since we do not have the X-ray coordinates of L1210 DHFR we cannot relate the results directly to a graphics model. However, eq 14 can be compared with eq 17 for chicken DHFR where we have a graphics model.[15d] The parameters of eq 14 and 17 are about as close as one can expect for this type of work showing that DHFR from the two sources are very similar indeed.

Inhibition of Chicken Liver DHFR by Triazines VIII

$$\log 1/K_i = 1.01\pi'_3 - 1.16 \log (\beta \times 10\pi'_3 + 1) + 0.86\sigma + 6.33 \quad (17)$$

$$n = 59 \quad r = 0.906 \quad s = 0.267 \quad \pi_0 = 1.89$$

This equation differs somewhat from that for *L. casei*; there is no I term and an electronic term in σ appears. The triazines are much more potent inhibitors of vertebrate enzymes than bacterial enzyme (compare intercepts). The coefficient of 1 with π'_3 suggests complete desolvation of X in the huge hydrophobic cavity shown in view VIII. Substituents in the meta and para positions of the outer phenyl ring have little effect on log $1/K_i$ and hence, as one can see, would not make effective contact with the enzyme. The same coefficient of 1 also appears to occur. The same coefficient of 1 with π_3 also occurs in QSAR for DHFR from human, bovine, rat, L5178Y leukemia, and L1210 leukemia sources.[15h]

Comparing equations and their coefficients obtained via this phenomenological approach is not without difficulties. Collinearity or multiple collinearity among variables, say in one equation but not in another, can produce noncomparable results. Compensation in which one variable cancels another is an insidious problem which has long concerned those working with free energy relationships.[21] The more variables the more serious the problem. Still we are optimistic that our approach can be used to gain new insight into biological structure–activity relationships and QSAR can be used to develop coded 3-D working models of receptor sites of unknown structure for drug research.[30]

The graphics models of the enzyme active sites were built over the past few years with help from many of our colleagues, especially: J. M. Blaney, M. L. Connolly, T. E. Ferrin, C. Huang, and J. McClarin. We are all deeply indebted to Professor Robert Langridge without whose vision, perserverance, and hard work we would not have a powerful high-performance molecular graphics system. We are also indebted to many of our colleagues sited in ref 15 who played important roles in the formulation of the mathematical models. Our research was supported by NSF grant DMB-8518169 (C.H.) and NIH grant R.R. 1081 (R.L.) and by an R. Nelson Smith faculty grant from the Seaver Science Research Fund of Pomona College (C.H.).

(30) Klein, T. E.; Huang, C.; Ferrin, T. E.; Langridge, R.; Hansch, C. In *Artificial Intelligence Applications in Chemistry*; Priece, T. H., Hohne, B. A., Eds.; ACS Symposium Series 306; American Chemical Society: Washington, DC, 1986; p 147.