1981

1981 RHO CHI LECTURE AWARD

Presented in St. Louis, MO., March 29, 1981

Dihydrofolate Reductase Inhibition.

A Study in the Use of X-Ray Crystallography, Molecular Graphics and Quantitative Structure-Activity Relations in Drug Desi gn

By CORWIN HANSCH
Pomona College, Claremont, California 91711

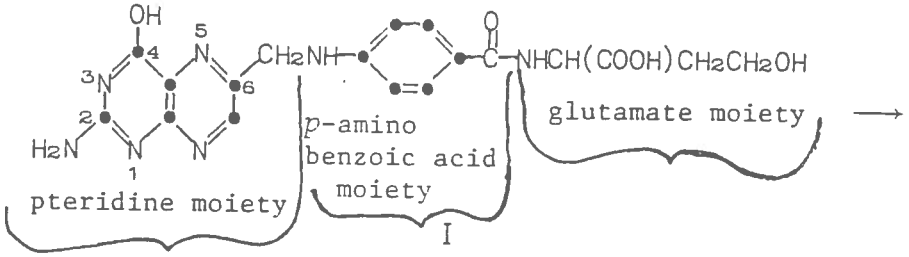
Our group at Pomona College has spent the last 20 years, as have many others, in trying to understand how small organic molecules interact with macromolecules and macromolecular systems. This is a central problem in drug research; in fact, it is a central problem in understanding life itself. Through their interaction with macromolecules, small molecules are converted into macromolecules which then construct themselves into cells which form the whole organisms.

The approach we have taken stems from the thinking of Bronsted, Hammett, Meyer and Overton and others; that is, one can use simple systems and statistical analysis to model chemical and biochemical processes too complex to tackle with a rigorous deterministic frontal assault. It has come to be called quantitative structure-activity relationships (QSAR) for lack of a better term.

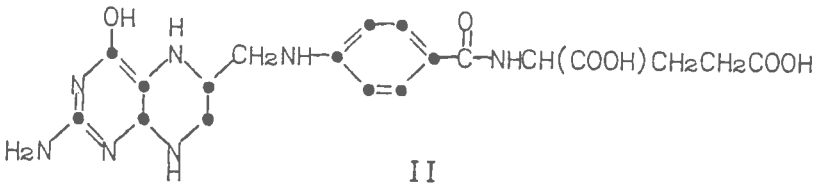
Meyer and Overton, around the turn of the century, observed a parallel between the narcotic action of simple alcohols, esters, ethers, ketones, etc. on tadpoles and fish and their oil/water partition coefficients. Lacking good analytical techniques, they did not attempt to measure many partition coefficients and never attempted to place their discussion in numerical terms, but what they did discover was a means for defining the so-called hydrophobic interaction; that is, the tendency of nonpolar molecules to partition out of water into apolar lipid pools or apolar pockets in macromolecules. Their work was greatly extended by the Finnish botanist Collander who showed that the rate at which organic compounds diffused through plant cells depended on their oil/water partition coefficients. About 1935, Hammett showed that ionization constants of benzoic acids could be used to define a numerical scale for the electronic effect of substituents on organic reactions. Taft then showed, about 1955, that a scale for the steric effect of substituents could be developed by using the rates of hydrolysis of XCH_2-C00Et. With these three numerically-defined scales — σ for electronic effects, E_S for steric effects, and for hydrophobic effects — one could use regression analysis to separate the roles of these different effects on the interaction of organic compounds with macromolecules[1].

There are two major problems in drug design from the QSAR point of view. One is how a drug finds its way to its site of reaction in an animal by a random walk from its site of administration. The other is how the drug interacts with a crucial receptor which then initiates a series of highly complex events which may result in curing an animal of a disease or killing it. Using the simple ideas of Meyer, Overton and Hammett with the powerful tool of regression analysis, it is already possible to obtain crude solutions to why hundreds of different members of a congeneric set of drugs produce different biological responses in a system as complex as a mouse infected with malaria[2] or cancer.[3] However, we cannot be satisfied by crude solutions. At the current stage in the development of QSAR we believe that the most difficult problem to understand is the reaction of a drug with its receptor. In approaching such a problem, one wants to study the pure receptor in isolation to simplify the problem. Because most drug receptors are not yet known, we have elected to study enzymes whose structures have been established by X-ray crystallographic analysis. A most

interesting example in this class is dihydrofolate reductase (DHFR). It has already been shown that the selective inhibition of this drug can be used to control many types of bacterial infection with the drug trimethoprim. Also of great interest, the drug methotrexate (MTX, a DHFR inhibitor) is an important anticancer agent. Recent advances in the techniques of administration of this drug have made it one of the most important in the treatment of a variety of cancers. The mechanism of action of MTX is becoming clearer.[4] DHFR is a crucial enzyme for all forms of life and plays an essential role in the synthesis of bases for the production of DNA. Thus, inhibition of DHFR prevents cell growth. DHFR and its cofactor are necessary to convert folic acid I into tetrahydrofolic acid II. Without tetrafolic acid,

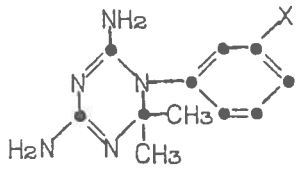


I

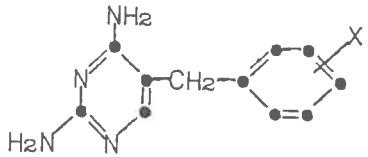


II

the essential bases for DNA cannot be made. The problem for the medicinal chemist is to make an inhibitor of bacterial DHFR which will not too seriously inhibit human DHFR. Fortunately, there are rather large differences between human DHFR and bacterial DHFR so that Barbara Roth and George Hitchings discovered the important drug trimethoprim which selectively inhibits the bacterial enzyme. A most intriguing question which has not yet been answered is, does normal human DHFR differ from DHFR in human tumors and, if so, is the difference significant enough so that drugs could be designed to selectively inhibit tumor DHFR? This problem could be answered by isolation and sequencing the amino acids in enzyme from the two different sources. It is surprising that no one has yet done this. However, even if there are no differences in the enzymes, it has been clearly established that MTX and DHFR inhibitors are active against a variety of human tumors.[5,6] We have therefore undertaken the study of the design of DHFR inhibitors of type III and IV, in part to attempt to discover better chemotherapeutic agents, but more important,



III



IV

to understand the interaction of small molecules and macromolecules. DHFR was selected for study because the structure for enzyme from *E. coli* and *L. casei* cells

has been established and the structure of enzyme from other sources is under very active investigation.[7] Another reason for selecting DHFR is that it is clear from the extensive studies on congeners III by B. R. Baker and his students that QSAR can be readily formulated from inhibition studies.[8]

While it is most attractive to study highly purified DHFR with a set of inhibitors rather than studying the same inhibitors in whole animals where one is plagued by distribution and metabolism problems, we do not yet have enough experience to know how valuable QSAR obtained with isolated enzyme will be in helping one estimate activity of the inhibitors in whole animals. Our approach is to study III and IV first on purified enzyme, then in cell culture, and finally in mice. We hope in this way to sort out some of the major problems in the design of drugs starting with enzymes rather than mice. One of our first concerns has been one that has long vexed biochemists: does the isolated enzyme in buffer solution have the same conformation and react the same way as enzyme in the living cell? Equations 1-3 for inhibitors III yield some QSAR insight on this problem.

Inhibition of DHFR from *L. casei*

$$\log 1/K_{i(app)} = 0.46\pi' - 0.57 \log(\beta.10^{\pi'} + 1) + 1.38I + 3.16 \quad (1)$$

$n = 44$; $r = 0.947$; $s = 0.333$; $\pi_0 = 4.4$

Inhibition of Methotrexate-Resistant *L.casei* Cell Culture

$$\log 1/C = 0.43\pi + 1.06I - 0.49MR_Y + 3.38 \quad (2)$$

$n = 37$; $r = 0.964$; $s = 0.268$

Inhibition of Methotrexate-Sensitive *L.casei* Cell Culture

$$\log 1/C = 0.80\pi' - 1.06 \log(\beta.10^{\pi} + 1) + 0.80I - 0.94MR_Y + 4.37 \quad (3)$$

$n = 34$; $r = 0.929$; $s = 0.371$; $\pi_0 = 2.9$

In eq 1, K_i is the inhibition constant which in effect is the concentration of III causing 50% inhibition of the DHFR, n represents the number of data points (number of congeners tested), r is the correlation coefficient, and s is the standard deviation from the regression equation. The parameter π is the hydrophobic constant for X of III obtained from octanol/water partition coefficients.[8] The $'$ indicates that for about a dozen cases where X = 3-$CH_2ZC_6H_4$-Y (Z = 0 or NH), π' for Y is assigned the value of 0 (i.e., π' $CH_2ZC_6H_4$-Y = π' $CH_2ZC_6H_5$). The reason for this assignment is that Y appears to have no significant effect on K_i regardless of whether X is hydrophobic, hydrophilic, large (i.e., *t*-butyl), or small. We assumed that this meant that Y was held in such a way that it could not contact the enzyme. Some typical examples of 3-X where π takes its normal value are : SO_2NH_2, $CONH_2$, OH, $COCH_3$, NO_2, CF_3, Cl, I, $OCH_2C_6H_3$-3′,4′-Cl_2, $OCH_2C_6H_5$, $O(CH_2)_nCH_3$ (n = 3, 8, 10, 11, 12, 13). Some examples of Y are: H, 3-Cl, 3-CN, 3-CH_2OH, 3-CH_3, 4-SO_2NH_2, 3-$NHCONH_2$, 3-$NHCSNH_2$, 3-$C(CH_3)$, $CH(CH_3)_2$.

The two terms in π' constitute the bilinear model of Kubinyi for describing the relationship between the inhibitory power of X and its hydrophobicity (π). If I is held constant in eq 1, then inhibitory power increases linearly with slope of 0.46 until one nears the value of π_0 (determined by the parameter β; eq 1 is a nonlinear equation so that β is evaluated by an iterative procedure via the computer); at this point the slope gradually changes to -0.11 (0.46 - 0.57) with activity continuing to decrease linearly - hence the term bilinear. In the case of eq 1, the right hand slope of the bilinear equation is so near 0 (-0.11) that we cannot be sure it differs from 0; therefore, the assumption is made that large hydrophobic groups extend beyond the active site into an aqueous phase so that the part of the substituent $> \pi_0$ does not affect K_1. The parameter π_0 represents the limit of π for which one can get increasing inhibitory power (binding ability) by simply making X more hydrophobic.

C in eq 2 and 3 represents the concentration of III needed to produce 50% inhibition in the growth rate of *L. casei* cell culture. The experimental work for these equations comes from the laboratory of E. A. Coats and C. S. Genther of the University of Cincinnati. The enzyme used to formulate eq 1 is extracted from the same kind of

cells used to develop eq 2. Despite the similarity of the DHFR, eq 2 is grossly different from eq 1. We have not yet made π hydrophobic enough to establish π_0 with certainty. An inspection of the data in hand suggests that π_0 is about 6; much greater than for eq 1. In eq 2, π is used instead of π' so that the hydrophobic effect of Y is making itself evident. With cells, the inhibitor must cross hydrophobic membranes and therefore Y displays a hydrophobic interaction.

The coefficient with I is smaller in eq 3 than in either eq 1 or 2 which suggests that the binding region of $-CH_2OC_6H_4-$ is somewhat different *in vivo* than *in vitro*.

A pronounced difference in eq 2, as well as in eq 3, is the terms in MR_Y. The molar refractivity (MR) of a substituent (Y) is closely related to its molar volume.[9] From this it follows that the negative coefficient with MR_Y implies a steric effect of Y in the living cell not seen in this isolated enzyme. Since π_Y and MR_Y are quite orthogonal vectors, this effect has nothing to do with a classical hydrophobic effect. It would appear that the window open to Y in isolated DHFR is partially blocked in DHFR in live bacteria cells by some kind of movable barrier. Since all activity is not lost when Y is a rather large group, flexibility of the barrier allows Y enough space so that the inhibitor can fit onto the DHFR molecule. To our knowlege, this is the first instance where a specific inference has been made about the difference in structure between an isolated enzyme and the enzyme in the living cell. The cause of this difference may be due to the DHFR being adsorbed onto a membrane or another macromolecule in the cell. We believe that QSAR offers a means for comparison of enzymes in the two different states; possibly the only general means when the enzyme is in the cell or whole animal.

Equation 3 is more similar to eq 1 than to eq 2. Of course here we cannot be sure the basic structure of the DHFR is the same as that upon which eq 1 is based. The normal *L. casei* organism has been used for eq 3, while the isolated enzyme comes from the resistant organism used to obtain eq 2. We were surprised to find that π' gave a slightly better correlation in eq 3 than π even though membrane interactions are involved. This may be artifactual, or it may mean the membrane system in the sensitive cells is rather simple compared to the resistant cells. The coefficient with I is considerably smaller than in eq 1, suggesting a definite change in the region where $-CH_2OC_6H_4-$ binds. we again find an MR_Y term with a negative coefficient, implying even more restriction of binding by Y. In summary then, correlation eq 1-3 give us insight about the state of DHFR *in vivo* and *in vitro* and show us a surprising difference between resistant and sensitive cells in their response to hydrophobic drugs.

Equations 4 — 7 allow us to compare the action of inhibitors III on mammaliam systems with the bacterial systems (eq 1-3). The work on tumor cell culture has been carried out in T. Khwaja's laboratory at the USC Cancer Center with the help of C. D. Selassie of our group. The murine tumor enzyme used to establish eq 5 was supplied to us by J. Bertino of Yale University.

Inhibition of Purified Bovine Liver DHFR

$$\log 1/k_{i(app)} = 1.08\pi - 1.19\log(\beta \bullet 10^{\pi} + 1) + 7.27 \quad (4)$$

$n = 38$; $r = 0.903$; $s = 0.288$; $\pi_0 = 1.62$

Inhibition of Murine Leukemia (L5178Y) DHFR

$$\log 1/K_{i(app)} = 1.13\pi' - 1.33\log(\beta \bullet 10^{\pi'} + 1) + 0.42I + 6.44 \quad (5)$$

$n = 38$; $r = 0.920$; $s = 0.315$; $\pi_0 = 1.44$

50% Inhibition of Leukemia (L5178Y) Cell Culture

$$\log 1/C = 1.32\pi - 1.70\log(\beta \bullet 10^{\pi} + 1) + 0.44I + 8.10 \quad (6)$$

$n = 37$; $r = 0.929$; $s = 0.274$; $\pi_0 = 0.76$

50% Inhibition of L5178Y Cells Resistant to Methotrexate

$$\log 1/C = 0.57\pi - 0.15MR - 0.35\log(\beta \bullet 10^{\pi} + 1) + 5.12 \quad (7)$$

$n = 42$; $r = 0.932$; $s = 0.288$

Equation 4 is quite different from eq 1 in several important respects: the coefficient with π' is much larger, π_0 is much smaller, and the term in I is absent. One might expect that the hydrophobic pocket in *L. casei* DHFR is larger than bovine because of comparative sizes of π_0; however, the larger coefficient with π' in eq 4 means that the increase in activity attainable by increasing π from 0 to $\pi = \pi_0$ ($1.08 \times 1.62 = 1.75$) is not generally different from that for eq 1 ($0.46 \times 4.4 = 2.02$). This suggests similar size hydrophobic regions, but a differrent mode in the desolvation process of X or a greater rigidity in one enzyme compared to the other.

Equation 4 differs from the corresponding bacterial equation in that it lacks the term in I. This is important because I makes a large contribution to the $1/K_i$ in eq 1.

Inhibition of the isolated murine enzyme (eq 5) is similar to bovine enzyme as far as hydrophobicity goes. The point of interest in eq 5 is the term in I which is much smaller than the corresponding term in eq 1 and which at first we thought might be an artifact. In testing the inhibitors on the leukemia cell culture for cells sensitive to MTX (eq 6), we find the same term in I which is rather convincing evidence that this small effect is real for the murine system.

Equation 6 resembles eq 5 for the leukemia system just as eq 3 resembles eq 1 for the *L. casei* system. In each system we find π_0 to decrease in the cell culture. The cells resistant to MTX likewise give a much different equation. There is a bilinear term in eq 7 but its slope is positive ($0.57 - 0.35 = 0.22$), not negative; hence π_0 cannot be defined without testing more hydrophobic cogeners. It appears from an inspection of the data that π_0 is near 6; a marked contrast compared to 0.8 of eq 6. Thus both cell types, bacterial and mammalian, appear to use the same mechanism to protect themselves from MTX. As MTX is an ionic, very hydrophilic drug, cells may use this mechanism to guard against hydrophilic toxicants.

Crystallography and Graphics

In the formulation of QSAR, one allows inhibitors to interact with enzyme, cell or animal and one attempts to deduce a structure-activity relationship from computer-assisted analysis of the perturbations obtained. Since QSAR is still in the developmental state, one would like to have an independent means for checking up on QSAR results. We have embarked on a joint project with D. A. Matthews and J. Kraut at UCSD to study our inhibitors bound to DHFR. The UCSD group has undertaken an extensive study of the crystallography of DHFR.[10] While getting the coordinates of each atom in the protein is of enormous help, the study of the resulting data is not easy. Constructing models of the protein is expensive and time-consuming. Moreover, the models become eroded with handling so that one can no longer be sure of the details of the spatial relationships. Consequently, crystallographers have turned to the use of computer graphics for visualization of protein structure. A huge advance has recently been made by Langridge's group at the University of California Medical Center in San Francisco.[11] The stereo views I-III have been made by J. M. Blaney using Langridge's molecular graphics system. Stereo view I is conventional graphics in color. The blue lines are the backbone protein structure for DHFR from *L. casei*. The X-ray crystallographic coordinates provide the information for the protein structure with MTX bound to it. The glutamate end of MTX projects out toward the viewer.

Two other drugs are also included but the coordinates for them are not known. They were positioned by placing their 2, 4-diaminopyrimidine portions over the coordinates for the corresponding atoms of MTX. The yellowish structure is the antibacterial trimethoprim and the lavender structure is a triazine III where X = $CH_2OC_6H_4$-3′ -Cl. The view shows that there may be enough space in the region near the active site to accommodate this and, more important, have the 3 -Cl project beyond the enzyme as called for by eq 1.

A fascinating technique has been used in stereo view II to place a surface on the protein which shows the space available for a ligand to fit into. This technique was worked out by M. L. Connolly[11] and consists of directing the computer to "roll" an idealized water molecule over the protein. A color-coded dot is put down at each point where the water molecule comes within a van der Waals radius of a protein atom. The yellow lines are fragments of protein backbone surrounding the active site where folic acids, its analogs and inhibitors bind.

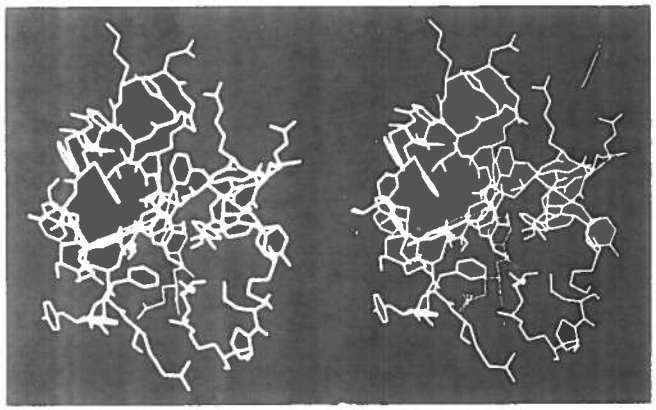
In stereo view II the channel toward the bottom of the picture is where the glutamate end of MTX falls. There is a narrowing in the center of the channel where the para amino benzoic acid moiety binds. Above this our triazine shown in lavender in view I now has placed on it a van der Waals surface of red dots. It is clear from this that our compound would fit into the region shown in view I, but the surfaces are of enormous help in seeing just how tight the fit is. It is also clear that the Cl does indeed project beyond the enzyme surface. As yet, this is speculation and we won't know the precise answer until Kraut and Matthews finish the crystallography of this congener or a similar one bound to DHFR.

Stereo view III shows how molecular graphics can be used to compare the active site of two different forms of DHFR. The blue lines in this picture represent the protein backbone of DHFR from *L. casei* and the orange lines are for *E. coli* DHFR. The superimposition was achieved by placing the structures from X-ray coordinates on separate parts of the computer screen simultaneously. They were then moved together so that the pteridine rings of MTX bound to each enzyme were exactly superimposed. Note the two fused hexagons of pteridine rings slightly left of center. One of the enzymes contained in addition to MTX is the cofactor NADPH whose skeleton is depicted in lavender. The pyridine moiety of the NADPH is nicely positioned to transfer a H to the 5,6-position of 7,8-dihydrofolic acid. The glutamate ends of the two MTXs in view III project toward the viewer (light yellow and blue). The benzene rings of the NHC_6H_4CONH moieties are only slightly out of parallel, showing the great similarity of the positioning of MTX in the two different active sites. In this sense the two enzymes are quite similar; however, there are large differences which affect the structure-activity relationships of inhibitors interacting with the two different systems. Four different residues which *L. casei* DHFR has and *E. coli* lacks are shown in light green. At the top of the picture and slightly left of center Leu-19 projects into the active site of *L. casei*. To the left and slightly above is the green his-22 which *E. coli* DHFR lacks. Slightly to the right of center one can see Phe-49 and back toward the viewer from Phe-49 is Pro-50.

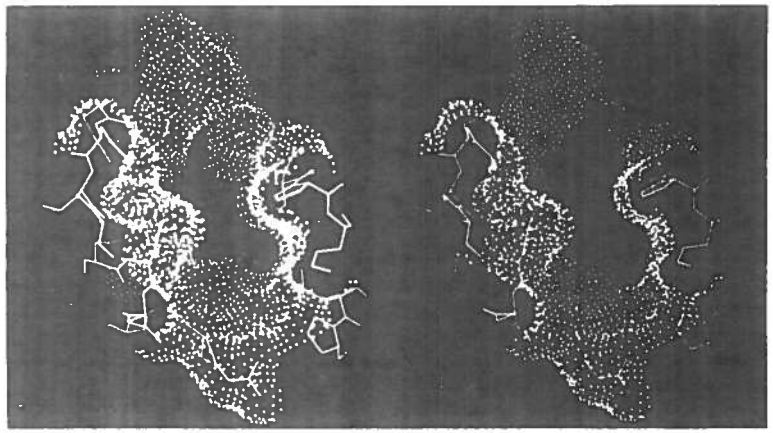
These pictures display the great promise of Langridge's graphics system for the interpretation of data obtained via X-ray crystallography. Of greatest importance is the surface representation illustrated in view II. Without such facility, it is very difficult and tedious to assess the possibility of ligand fit to protein.

Although we do not as of the present have crystallographic data on enough of our inhibitors and several enzymes to draw valuable general conclusions, the gross features of the active site are clear enough so that one can easily see that it would be silly to make certain kinds of derivatives. This alone can save an enormous amount of synthetic and testing effort. Some more subtle features appear plausible but will need more study for confirmation. For example, π' is assigned a value of π for $CH_2OC_6H_5$ for groups $CH_2OC_6H_4$-Y even though Y may be very hydrophobic or hydrophilic. This value of π is 1.66 which is far below π_0 of 4.4. For example, when Y = Cl, then π for the whole substituent would only be 1.66 + 0.71 = 2.37. One would think this should fit into the pocket. An explanation is that phenyl of 3-$CH_2OC_6H_4$- fits neatly into the narrow slot seen in view II where the strong hydrophobic interaction holds it, forcing Y to project beyond the surface. Flexible groups such as $O(CH_2)_nCH_3$ can coil into other parts of the pocket and make hydrophobic contact of up to n = 9. These alkyl residues do not appear to lock into the channel for the benzene ring. We shall soon have the crystallographic results to see what in fact is happening.

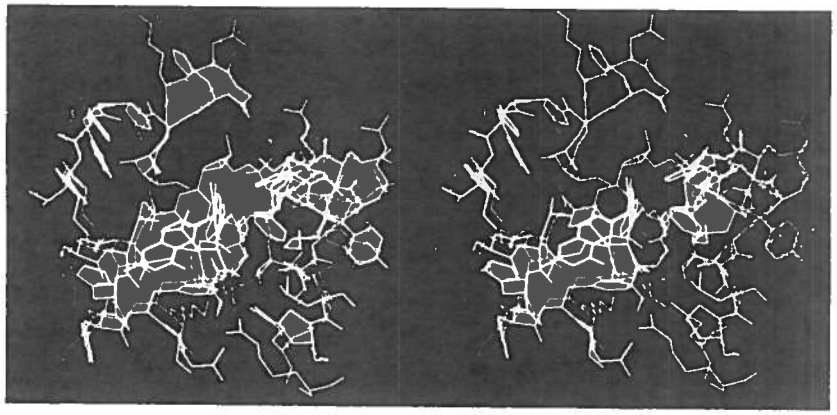
While crystallography and graphics are going to play a very important part in drug design, they cannot substitute for the information one gains from probing the enzyme of DNA molecule with a well developed set of inhibitors. For example, crystallography may not be refined enough to account for the *I* terms of eq 1-3. Of course there is no way to predict the MR terms of eq 2 and 3. Capitalizing on such small differences obtained via QSAR will be helpful in designing more selective drugs. What is needed now and what we are trying to do is to make a study of a well varied set of inhibitors on various forms of isolated enzyme, on cell culture, and in whole animals to obtain basic knowledge about the pitfalls of extrapolating from enzyme studies to the final drug.



STEREO VIEW I



STEREO VIEW II



STEREO VIEW III

EDITOR'S NOTE: Colors described in text for stereo views were not reproducible for this publication.

ACKNOWLEDGMENT. The experimental work for eq 1-7 was done by E. A. Coats, S. W. Dietrich, C. S. Genther, Zong-ru Guo and Cynthia D. Selassie. The details will be published elsewhere. The molecular graphics analysis was done by J. M. Blaney in the Computer Graphics Laboratory (supported by NIH RR-1081) of Robert Langridge at the University of California at San Francisco. The photography is by Paul K. Weiner.

REFERENCES

1. C. Hansch, *Acc. Chem. Res* 2, 232 (1969)
2. K. H. Kim, C. Hansch, J. Y. Fukunaga, E. E. Steller, P. Y. C. Jow, P. N. Craig, and J. Page, *J. Med. Chem.* 22, 366 (1979)
3. C. Hansch, A. Leo, C. Schmidt, P. Y. C. Jow, and J. A. Montgomery, *J. Med. Chem.* 23, 1095 (1980).
4. Y. M. Wang and T. L. Loo, *Cancer Bull.* 33, 49 (1981)
5. M. P. Sullivan, *Cancer Bull.* 33, 54 (1981).
6. N. Jaffe, *Cancer Bull.* 33, 59 (1981).
7. D. J. Baker, C. R. Beddell, J. N. Champness, P. J. Goodford, F. E. A. Norvington, D. R. Smith, and D. K. Stammers, *FEBS Lett.* 126, 49 (1981).
8. C. Silipo and C. Hansch, *J. Am. Chem. Soc.* 97, 6849 (1975).
9. C. Hansch and A. Leo, "Substituent Constants for Correlation Analysis in Chemistry and Biology," Wiley-Interscience, New York, 1979,
10. D. A. Matthews, A. A. Alden, J. T. Bolin, D. J. Filman, S. T. Freer, R. Hamlin, W. G. J. Hol, R. L. Kisliuk, E. J. Pastore, L. T. Plante, N. H. Xuong, and J. Kraut, *J. Biol. Chem.* 253, 6946 (1978).
11. R. Langridge, T. E. Ferrin, I. D. Kuntz, and M. L. Connolly, *Science* 212, 661 (1981).