# Use of Physicochemical Parameters in Distance Geometry and Related Three-Dimensional Quantitative Structure–Activity Relationships: A Demonstration Using *Escherichia coli* Dihydrofolate Reductase Inhibitors

Arup K. Ghose and Gordon M. Crippen*

*Department of Chemistry, Texas A&M University, College Station, Texas 77843.* *Received May 4, 1984*

In earlier distance geometry related three-dimensional quantitative structure–activity relationships (Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1984**, *27*, 901) the interactions of the ligand atom or group with the receptor site were evaluated empirically by using mathematical optimization techniques, without considering their physicochemical properties. In the present work we show how to use various physicochemical parameters in our three-dimensional receptor mapping. We have developed a model for *E. coli* DHFR using the inhibition data of 25 pyrimidines and 14 triazines. It gave a correlation coefficient of 0.893 and standard deviation of 0.530. It successfully predicted the binding data of five pyrimidines and five triazines.

We have reported[1-4] the success of distance geometry in developing three-dimensional receptor models for various dihydrofolate reductase (DHFR) inhibitors. The results are in general agreement with the X-ray crystallographic data[5-7] of some DHFR–inhibitor complexes, and our approach showed that for rat liver DHFR, a single three-dimensional receptor model can not only fit the binding data of a large number of inhibitors but it can also successfully predict the binding data of a large variety of compounds.[4] One of them was a member of a new class not included in the original data set. One shortcoming of our approach was that like Free–Wilson analysis, we used atom types to express their interaction with the receptor site. That approach may suffer from the following drawbacks: (i) it is difficult to get any idea of interaction of a foreign atom not studied in the original data set, and (ii) physicochemically similar atoms may be assigned very different interactions or vice versa, thereby posing problems in the physical interpretation of the model. In our earlier work[4] it was mentioned that these interactions might as well be correlated with physicochemical parameters. The objective of the present work is to show a method of using the physicochemical parameters in distance geometry or comparable three-dimensional quantitative structure–activity relationships (QSAR). We also want to make a critical comparison of our method with other QSAR approaches to indicate its greater versatility.

In the present work a three-dimensional receptor model for *Escherichia coli* DHFR was developed from the inhibition data of 25 pyrimidines[8] and 14 triazines.[9] The model successfully predicted the binding data of five pyrmidines and five triazines. The number of compounds was kept to a minimum, in contrast to all our previous studies, to demonstrate the success of the method in a limited data set. The compounds having substituents with more than four important rotatable torsion angles were avoided to keep the computational time to a manageable level. By the term "important" we mean those torsion angles whose rotation alter the position of non-hydrogen atoms. Ten compounds were arbitrarily omitted from these acceptable ones to test the predictive power of the model.

## Methods

The method followed in the present work is similar to that described in our earlier work[4] except for using physicochemical parameters of the ligand atoms to correlate their interaction with the site pockets. A concise description of the method follows: (i) Construct each ligand molecule satisfying crystallographic bond lengths and bond angles. (ii) Construct some site pockets relative to the ligand molecules on the basis of an active conformation hypothesis. Some or all of the atoms of the active conformation may be assumed to occupy the site pockets to give its interaction. In other words, these atoms are representative of the site pockets (see ref 4 for a detailed explanation). (iii) Evaluate the geometrically allowed binding modes. By the term binding mode we mean which atom of the ligand goes to which site pocket. Due to the internal rotations around the various dihedral angles as well as the rigid rotations and translations of the ligand relative to the receptor, a large number of geometrically feasible binding modes are possible, unless such motions are restricted by the steric requirements of the receptor site.[10] (iv) When an atom (or atoms) of the ligand enters a site pocket, it interacts with the site, and the interaction energy is a function of one or more physicochemical parameters of the atom. Assuming a linear function for the interaction, the binding energy of a particular binding mode may be given by

$$E_{\text{calcd}} = -CE_c + \sum_{i=1}^{n_s} \sum_{j=1}^{n_p} [C_{i'j} \sum_{k=1}^{n_o} P_j(t_k)] \qquad (1)$$

where $E_c$ is the energy of the conformation under consideration, $C$'s are the coefficients to be determined by quadratic programming, $i'$ is the type of the site $i$, $n_s$ represents the number of site pockets, $n_p$ represents the number of parameters we want to correlate with that site

(1) Crippen, G. M. *J. Med. Chem.* **1979**, 22, 988.
(2) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1982**, *25*, 892.
(3) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1983**, *26*, 996.
(4) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1984**, *27*, 901.
(5) Volz, K. W.; Matthews, D. A.; Alden, R. A.; Freer, S. T.; Hansch, C.; Kaufman, B. T.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 2528.
(6) Bolin, J. T.; Filman, D. J.; Mathews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650.
(7) Filman, D. J.; Bolin, J. T.; Mathews, D. A.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13 663.
(8) Li, R. L.; Dietrich, S. W.; Hansch, C. *J. Med. Chem.* **1981**, *24*, 538.
(9) Baker, B. R. *J. Med. Chem.* **1967**, 10, 912.
(10) A three-dimensional coordinate manipulation was used to evaluate the binding modes as reported earlier.[4] An algorithm for more exhaustive search of binding modes for flexible molecules using distance geometry[11] which satisfies all the requirements of three-dimensional space has been worked out and will be published elsewhere.
(11) Kuhl, F. S.; Crippen, G. M.; Friesen, D. K. *J. Comput. Chem.* **1984**, *5*, 24.

0022-2623/85/1828-0333$01.50/0 

pocket interaction, $n_o$ represents the number of atoms occupying that site pocket, and $P_j$ represents the $j$th physicochemical parameter of the atom of type $t_k$. Although most physicochemical parameters in current use are assigned for some common groups, in the present work the atomic contribution to any physicochemical parameter has been used, a detailed discussion of which is given at the end of this section.

A site pocket may be attractive or repulsive depending on the nature of the receptor atoms surrounding the pocket and the ligand atoms entering the pocket. Attraction, for example, may arise from the entrance of a hydrophobic group in a pocket surrounded by hydrophobic atoms. Repulsion may result from the entrance of a hydrophilic group in a hydrophobic pocket or vice versa. This type of repulsion is "soft"; a ligand molecule may tolerate that type of repulsion in the binding process provided it has strong attraction to other regions. In our model building process it may be necessary to create repulsive pockets of another type. The repulsion here is "hard", intolerable. A ligand cannot bind if any of its atoms enters that type of pocket. Physically, those pockets represent the regions occupied by the receptor atoms. In that sense the term "pocket" is somewhat misleading here. More appropriate terms are excluded volumes or receptor volumes.

The optimization procedure in our approach is quite different from the other QSAR methods. In other QSAR methods[12,13] there is only one expression for the calculated inhibition data. In such methods the optimization procedure is the simple least-squares fit of the observed and calculated inhibition data. These are "static QSAR", since the single expression for calculated binding energy is based on a fixed position, orientation, and conformation of the ligand molecule in the active site. Although that is compatible with the lock and key interpretation of the active receptor site and the ligand, recent spectroscopic evidence suggests that some structural changes in the ligand molecule may lead it to bind in a different orientation.[14] The present approach is "dynamic QSAR", since it explores the various binding possibilities arising from the rigid rotations, translations, and internal rotations. The optimization here may be done in two fundamentally equivalent ways. Among the various geometrically feasible binding modes, the actual mode should be the one having the maximum binding energy. In the first approach it is assumed that the actual (fixed) binding modes are known. In such case the various proportionality constants of the expression for calculated binding energy should be adjusted in such a way that

$$\sum_{i=1}^{n}(E_{\text{fixed}} - E_{\text{obsd}})^2 = \text{minimum} \qquad (2a)$$

under the constraints

$$E_{\text{fixed}} > E_{\text{rest}} \qquad (2b)$$

where $E_{\text{fixed}}$ represents the calculated binding energy of the fixed mode and $E_{\text{rest}}$ that of the other geometrically feasible binding modes, and $E_{\text{obsd}}$ is the experimental binding energy. The summation is over the total number of molecules, $n$. This constrained solution is exactly equivalent to the standard quadratic programming problem.[15,16] The difficulty of this approach, however, is the choice of the fixed mode, since initially the various proportionality constants in the expression for calculated binding energy are not known. The intuitive choices often are not mathematically consistent in the quadratic programming algorithm. This situation leads us to develop the second optimization procedure where the fixed binding mode, if necessary, is altered to another geometrically feasible binding mode. That alteration may be made in an automated way by using the optimization technique reported earlier.[17] Here the altered binding mode corresponds to the one having the least root mean squared (rms) deviation in the ongoing optimization.

The constraints always have some damaging effect on the rms deviation. However, it is an essential requirement for any receptor mapping procedure using a flexible molecule. Consideration of flexibility on the other hand often shows some obvious faults in our initial hypothesis on which the model is based. Those faults otherwise go unnoticed in "static QSAR" (see Results and Discussion section).

The assignment of the atomic contribution to any physicochemical parameter is extremely difficult, since many subtle structural changes influence the physicochemical properties markedly. These factors can be taken into account by using a large number of atom types to differentiate the various structural environments. There are several structural features that need to be considered: charge density, hybridization, extent of substitution, etc. With our limited resources only a few of the numerous possibilities have been explored and the one having the best statistical significance has been reported here. We worked mainly on hydrophobic atomic values and assumed that simple addition of the atomic contributions will reproduce the partition coefficients of the molecule in a water–octanol system. The same atom classification was used for molar refractivity, and the atomic contributions were evaluated from the values reported by Martin,[12] with correction for multiple bonds by adding half of the multiple bond contribution to the atoms sharing it. However, in the long run it may be necessary to use a different classification for different physicochemical properties.

The atomic values of the various physicochemical properties are necessary for two reasons: (i) in distance geometry or related three-dimensional QSAR we are concerned about the positions of atoms in space relative to the active receptor site, and (ii) a large substituent in a molecule may not experience the same type of interaction at the receptor site, and some parts may even escape any direct interaction. In the present work only a linear form of the interaction was used. As long as the interactions are linear functions of the coefficients to be determined, any form of representation can be used in the framework of the quadratic programming algorithm.

## Results and Discussion

**Development of Physicochemical Parameters.** First the atomic contribution to hydrophobic parameters has been assigned. Hansch et al.[18] retained constant fragment

(12) Martin, Y. C. "Quantitative Drug Design"; Marcel Dekker: New York, 1978.
(13) Topliss, J. G. "Quantitative Structure–Activity Relationships of Drugs"; Academic Press: New York, 1983.
(14) Matthews, D. A.; Alden, R. A.; Bolin, J. T.; Filman, D. J.; Freer, S. T.; Hamlin, R.; Hol, W. G. J.; Kisliuk, R. L; Pastore, E. J.; Plante, L. T.; Xuong, N.; Kraut, J. *J. Biol. Chem.* **1978**, *253*, 6946.
(15) Ravindran, A. *Commun. ACM* **1972**, *15*, 818.
(16) Ravindran, A. *Commun. ACM* **1974**, *17*, 157.
(17) Ghose, A. K.; Crippen, G. M. In "Proceedings of the 4th European Symposium on Chemical Structure and Biological Activity: Quantitative Approaches"; Dearden, J. C., Ed.; Elsevier: Amsterdam, The Netherlands, 1982; p 99.
(18) Hansch, C.; Leo, A. "Substituent Constants for Correlation Analysis in Chemistry and Biology"; Wiley: New York, 1979.

**Table I.** Classification of Atoms and Their Contributions toward Hydrophobicity and Molar Refractivity

| type | description | hydrophobic contribution | *t* test[a] | molar refractivity |
|---|---|---|---|---|
| 1 | C, in $CH_3$ | 0.2101 | 100.00 | 2.42 |
| 2 | C, in $-CH_2-$ | 0.1182 | 99.78 | 2.42 |
| 3 | *tert*-C, -CH< | -0.4419 | 100.00 | 2.42 |
| 4 | neo-C, >C< | -0.7538 | 100.00 | 2.42 |
| 5 | ethylenic C, $=CH_2$ | 0.7274 | 100.00 | 3.285 |
| 6 | ethylenic C, =CH- | 0.5011 | 100.00 | 3.285 |
| 7 | ethylenic C, =C< | -0.5235 | 99.99 | 3.285 |
| 8 | acetylenic C, ≡CH | 0.7647 | 100.00 | 3.620 |
| 9 | acetylenic C, ≡C- | 0.1661 | 80.35 | 3.620 |
| 10 | aromatic C, CH | 0.3992 | 100.00 | 3.285 |
| 11 | aromatic C, C- | 0.1263 | 99.15 | 3.285 |
| 12 | carbonyl C, >CO | -1.6828 | 100.00 | 3.285 |
| 13 | carboxyl and derivatives, -COX | -1.8158 | 100.00 | 3.285 |
| 14[b] | H, to[c] $C_{sp^3}{}^0$ | 0.1325 | 100.00 | 1.03 |
| 15 | H, to $C_{sp^3}{}^1$, $C_{sp^2}{}^0$ | -0.0929 | 100.00 | 1.03 |
| 16 | H, to $C_{sp^3}{}^2$, $C_{sp^2}{}^1$, $C_{sp}{}^0$ | -0.5860 | 100.00 | 1.03 |
| 17 | H, to $C_{sp^3}{}^3$, $C_{sp^2}{}^2$, $C_{sp^2}{}^3$ | -0.4524 | 99.05 | 1.03 |
| 18 | H, to electronegative atoms and active α-hydrogens | -0.1041 | 100.00 | 1.03 |
| 19 | O, alcoholic | -0.5664 | 100.00 | 1.52 |
| 20 | O, phenolic, enolic carboxylic | 0.2300 | 96.50 | 1.52 |
| 21 | O, double bonded | 1.6024 | 100.00 | 3.075 |
| 22 | O, aliphatic ether | -0.1129 | 62.22 | 1.64 |
| 23 | O, aromatic ether | 0.3933 | 97.69 | 2.18 |
| 24 | O, unused | | | |
| 25 | N, aliphatic amine | -0.2578 | 95.25 | 2.47 |
| 26 | N, aromatic amine amide | -0.3384 | 100.00 | 3.69 |
| 27 | N, unused | | | |
| 28 | N, imino, nitrile | 0.7620 | 100.00 | 3.00 |
| 29 | N, unused | | | |
| 30 | N, in pyridine or pyrimidine | -1.0365 | 100.00 | 2.50 |
| 31 | N, aromatic nitro | -3.1288 | 100.00 | 1.30 |
| 32 | N, aliphatic nitro | -3.2605 | 100.00 | 1.30 |
| 33 | N, nitroso aromatic or amino | -0.9942 | 100.00 | 2.125 |
| 34 | N, unused | | | |
| 35 | F, to $C_{sp^3}{}^1$ | 0.4234 | 98.51 | 1.05 |
| 36 | F, to $C_{sp^3}{}^2$ | 0.6489 | 100.00 | 1.05 |
| 37 | F, to $C_{sp^3}{}^3$ | 0.5912 | 100.00 | 1.05 |
| 38 | F, to aromatic or ethylenic C | 0.6111 | 99.98 | 0.92 |
| 39 | Cl, to $C_{sp^3}{}^1$ | 1.0171 | 100.00 | 5.93 |
| 40 | Cl, to $C_{sp^3}{}^2$ | 1.1083 | 100.00 | 5.93 |
| 41 | Cl to $C_{sp^3}{}^3$ | 0.9183 | 100.00 | 5.93 |
| 42 | Cl, to aromatic or ethylenic C | 0.9272 | 100.00 | 6.03 |
| 43 | Br, to $C_{sp^3}{}^1$ | 0.7254 | 99.95 | 8.80 |
| 44 | Br, to $C_{sp^3}{}^2$ | 1.1999 | 100.00 | 8.80 |
| 45 | Br, to $C_{sp^3}{}^3$ | unused | | |
| 46 | Br, to aromatic or ethylenic C | 1.2072 | 100.00 | 8.88 |
| 47 | I, to $C_{sp^3}{}^1$ | 1.5785 | 100.00 | 13.90 |
| 48 | I, to $C_{sp^3}{}^2$ | unused | | |
| 49 | I, to $C_{sp^3}{}^3$ | unused | | |
| 50 | I, to aromatic or ethylenic C | 1.4603 | 100.00 | 13.90 |
| 51 | S, thiol | 0.6689 | 99.31 | 7.74 |
| 52 | S, thioether | 1.0100 | 99.99 | 9.45 |
| 53 | S, -SO- | -2.6082 | 100.00 | 5.50 |
| 54 | S, $-SO_2-$ | -3.9719 | 100.00 | 6.70 |
| 55 | S, thioketone | 2.0300 | 100.00 | |
| 56[d] | correction for ion pair | -5.0832 | 100.00 | |

[a] Level of significance of each contribution. [b] The superscript indicates the formal oxidation number, i.e., the number of more electronegative atoms attached to the carbon. [c] Description of the atom to which it is attached. [d] We have yet to decide how this correction factor is to be distributed among various atoms of the ion pair.

values for fundamental structural elements and then introduced some corrections for some structural moieties. In order to include the correction factors of the Hansch approach, a large number of atom types has been used.

Branching, hybridization, and charge densities were used to classify the atoms. One such classification which gave good *F*-test and *t*-test values for significance is shown in Table I. Initially the carbons were classified in terms of their hybridization and formal oxidation number (the number of more electronegative atoms attached), and the hydrogens were classified in terms of their point of attachment. For example, those attached to saturated carbon, ethylenic carbon, aromatic carbon, acetylenic carbon, and electronegative atoms were given different classes. Other atoms were classified similarly, as given in Table I. Although that classification gave an almost identical correlation coefficient and *F* test for significance of the overall fit, many individual parameters gave poor *t*-test results. Also the corresponding atomic hydrophobicity parameters failed to indicate the changes in partition coefficients of isomeric hydrocarbons differing in extent of branching. In the second try the carbons were classified mainly in terms of their hybridization and number of non-hydrogen substituents. Since electron charge density has considerable effect on its hydrophobicity, that feature was built into the classification of hydrogen atoms, where

**Table II.** The Observed and Calculated Partition Coefficients of the Compounds Used To Construct the Hydrophobic Atomic Contributions

| no. | compd | log *P* obsd | log *P* calcd | Δ(calcd–obsd) | no. | compd | log *P* obsd | log *P* calcd | Δ(calcd–obsd) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $CHF_2Cl$ | 1.08 | 1.21 | 0.13 | 50 | $C_6H_5SO_2NH_2$ | 0.31 | 0.34 | 0.03 |
| 2 | $CHFCl_2$ | 1.55 | 1.53 | −0.02 | 51 | $C_6H_5SO_2CH_3$ | 0.48 | 0.79 | 0.31 |
| 3 | $CHCl_3$ | 1.97 | 1.86 | −0.11 | 52 | $C_6H_5SCH_3$ | 2.74 | 2.60 | −0.14 |
| 4 | $CH_2Cl_2$ | 1.25 | 1.16 | −0.09 | 53 | $C_6H_5SOCH_3$ | 0.55 | 0.55 | 0.00 |
| 5 | $CH_2F_2$ | 0.20 | 0.24 | 0.04 | 54 | 2,5-dihydrofuran | 0.46 | 0.46 | 0.00 |
| 6 | HCOOH | −0.54 | −0.54 | 0.00 | 55 | $C_6H_5OCH_3$ | 2.08 | 1.98 | −0.10 |
| 7 | $CH_3Br$ | 1.19 | 0.66 | −0.53 | 56 | *m*-$CH_3C_6H_4OH$ | 1.97 | 2.21 | 0.24 |
| 8 | $CH_3Cl$ | 0.91 | 0.95 | 0.04 | 57 | $(CH_3)_2CHNH_2$ | 0.26 | 0.21 | −0.05 |
| 9 | $CH_3F$ | 0.51 | 0.35 | −0.16 | 58 | $CH_3(CH_2)_2NH_2$ | 0.48 | 0.46 | −0.02 |
| 10 | $CH_3I$ | 1.51 | 1.51 | 0.00 | 59 | $C_6H_5NH_2$ | 0.90 | 1.11 | 0.21 |
| 11 | $CH_3NO_2$ | −0.34 | −0.16 | 0.18 | 60 | *m*-$CH_3C_6H_4N(CH_3)_3I$ | −2.31 | −1.88 | 0.43 |
| 12 | $CH_4$ | 1.09 | 0.74 | −0.35 | 61 | *m*-$C_6H_4(NO_2)_2$ | 1.49 | 1.63 | 0.14 |
| 13 | $H_2NCONH_2$ | −1.09 | −1.31 | −0.22 | 62 | $C_2H_5NO_2$ | 0.18 | 0.46 | 0.28 |
| 14 | $H_2NCSNH_2$ | −1.06 | −0.88 | 0.18 | 63 | $[(CH_3)_2CH]_2NNO$ | 1.63 | 1.63 | 0.00 |
| 15 | $CH_3OH$ | −0.71 | −0.74 | −0.03 | 64 | $HOOCC_6H_3$, 2-OH, 5-NO | 2.34 | 1.94 | −0.40 |
| 16 | $CH_3NH_2$ | −0.57 | −0.53 | 0.04 | 65 | $CH_2OHCH_2F$ | −0.92 | −0.38 | 0.54 |
| 17 | $CF_2ClCF_2Cl$ | 2.82 | 2.69 | −0.13 | 66 | $CH_3CHF_2$ | 0.75 | 0.88 | 0.13 |
| 18 | $CCl_2{=}CCl_2$ | 2.60 | 2.66 | 0.06 | 67 | $C_6H_5F$ | 2.27 | 2.27 | 0.00 |
| 19 | $CF_3CF_3$ | 2.00 | 2.04 | 0.04 | 68 | $C_2H_5Cl$ | 1.43 | 1.56 | 0.13 |
| 20 | $CHBrClCF_3$ | 2.30 | 2.30 | 0.00 | 69 | *m*-$IC_6H_4NO_2$ | 2.94 | 3.01 | 0.07 |
| 21 | $CHCl{=}CCl_2$ | 2.29 | 2.17 | −0.12 | 70 | $HSCH_2COOH$ | 0.09 | 0.39 | 0.30 |
| 22 | $CH{\equiv}CH$ | 0.37 | 0.36 | −0.01 | 71 | 5-iodouracil | 0.04 | 0.03 | −0.01 |
| 23 | $CCl_3CONH_2$ | 1.04 | 1.24 | 0.20 | 72 | $(C_2H_5)_2S$ | 1.95 | 2.09 | 0.14 |
| 24 | $CF_2{=}CH_2$ | 1.24 | 1.24 | 0.00 | 73 | ethylenethiourea | −0.66 | −0.84 | −0.18 |
| 25 | $CF_3CONH_2$ | 0.12 | 0.26 | 0.14 | 74 | $CH_2{=}CH{=}CH_2$ | 1.45 | 1.25 | −0.20 |
| 26 | $CH_2BrCOOH$ | 0.41 | 0.55 | 0.14 | 75 | $CH_3C{\equiv}CH$ | 0.94 | 0.95 | 0.01 |
| 27 | $CCl_3CH_3$ | 2.49 | 2.61 | 0.12 | 76[a] | $NH{=}C{=}NH$ | −0.50 | −0.50 | 0.00 |
| 28 | $CCl_3CH_2OH$ | 1.42 | 1.26 | −0.16 | 77 | $CHCl_2CH_3$ | 1.79 | 1.80 | 0.01 |
| 29 | $CF_3CH_2OH$ | 0.41 | 0.28 | −0.13 | 78 | $CHCl_2CH_2OH$ | 0.37 | 0.45 | 0.08 |
| 30 | $CH_2{=}CH_2$ | 1.13 | 1.08 | −0.05 | 79 | $CH_2{=}CHCH_3$ | 1.77 | 1.56 | −0.21 |
| 31 | $CH_2BrCONH_2$ | −0.52 | −0.12 | 0.40 | 80 | $CH_2{=}CHC_2H_5$ | 2.40 | 1.94 | −0.46 |
| 32 | pyridine | 0.62 | 0.50 | −0.12 | 81 | $C_6H_5Cl$ | 2.46 | 2.59 | 0.13 |
| 33 | 2-Br-$C_5H_4N$ | 1.38 | 1.52 | 0.14 | 82 | *m*-$IC_6H_4OH$ | 3.00 | 3.06 | 0.06 |
| 34 | 4-Br-$C_5H_4N$ | 1.54 | 1.52 | −0.02 | 83 | $CH_3COOH$ | −0.24 | −0.19 | 0.05 |
| 35 | $CH_3COCH_3$ | −0.24 | −0.28 | −0.04 | 84 | $CH_3C{\equiv}CCH_3$ | 1.46 | 1.55 | 0.09 |
| 36 | $CH_3COCH_2CH_3$ | 0.29 | 0.33 | 0.04 | 85 | $C_2H_5OH$ | −0.30 | −0.13 | 0.17 |
| 37 | $CH_3CH_2OCH_2CH_3$ | 0.77 | 0.97 | 0.20 | 86 | $CH_3OCH_3$ | 0.10 | −0.25 | −0.35 |
| 38 | $C_2H_5OCH_2CH_2OH$ | −0.54 | −0.38 | 0.16 | 87 | $CH_3COOCH_3$ | 0.18 | 0.01 | −0.17 |
| 39 | $CH_2{=}CHCN$ | −0.92 | −0.10 | 0.82 | 88 | $CH_2{=}CHOC_2H_5$ | 1.04 | 1.39 | 0.35 |
| 40 | $CH_3CN$ | −0.34 | −1.16 | −0.82 | 89 | $C_2H_5NH_2$ | −0.13 | 0.07 | 0.20 |
| 41 | $CH_3COOC_2H_5$ | 0.70 | 0.62 | −0.08 | 90 | $H_2NCH_2CHOHCH_3$ | −0.96 | −1.13 | −0.17 |
| 42 | $(CH_3)_3NC_2H_5I$ | −3.00 | −3.43 | −0.43 | 91 | $CH_3CONHCH_3$ | −1.05 | −0.83 | 0.22 |
| 43 | pyrrole | 0.75 | 0.78 | 0.03 | 92 | *m*-$ClC_6H_4NH_2$ | 1.88 | 1.86 | −0.02 |
| 44 | $C_6H_5NO_2$ | 1.85 | 1.73 | −0.12 | 93 | *m*-$CH_3C_6H_4NO_2$ | 2.40 | 2.16 | −0.24 |
| 45 | $C_6H_5NO$ | 2.01 | 2.27 | 0.26 | 94 | $CH_2OHCH_2NO_2$ | −0.42 | −0.88 | −0.46 |
| 46 | $(CH_3)_2NNO$ | −0.57 | 0.13 | 0.70 | 95 | $H_2NCON(NO)CH_3$ | −0.16 | −0.56 | −0.40 |
| 47 | $C_6H_5Br$ | 2.99 | 2.87 | −0.12 | 96 | $CH_2FCH_2N(NO)CONHCH_2CH_2Cl$ | 0.95 | 0.78 | −0.17 |
| 48 | $C_6H_5I$ | 3.25 | 3.12 | −0.13 | 97 | *n*-$C_5H_{11}F$ | 2.33 | 2.11 | −0.22 |
| 49 | $C_6H_5SH$ | 2.52 | 2.22 | −0.30 | 98 | $C_6H_5NHSO_2CH_2F$ | 1.35 | 1.01 | −0.34 |

[a] Partition coefficient in Pr. pentanols, corrected to estimated octanol-water values.

the charge density of its point of attachment was considered (see Table I). The only difference between the second classification and the third one (Table I) was the active α-hydrogens. They were not given any special class in the earlier classification. However, it was found that the hydrophobicity of many compounds having that type of hydrogen was overestimated. The active α-hydrogens were therefore classified with the hydrogens attached to electronegative atoms.

The hydrophobic contributions as given in Table I were generated from the partition coefficients of 98 compounds[18] in water–octanol system by using the least-squares technique. The structure of these compounds and their observed and calculated partition coefficients are given in Table II. We selected these compounds because they contain the atom environments used in the various dihydrofolate reductase inhibitors. The calculated values have a standard deviation of 0.240 and a correlation coefficient of 0.980, and satisfied the *F* test at the level of 99.99%. The molar refractivity of the various atoms, as used in the present work, is given in Table I.[12]

The hydrophobic contribution in Table I indicates that for the carbons of similar hybridization, substitution decreases the hydrophobicity. This is an expected phenomenon, since (i) the surface area for interaction with solvent decreases with substitution and (ii) for carbon the major interaction is hydrophobic in nature. The same reason may partly be responsible for the increase in hydrophobicity in the order $C_{sp^3} < C_{sp^2} < C_{sp}$. That charge density has great influence on the hydrophobicity is evident from the substantially decreased values for carbonyl and carboxyl carbons. For oxygen the major interaction is hydrophilic and therefore the surface effect acts in the opposite direction. Thus Alcoholic oxygen is more hydrophilic than aliphatic ether oxygen and phenolic oxygen is more hydrophilic than aromatic ether oxygen. Phenolic oxygen is more hydrophobic than alcoholic oxygen, since the electron delocalization decreases its hydrogen bonding capacity. The reason for the strong hydrophobic nature of double-bonded oxygen is not obvious. For amino and

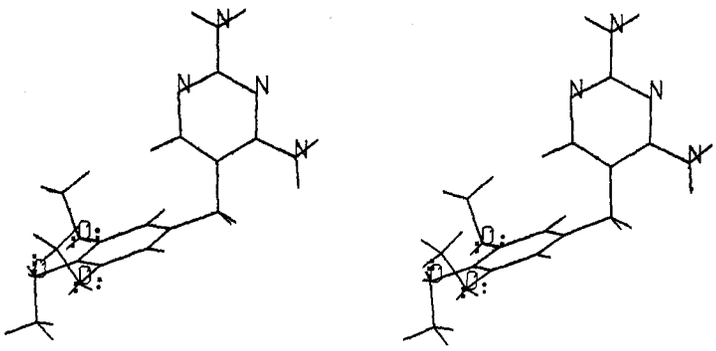


**Figure 1.** Stereo view of the minimum-energy conformation of trimethoprim.

**Table III.** Observed and Calculated[a] *E. coli* Dihydrofolate Reductase Inhibition Data[b] of Various Pyrimidines and Triazines Used To Construct the Model

| no. | substituents | log $1/C_{obsd}$ | log $1/C_{calcd}$ | Δ(calcd-obsd) |
|---|---|---|---|---|
| | Pydimidines (IV) | | | |
| 1 | 3′,5′-$(CH_2OH)_2$ | 5.31 | 6.60 (5.77) | 1.29 (0.46) |
| 2 | H | 6.18 | 5.91 (5.95) | −0.27 (−0.23) |
| 3 | 4′-$NO_2$ | 6.20 | 6.48 (6.43) | 0.28 (0.23) |
| 4 | 3′-$CH_2OH$ | 6.28 | 6.38 (6.32) | 0.10 (0.04) |
| 5 | 4′-$NH_2$ | 6.30 | 6.54 (6.48) | 0.24 (0.18) |
| 6 | 4′-Cl | 6.45 | 6.77 (6.66) | 0.32 (0.21) |
| 7 | 3′,4′-$(OH)_2$ | 6.46 | 6.29 (6.29) | −0.17 (−0.17) |
| 8 | 3′-OH | 6.47 | 6.03 (6.08) | −0.44 (−0.39) |
| 9 | 4′-$CH_3$ | 6.48 | 6.32 (6.30) | −0.16 (−0.18) |
| 10 | 4′-$OCF_3$ | 6.57 | 6.52 (6.46) | −0.05 (−0.11) |
| 11 | 3′-$CH_2OCH_3$ | 6.59 | 6.38 (6.32) | −0.21 (−0.27) |
| 12 | 3′-Cl | 6.65 | 7.14 (6.48) | 0.49 (−0.17) |
| 13 | 3′-$CH_3$ | 6.70 | 6.39 (6.48) | −0.31 (−0.22) |
| 14 | 4′-$N(CH_3)_2$ | 6.78 | 6.78 (6.67) | 0.00 (−0.11) |
| 15 | 4′-$OCH_3$ | 6.82 | 6.52 (6.46) | −0.30 (−0.36) |
| 16 | 4′-Br | 6.82 | 7.25 (7.07) | 0.43 (0.25) |
| 17 | 4′-$NHCOCH_3$ | 6.89 | 6.93 (6.80) | 0.04 (−0.09) |
| 18 | 3′-$OSO_2CH_3$ | 6.92 | 6.77 (6.91) | −0.15 (−0.01) |
| 19 | 3′-$OCH_3$ | 6.93 | 6.64 (7.25) | −0.29 (0.32) |
| 20 | 3′-Br | 6.96 | 7.96 (7.14) | 1.00 (0.18) |
| 21 | 3′-$CF_3$ | 7.02 | 6.79 (7.11) | −0.23 (0.09) |
| 22 | 3′-$CF_3$, 4′-$OCH_3$ | 7.69 | 7.40 (7.62) | −0.29 (−0.07) |
| 23 | 3′, 4′-$(OCH_3)_2$ | 7.72 | 7.25 (7.76) | −0.47 (0.04) |
| 24 | 3′, 5′-$(OCH_3)_2$ | 8.38 | 7.69 (8.30) | −0.69 (−0.08) |
| 25 | 3′, 4′, 5′-$(OCH_3)_3$ | 8.87 | 8.30 (8.80) | −0.57 (−0.07) |
| | Triazines (V) | | | |
| 26 | $C_6H_4$-4′-COOH | 2.72 | 3.59 (3.55) | 0.87 (0.83) |
| 27 | $C_6H_4$-3′-COOH | 3.89 | 5.08 (3.81) | 1.19 (−0.08) |
| 28 | $CH_3$ | 4.32 | 4.77 (4.47) | 0.45 (0.15) |
| 29 | $C_6H_4$-4′-$COOC_2H_5$ | 4.42 | 3.59 (3.55) | −0.83 (−0.87) |
| 30 | $CH_2C_6H_5$ | 4.47 | 4.99 (5.43) | 0.52 (0.96) |
| 31 | $C_6H_4$-4′-$C_6H_5$ | 5.24 | 5.74 (5.71) | 0.50 (0.47) |
| 32 | $(CH_2)_2CH_3$ | 5.33 | 4.89 (4.92) | −0.44 (−0.41) |
| 33 | $C_6H_4$-3′-$OCH_3$ | 5.41 | 5.08 (5.49) | −0.33 (0.08) |
| 34 | $(CH_2)_2C_6H_5$ | 5.42 | 5.92 (5.10) | 0.50 (−0.32) |
| 35 | $C_6H_5$ | 5.52 | 5.28 (5.58) | −0.24 (0.06) |
| 36 | $C_6H_4$-3′-$CF_3$ | 5.64 | 5.08 (5.65) | −0.56 (0.01) |
| 37 | $C_6H_4$-3′-Cl | 6.22 | 5.08 (5.89) | −1.14 (−0.33) |
| 38 | $C_6H_3$-3′, 4′-$Cl_2$ | 6.80 | 6.69 (6.78) | −0.11 (−0.02) |
| 39 | $(CH_2)_3C_6H_5$ | 7.20 | 7.20 (7.20) | 0.00 (0.00) |

[a] Values within parentheses represent the calculated values if no alternate binding modes are kept in the model. [b] *C* represents the concentration of the inhibitor causing 50% inhibition.

amide nitrogens we did not consider the influence of substitution. However, such distinction may eventually be necessary. Formal oxidation number of carbon holding the halogen affects its hydrophobicity (see Table I, types 35–37 and 39–41). The hydrophobicity goes through a maximum when the oxidation number is 2. Although the reason is not very obvious, it seems to be due to two opposing factors, increase in hydrophobicity due to decreased charge separation and decrease in hydrophobicity due to crowding.

It is possible that the values we assigned for various atoms may not account for all structural features that affect the hydrophobicity. Further research is necessary to generate more widely applicable atomic contributions of various physicochemical properties.

During the preparation this manuscript, Broto et al.[19] published the atomic contributions of partition coefficients. Their classification was somewhat different from ours, since they did not explicitly include hydrogen. Their heteroatomic classification was more exhaustive where they considered various conjugation effects. Their carbon classification resembles our classification in many respects. Broto et al. used a large number of atom types (222) and also a large number of compounds (1868) to evaluate their atomic contribution. However, since they did not classify or separate the hydrogen atom contribution, their atomic (which is not truly atomic in many cases) values cannot be compared with those reported here. For example, the methyl group in methyl and ethyl chlorides are different in our classification but not in theirs. Hydrogen plays an important role in the ligand binding process. If attached to an electronegative atom it may form a hydrogen bond with the receptor site. A hydrogen attached to carbon may not interact strongly with an attractive or soft repulsive site pocket, but if it goes to a sterically blocked region, that mode should be forbidden. Therefore, in our model building process we ought to keep hydrogen atoms in the ligand molecule explicity.

**Development of *E. coli* DHFR Receptor Model.** Twenty-five 2,4-diamino-5-(substituted benzyl)pyrimidines[8] and 14 2,4-diamino-5-substituted-triazines[9] (Table III) were used to construct the *E. coli* DHFR receptor model. However, in quantitative structure–activity relationships, there is always danger of using the biological data from different sources, since as we found earlier,[4] the same compounds were reported to have different $I_{50}$ values by different authors. For instance, the trimethoprim inhibition reported by Li et al.[8] and Baker[9] do not agree. The reasons for inconsistency may be the extent of purification of the enzyme and differences in the assay conditions. The value from Hansch's group was used in case of dispute.

The basis of the present receptor model started from our previous working experience on DHFR from different species,[2-4] from the conformational studies on some inhibitors of this species, and from the X-ray data on the trimethoprim and *E. coli* DHFR binary comples.[20,21] The crystallographic study by Volz et al.[5] showed that the

(19) Broto, P.; Moreau, G.; Vandycke, C. *Eur. J. Med. Chem.* 1984, *19*, 71.
(20) Baker, D. J.; Beddell, C. R.; Champness, J. N.; Goodford, P. J.; Norrington, F. E. A.; Smith, D. R.; Stammers, D. K. *FEBS Lett.* 1981, *126*, 49.

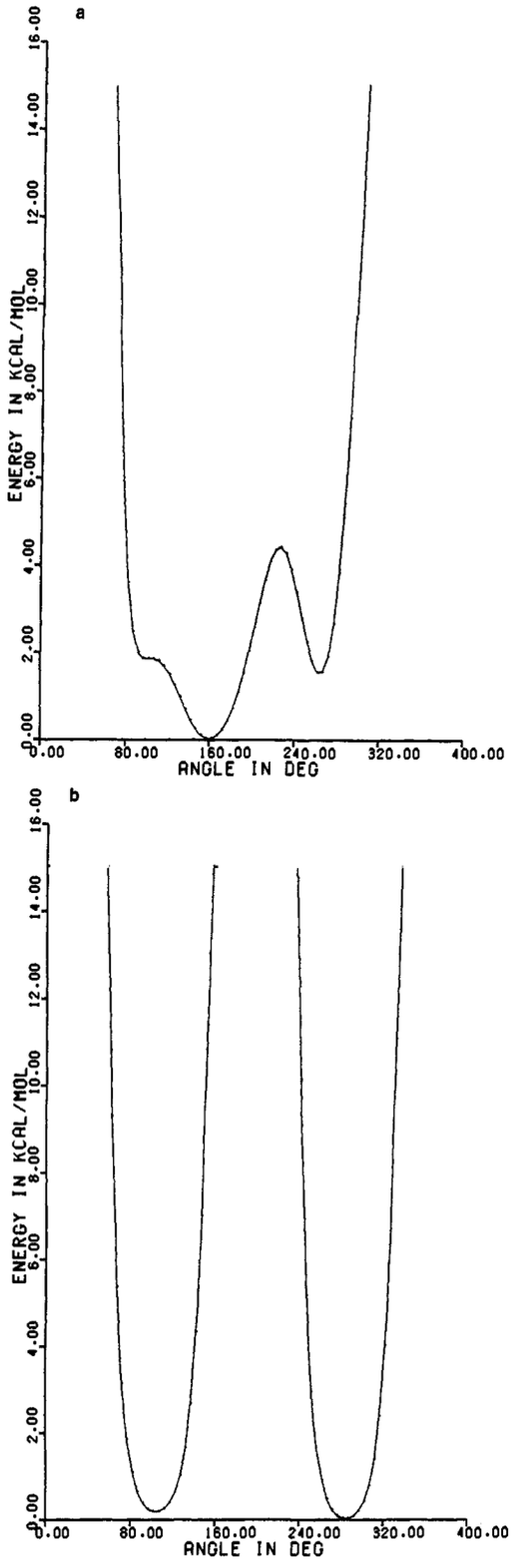



**Figure 2.** Trimethoprim (I: A = D = F = O, B = E = G = $CH_3$) energy relative to the global minimum, (a) varying $\omega_3$(C4–C5–C7–C1′), (b) varying $\omega_4$(C5–C7–C1′–C2′). All other dihedral angles are held fixed at the value of its global minimum-energy conformation.

2,4-diamino heterocyclic rings of different inhibitors occupy analogous position at the receptor site. That prompted us to develop a comon three-dimensional receptor model for rat liver DHFR. It explained and predicted the inhibition data of a large number of compounds. In the present work the same line has been maintained by allowing the heterocyclic ring to bind in the same mode. For that three hydrogen bonding site pockets, 1, 2, and 3, were generated. To create the site pockets for the substituents of the heterocyclic ring, the low-energy conformations of some highly active compounds were employed.

The constant valence structure conformational analysis (see ref 4 for details of the method) of trimethoprim (I: A = D = F = O, B = E = G = $CH_3$) showed the (probably global) minimum energy conformation at $\omega_1$(C5–C4–N–H) = 20°, $\omega_2$(N3–C2–N–H) = 10° $\omega_3$(C4–C5–C7–C1′) = 160°, $\omega_4$(C5–C7–C1′–C2′) = 285° (−75°),[22] $\omega_5$(C2′–C3′–A–B) = 90°, $\omega_6$(C3′–A–C–H) = 60°, $\omega_7$(C3′–C4′–D–E) = 270° (−90°), $\omega_8$(C4′–D–C–H) = 60°, $\omega_9$(C4′–C5′–F–G) = 90°, and $\omega_{10}$(C5′–F–C–H) = 60° (Figure 1). The energy changes with respect to two important dihedral angles ($\omega_3$ and $\omega_4$) when other dihedral angles are kept fixed at the values of their minimum energy conformation are shown in Figure 2. The molecule seems to be rather flexible in terms of the rotational energy around the dihedral angle $\omega_3$. It can attain any value between 90 and 270 at the cost of 4 kcal/mol or less. However, from the schematic illustration of Baker et al.[20] as well as from the proton and $^{13}C$ NMR studies of Birdsall et al.[21] we found that the conformation of trimethoprim bound to *E. coli* DHFR is very similar to its minimum-energy conformation. Their values for $\omega_3$ = 191° (−169°) and $\omega_4$ = 73° very closely resemble the enantiomeric form of our minimum-energy conformation. Therefore three site pockets (4, 5, and 6) were generated to bind the phenyl ring and six site pockets (7–12) to bind the three methoxy substituents in its minimum-energy conformation (Figure 3). When the other torsion angles are kept fixed at the values of the global minimum-energy conformation, the conformational behavior of the three methoxyl groups are as follows: For $\omega_5$ there are three minima at 0° (6.8 kcal/mol), 90° (0.0 kcal/mol), and 295° (0.5 kcal/mol). For $\omega_7$ there are two minima at 90° (21.7 kcal/mol) and 270° (0.0 kcal/mol). For $\omega_9$ there are three minima at 90° (0.0 kcal/mol), 180° (6.8 kcal/mol), and 240° (0.6 kcal/mol). The terminal methoxyl groups are quite flexible near 90°. For example, $\omega_5$ can attain any conformation from 55° to 110° at the cost of 1 kcal/mol or less. The middle methoxyl group is compariatively rigid; it can attain any conformation from 250° to 290° at the cost of 1 kcal/mol or less, but the energy increases sharply outside this range. From our earlier conformational studies[4] on phenyltriazines (II) we know that the rotation of the phenyl ring is highly restricted. The low-energy conformations are those in which the phenyl ring is almost perpendicular, $\omega$(C4–N5–C1′–C2′) = 75°, to the triazine ring. Therefore three site pockets (15–17) were generated to bind the phenyl ring and two site pockets (18 and 19) to bind its 3′- and 4′-substituents, respectively (Figure 4) at its minimum-energy conformation. In the triazines a remarkable increase in the DHFR inhibition is observed when a 3-phenyl propyl group is introduced at the 5-position (III: $n$ = 3). One obvious reason may be that the phenyl ring is going to some strong hydrophobic pocket. In order to get some idea about the location of the hydrophobic pocket, the conformational behavior of this molecule was studied. We found the global minimum energy conformation at $\omega_1$(C4–N5–C′1–C′2) = 275° (−85°),

(21) Birdsall, B.; Roberts, G. C. K.; Feeney, J.; Dann, J. G.; Burgen, A. S. V. *Biochemistry* **1983**, 22, 5597.

(22) For the sign convention for torsion angle, we followed Klyne and Prelog: Klyne, W.; Prelog, V. *Experientia* **1960**, *16*, 521. The negative equivalents of the torsion angles greater than 180° are presented within parentheses to make the enantiomeric conformations more obvious.

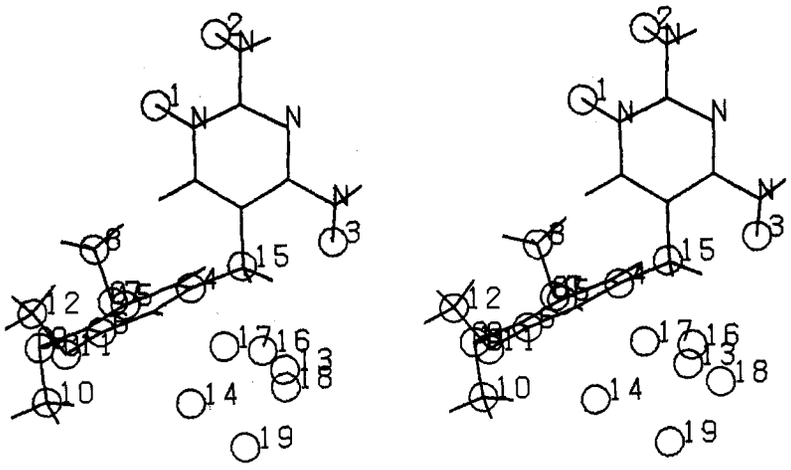


**Figure 3.** Stereo drawing of the minimum-energy conformation of trimethoprim superimposed over the site pockets. The radius of the circles representing the site pockets are half of their actual value, for the clarity of the picture.

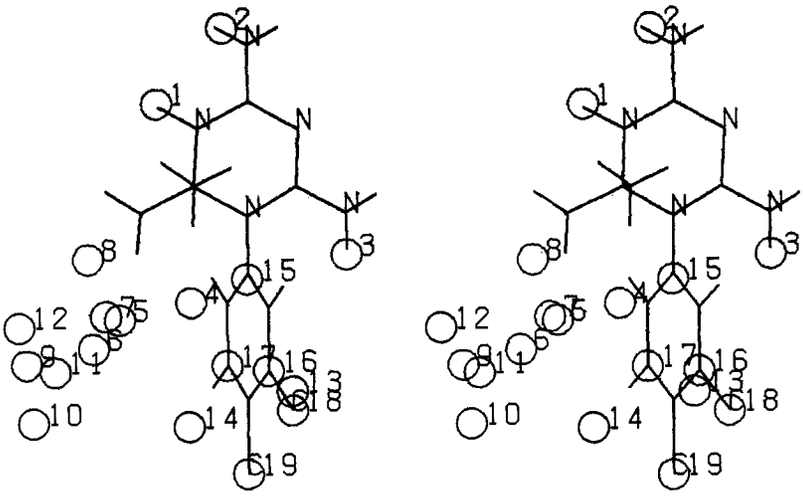


**Figure 4.** Stereo drawing of the minimum energy conformation of a triazine (II: A = B = Cl, D = H), superimposed over the site pockets.

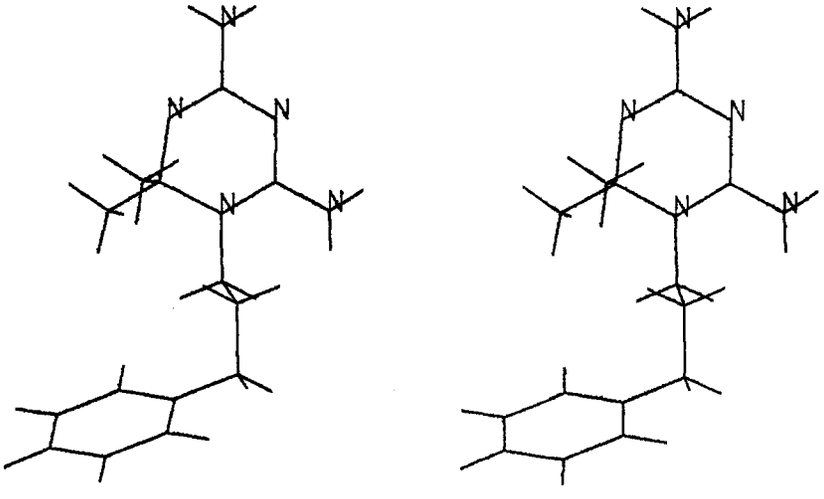
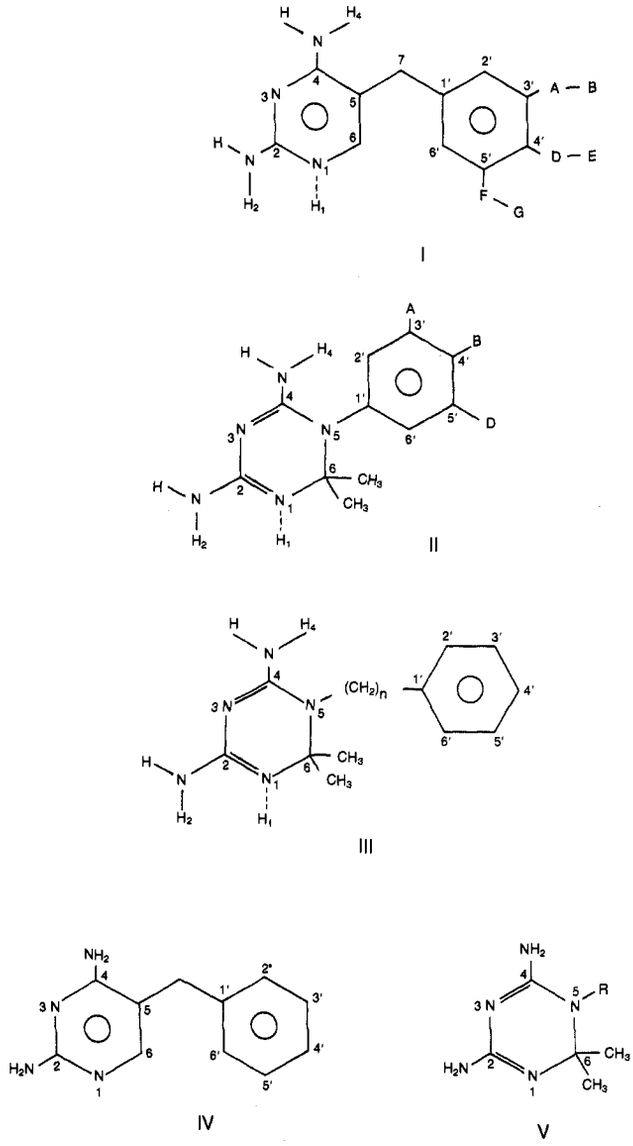


**Figure 5.** Stereo view of the minimum-energy conformation of triazine (III: $n$ = 3).

$\omega_2$(N5–C′1–C′2–C′3) = 180°, $\omega_3$(C′1–C′2–C′3–C1′) = 60°, and $\omega_4$(C′2–C′3–C1′–C2′) = 70° with the amino groups planar and the methyl groups staggered (Figure 5). The energy change with respect to any torsion angle when other angles are kept fixed at the global minimum-energy conformation values is as follows: For $\omega_1$ there are two minima at 70° (1.3 kcal/mol) and 275° (0.0 kcal/mol). Energy increases sharply as this bond is rotated in either direction. For $\omega_2$ there is only one minimum at 180°, but it can attain any conformation from 160° to 200° at the cost of 1 kcal/mol or less. Beyond that the energy increases sharply. For $\omega_3$ there are three minima at 60° (0.0 kcal/mol), 180° (0.8 kcal/mol), and 260° (4.0 kcal/mol). The two energy peaks separating these minima are approximately 4.5 kcal/mol above the global minimum. Due to the $c_2$ axis of symmetry in the phenyl ring, $\omega_4$ is rotated from 0° to



180°. Within this range it has one minimum at 70°. However, it can attain any value from 50° to 90° at the cost of 1 kcal/mol or less. Initially the hydrophobic site pockets were generated at its minimum-energy conformation. However, soon we realized that the position of the phenyl group in this conformation is very close to the phenyl group of the benzylpyrimidines. This leads to two distinct situations. If this pocket is considered to be strongly hydrophobic, then the phenyl group of the benzylpyrimidines goes to this pocket, thereby overestimating the binding energy. Alternatively, if this pocket is not very attractive to the phenyl group, it fails to account for the tight binding of this triazine. Therefore these site pockets (13 and 14) were relocated to a somewhat remote place corresponding to another local minimum-energy conformation with $\omega_3 = 180°$ (Figure 6). The purposes of the various site pockets are summarized in Table IV, and their relative positions are shown in Table V.

Since site pockets of type 1 were designed for hydrogen bonding, the interaction energies were correlated to the formal charge density over the hydrogens as is obtained from CNDO/2 calculation (see Table VI). However, since all these hydrogens had the same atom type, we ought to use the same value for their atomic charges. A charge of 0.170 au was assigned for these hydrogens. Site pockets of type 2 and 9 were assumed to be hydrophobic pockets. We took log $P$ as a measure of the hydrophobicity and

**Table IV.** Description of the Site Pockets

| sites | type | description |
|---|---|---|
| 1–3 | 1 | hydrogen-bonding site pockets to bind the protonated diamino heterocyclic ring |
| 4–6 | 2 | hydrophobic pocket to bind the phenyl ring of the benzyl pyrimidines |
| 7 | 3 | to bind the first atom of the 3′-substituents of the pyrimidines |
| 8 | 4 | to bind the second atom of the 3′-substituents of the pyrimidines |
| 9, 10 | 5 | to bind the first and second atoms of the 4′-substituents, respectively |
| 11, 12 | 6 | to bind the first and second atoms of the 5′-substituents, respectively |
| 13, 14 | 9 | hydrophobic pocket to bind the phenyl ring of the triazine **39** |
| 15–17 | 2 | hydrophobic pocket to bind the phenyl ring attached at the 5-position of the triazines |
| 18 | 7 | to bind the 3′-substituents of the triazines |
| 19 | 8 | to bind the 4′-substituents of the triazines |

**Table V.** Coordinates of the Site Points in Angstroms

| site[a] | $x$ | $y$ | $z$ |
|---|---|---|---|
| 1 | −0.9362 | −0.5776 | 0.0022 |
| 2 | 0.5869 | −2.3564 | −0.5912 |
| 3 | 3.4525 | 2.9353 | 0.2975 |
| 4 | −0.1467 | 4.0813 | −0.5214 |
| 5 | −1.7951 | 4.4870 | −2.4306 |
| 6 | −2.4090 | 5.1689 | −0.0462 |
| 7 | −2.1131 | 4.4021 | −3.7501 |
| 8 | −2.5449 | 3.0649 | −4.0155 |
| 9 | −3.9558 | 5.5357 | −1.9181 |
| 10 | −3.8004 | 6.9126 | −2.2719 |
| 11 | −3.2930 | 5.7127 | 0.8326 |
| 12 | −4.1262 | 4.6617 | 1.3287 |
| 13 | 2.1952 | 6.1432 | 5.3432 |
| 14 | −0.1903 | 6.9960 | 5.0124 |
| 15 | 1.1499 | 3.5122 | 0.0692 |
| 16 | 1.6371 | 5.6557 | −1.0163 |
| 17 | 0.6842 | 5.5490 | 1.3518 |
| 18 | 2.2037 | 6.5877 | −2.3720 |
| 19 | 1.1735 | 8.0771 | 0.2844 |

[a] The ligand molecules were superimposed by placing atom 1 of the heterocyclic rings (see structures IV and V) at the origin, atom 3 along the $x$ axis, and atom 5 in the $xy$ plane of the site coordinate frame. This defines the coordinate system used in this table. Any ligand atom coming within 0.7 Å of a site point is assumed to interact with the corresponding site pocket.

correlated the interactions with the atomic log $P$ values. For site pockets of type 5 we used molar refractivity, and for all other types of site pockets, both hydrophobicity and molar refractivity were used to correlate their interactions, assuming it to come partly from hydrophobic factors and partly from dispersive forces.

We initially supplied binding modes of the various molecules (Table VII) satisfying the purposes of the various site pockets. We tried to optimize the agreement between observed and calculated log (1/C) by using these binding modes and allowed switching over to another geometrically feasible binding mode only in the final step when it could not optimize the supplied mode. Switching to other mode was based on our earlier algorithm.[17] It is always advantageous to optimize the initially supplied binding modes, since that indicates any inherent faults in the model. As discussed earlier, the relocation of site pockets 13 and 14 was required since otherwise the optimization procedure would satisfy the demands of various pyrimidines by converting these site pockets to weakly attractive. Another somewhat different situation is observed with molecule **27**. Comparing its binding energy with that of molecule **35** (Table III), one can conclude that the carboxyl group at the 3′-position in **27** experiences

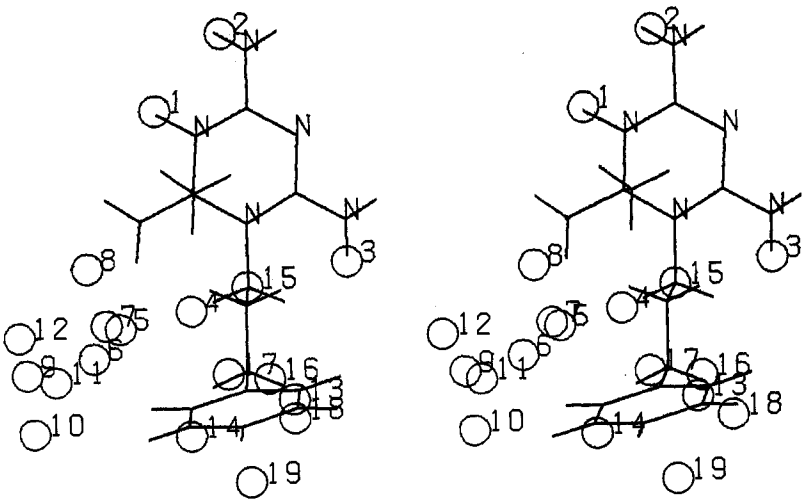


**Figure 6.** Stereo drawing of a local minimum-energy conformation of the triazine (III: $n$ = 3) superimposed over the site pocket.

**Table VI.** CNDO/2 Valence Shell Gross Atomic Charges on the Protonated Triazine and Pyrimidine Rings

| pyrimidine (I) | | triazine (II) | |
|---|---|---|---|
| atom | charge | atom | charge |
| N1 | 5.095 | N1 | 5.225 |
| H1 | 0.802 | H1 | 0.843 |
| C2 | 3.589 | C2 | 3.546 |
| N2 | 5.249 | N2 | 5.230 |
| H2 | 0.836 | H2 | 0.831 |
| H2′ | 0.823 | H2′ | 0.816 |
| N3 | 5.260 | N3 | 5.356 |
| C4 | 3.648 | C4 | 3.549 |
| N4 | 5.226 | N4 | 5.240 |
| H4 | 0.832 | H4 | 0.832 |
| H4′ | 0.842 | H4′ | 0.818 |
| C5 | 4.128 | N5 | 5.214 |
| H5[a] | 0.920 | H5[a] | 0.860 |
| C6 | 3.798 | C6 | 3.746 |
| H6 | 0.952 | C[b] | 4.082 |
| | | H[b] | 0.954 |
| | | H[b] | 0.952 |
| | | H[b] | 0.972 |
| | | C[b] | 4.069 |
| | | H[b] | 0.960 |
| | | H[b] | 0.960 |
| | | H[b] | 0.951 |

[a] 5-Unsubstituted compounds were used for this calculation. [b] The two methyl groups at the 6-position.

some soft repulsion. Any optimization procedure that does not consider the possibilities of alternate geometrically feasible binding modes will assign a repulsive interaction at the receptor site of this group with the corresponding site pocket. In reality as well as in our approach, the phenyl ring may undergo 180° rotation to produce an almost equally stable conformation[4] in which it can avoid such interaction (Figure 7), thereby giving higher binding energy. Since low binding energy is observed, one may wonder how to get the fit. There are several ways to explain the low binding energy of this molecule: (i) such rotation may be forbidden due to steric overlap of the carboxyl group with the receptor site in its alternate conformation, (ii) in the alternate conformation too, the carboxyl group goes to an equally repulsive site pocket, (iii) there is no direct interaction of this group with the receptor site, but rather the influence of this group on the charge density of the aromatic ring may be indirectly responsible. One of these possibilities may be confirmed by the future syntheses. For example, if simultaneous substitution at 3′- and 5′-positions markedly decreases the binding energy irrespective of their nature, the steric overlap may be responsible. If simultaneous substitution at 3′- and 5′-positions with groups having characteristics opposite to carboxyl group increases the activity markedly, the second interpretation may hold and so on. In a thorough receptor mapping, this information is essential and should be given due consideration in the process of model building.

The initially given binding modes and the final optimized binding modes of the various molecules are summarized in Table VII. The optimization was completed only after setting 11 inequality constraints. In addition to that, there were some preassigned lower limits on the coefficients. Site points of type 1 had a lower limit of 0.0, and all other coefficients had a lower limit of −10.0. These limitations were set to keep the interactions physically realistic and the basic structural moiety attractive to the receptor site. The optimized values of these coefficients are shown in Table VIII.

These coefficients gave a correlation coefficient of 0.893 and standard deviation of 0.530. The constraints in general worsen the standard deviation, since they disallow some parts of the coefficient space. For example, if the alternate binding modes are disregarded, the initial fixed binding modes (Table VII) give a correlation coefficient of 0.961 and a standard deviation of 0.326. However, constraints are unavoidable if the alternate binding modes of the molecules are considered. The statistics of the present studies are summarized in Table IX.

Table III shows that benzyltriazine (**30**) has quite a different binding energy from benzylpyrimidine (**2**). The present study explains this by giving them quite different binding modes. The reason is the difference in conformational behavior of the triazine from the pyrimidine. The minimum-energy conformation of this triazine, $\omega_1$(N5–C4–N–H) = 0°, $\omega_2$(N3–C2–N–H) = 0°, $\omega_3$(C4–N5–C′1–C1′) = 95°, and $\omega_4$(N5–C′1–C1′–C2′) = 120° (Figure 8), is quite different from that of the pyrimidines (Figure 1), and additionally the rotational behavior of the triazine around the methylene group (Figure 9) is quite different from the pyrimidine (Figure 2). One reason for such conformational differences is the dimethyl substituents at the 6-position of the triazine ring. This interpretation gives a plausible explanation for the fact that substitution at the 6-position of the pyrimidines decreases its activity markedly.[23] Such substitution decreases the stability of the active conformation.

(23) Roth, B.; Aig, E.; Lane, K.; Rauckman, B. S. *J. Med. Chem.* 1980, *23*, 535.

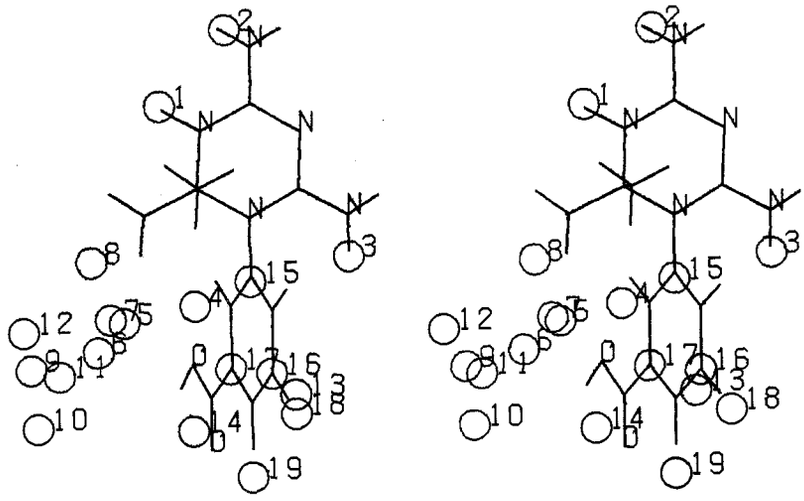


**Figure 7.** An alternative geometrically feasible binding mode of the triazine (**27**) to avoid the repulsive interaction of the carbonyl group with site 18.

**Table VII.** The Initially Supplied and the Final Best Fitted Binding Modes[a]

| | sup/obsd | site points | | | | | | | | | site points | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| no. | (I/II) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | no. | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 1 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 1 | D | | F | G | | | C7 | | | | |
| 2 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 2 | D | | F | | | | C7 | | | | |
| 3 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 3 | D | E | F | | | | C7 | | | | |
| 4 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 4 | D | | F | | | | C7 | | | | |
| 5 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 5 | D | E | F | | | | C7 | | | | |
| 6 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 6 | D | | F | | | | C7 | | | | |
| 7 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 7 | D | E | F | G | | | C7 | | | | |
| 8 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 8 | D | | F | | | | C7 | | | | |
| 9 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 9 | D | E | F | | | | C7 | | | | |
| 10 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 10 | D | E | F | | | | C7 | | | | |
| 11 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 11 | D | | F | | | | C7 | | | | |
| 12 | I | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 12 | D | | F | | | | C7 | | | | |
| | II | H1 | H2 | H4 | C1′ | C3′ | C5′ | | A | | D | | | | | | C7 | | | | |
| 13 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 13 | D | | F | | | | C7 | | | | |
| 14 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 14 | D | E | F | | | | C7 | | | | |
| 15 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 15 | D | E | F | | | | C7 | | | | |
| 16 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 16 | D | | F | | | | C7 | | | | |
| 17 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 17 | D | E | F | | | | C7 | | | | |
| 18 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 18 | D | | F | | | | C7 | | | | |
| 19 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 19 | D | | F | | | | C7 | | | | |
| 20 | I | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | | 20 | D | | F | | | | C7 | | | | |
| | II | H1 | H2 | H4 | C1′ | C3′ | C5′ | | A | | D | | | | | | C7 | | | | |
| 21 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 21 | D | | F | | | | C7 | | | | |
| 22 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 22 | D | E | F | | | | C7 | | | | |
| 23 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 23 | D | E | F | | | | C7 | | | | |
| 24 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 24 | D | | F | G | | | C7 | | | | |
| 25 | I/II | H1 | H2 | H4 | C1′ | C3′ | C5′ | A | B | 25 | D | E | F | G | | | C7 | | | | |
| 26 | I/II | H1 | H2 | H4 | | | | | | 26 | | | | | | | C1′ | C3′ | C5′ | A | B |
| 27 | I/II | H1 | H2 | H4 | | | | | | 27 | | | | | | | C1′ | C3′ | C5′ | A | B |
| 28 | I | H1 | H2 | H4 | H[b] | | | | | 28 | | | | | | | C′1 | | | | |
| | II | H1 | H2 | H4 | | | | | | 28 | | | | | | | C′1 | | | | |
| 29 | I/II | H1 | H2 | H4 | | | | | | 29 | | | | | | | C1′ | C3′ | C5′ | A | B |
| 30 | I/II | H1 | H2 | H4 | | | | | | 30 | | | | | | | C′1 | C2′ | | C′3 | |
| 31 | I/II | H1 | H2 | H4 | | | | | | 31 | | | | | | | C′1 | C′3 | C5′ | A | B |
| 32 | I | H1 | H2 | H4 | | | | | | 32 | | | | | | | C′1 | C′3 | | H[c] | |
| | II | H1 | H2 | H4 | | | | | | 32 | | | | | | | C′1 | | H[c] | | |
| 33 | I/II | H1 | H2 | H4 | | | | | | 33 | | | | | | | C1′ | C3′ | C5′ | A | B |
| 34 | I | H1 | H2 | H4 | | | | | | 34 | | | | | | | C′1 | C1′ | H6′ | C2′ | C5′ |
| | II | H1 | H2 | H4 | C′2 | | C2′ | | | | | | C3′ | | | | C′1 | | | | |
| 35 | I/II | H1 | H2 | H4 | | | | | | 35 | | | | | | | C1′ | C3′ | C5′ | A | B |
| 36 | I/II | H1 | H2 | H4 | | | | | | 36 | | | | | | | C1′ | C3′ | C5′ | A[d] | B |
| 37 | I | H1 | H2 | H4 | | | | | | 37 | | | | | | | C1′ | C3′ | C5′ | A | B |
| | II | H1 | H2 | H4 | | | | | | | | | | | | | C1′ | C3′ | C5′ | A | B |
| 38 | I | H1 | H2 | H4 | | | | | | 38 | | | | | | | C1′ | C3′ | C5′ | A | B |
| | II | H1 | H2 | H4 | | | | | | | | | | | | | C1′ | C3′ | C5′ | | B |
| 39 | I/II | H1 | H2 | H4 | | | | | | 39 | | | | | C3′ | C5′ | C′1 | | C′3 | | |

[a] Atom labels of the pyrimidines correspond to structure I and those of the triazines, structures II and III. [b] 5-Methyl hydrogen. [c] Terminal hydrogen of the *n*-propyl group. [d] One fluorine atom.

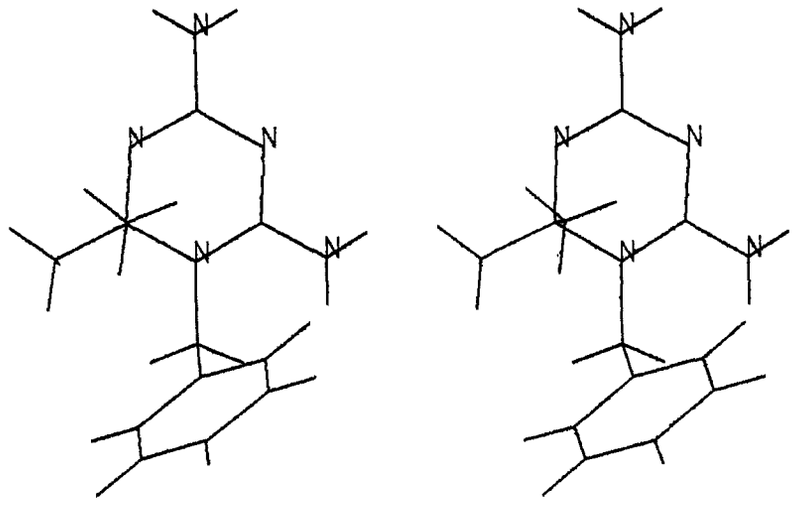


**Figure 8.** Stereo view of the minimum energy conformation of 5-benzyltriazine (**30**).

**Table VIII.** Proportionality Constants of the Interaction with Various Parameters[a]

| site type | hydrophobicity | molar refractivity | charge density |
|---|---|---|---|
| 1 | 0.0 | 0.0 | 9.0580 |
| 2 | 0.7127 | 0.0 | 0.0 |
| 3 | -0.2947 | 0.3339 | 0.0 |
| 4 | 0.2376 | 0.2620 | 0.0 |
| 5 | 0.0 | 0.1708 | 0.0 |
| 6 | 1.2189 | 0.1111 | 0.0 |
| 7 | 0.0 | -0.0002 | 0.0 |
| 8 | 1.1063 | 0.0969 | 0.0 |
| 9 | 3.0212 | 0.0 | 0.0 |

[a] Units of the constants are (unit of biological activity/unit of parameter).

The above conclusion suggests that the 5-benzyltriazines having no substituents at the 6-position, if chemically stable, may be tried as *E. coli* dihydrofolate reductase inhibitors. Substitution by a $CH_2CH_2C_6H_5$ group for one methylene hydrogen of trimethoprim or similar benzylpyrimidines of high activity may enhance the potency markedly.

The model was allowed to predict the biological data of five pyrimidines and five triazines (Table X). The predicted values gave a correlation coefficient of 0.941 and standard deviation of 0.893. Except for one compound (**40**) the predicted values were excellent. The binding modes of these compounds are given in Table XI. Note that a better fit to the original data is obtained when no alternate modes are allowed (Table IX, study II), but allowing alternate modes gives better predictions (Table IX, study III vs. IV).

The conformational energy was not used in the correlation equation for the following reasons: (i) Site points were generated in the neighborhood of minimum-energy conformations of some inhibitors. (ii) The conformational energy parameters involving many heteroatoms are not very refined. (iii) The conformational energies were evaluated with use of fixed valence structure formalism in order to keep the computation time to a manageable level. However, relaxation in bond length and bond angles often changes the relative energy of the conformations compared to the global minimum-energy conformation by substantial amount. (iv) The methotrexate conformation at the receptor site was found to have considerably higher energy compared to the calculated minimum-energy conformation[24] of the isolated molecule.

The present model is based on the idea that the diaminoheterocyclic ring of the pyrimidines and triazines binds at the same site of the receptor. This feature holds for some *Lactobacillus casei* DHFR inhibitors.[14] The X-ray structure of the binary complex of *E. coli* DHFR and two of its inhibitors, namely, trimethoprim[20] and methotrexate,[25] are known. However, the strain of the *E. coli* was different. The amino acid sequences of the two enzymes differ at three positions.[20] The difference Fourier electron density between the methotrexate and trimethoprim complexes shows that the pyrimdine ring of methotrexate may be slightly tilted relative to that in trimethoprim away from the Phe-31. Some other small differences are also observed in the amino acid residues of the enzymes. However, in absence of information regarding the structures of the pure enzymes, and the high-resolution X-ray data of the complexes using the enzyme from the same strain of *E. coli*, about all we can suggest is that trimethoprim and methotrexate interact at the same active site of the receptor. The present model is based on a similar hypothesis. Furthermore, the minimum-energy conformation of trimethoprim on which the present model is based is very similar to the X-ray crystallographic conformation at the receptor site.[21] Since the crystallographic coordinates of the trimethoprim and DHFR complex were not available to the authors, the site pockets are compared here with the crystal structure of methotrexate and DHFR complex. The site points 1, 2, 1nd 3 were placed so that they bind the protonated 2,4-diaminopyrimidine part of the methotrexate ring. That superposition showed three interesting features of the present site pockets: (i) All the site pockets are sterically accessible, i.e., they do not overlap with the receptor site atoms. (ii) Site points 1, 2, and 3 are surrounded by groups capable of hydrogen bonding. (iii) The hydrophobic pockets of type 2 and 9 are surrounded by various hydrophobic groups (see Figure 10).

The objective of any QSAR is the mapping of the receptor site, so that more meaningful changes in the ligand molecule may be made in future structure–activity studies. However, modeling a biological system is an extremely difficult problem. It is almost impossible to get a definite solution to this problem unless the large number of vari-

(24) Spark, M. J.; Winkler, D. A.; Andrews, P. R. *Int. J. Quantum Chem. Quantum Biol. Symp.* **1982**, *9*, 321.
(25) Matthews, D. A.; Alden, R. A.; Bolin, J. T.; Freer, S. T.; Hamlin, R.; Xuong, N.; Kraut, J.; Poe, M.; Williams, M.; Hoogsteen, K. *Science* **1977**, *197*, 452.

**Table IX.** Statistics of the Study

| study | | no. of compnds | no. of vars | no. of cnstrnts | correl coeff | SD | max error |
|---|---|---|---|---|---|---|---|
| I | original data with alternate binding modes | 39 | 13 | 11 | 0.893 | 0.530 | 1.29 |
| II | original data without alternate binding modes | 39 | 13 | 0 | 0.961 | 0.326 | 0.96 |
| III | test data (Table X) using the results of study I | 10 | 0 | 0 | 0.941 | 0.893 | 2.10 |
| IV | test data (Table X) using the results of study II | 10 | 0 | 0 | 0.779 | 2.627 | 7.86 |

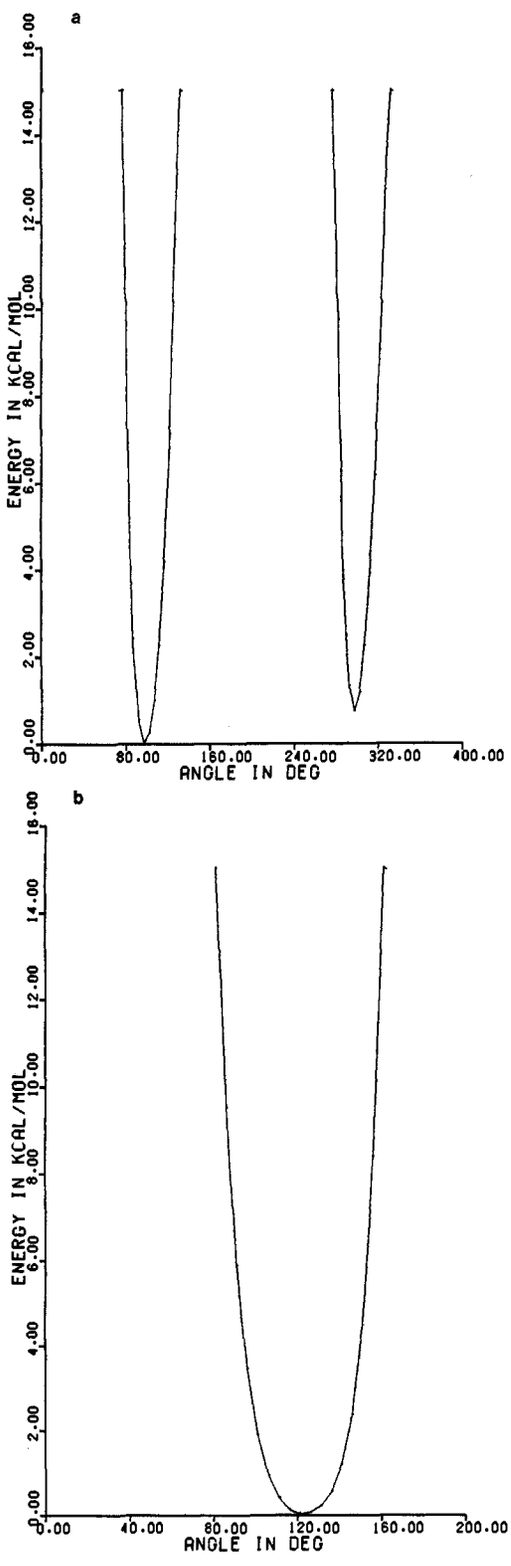



**Figure 9.** Triazine (III: $n = 1$) energy relative to the (probably) global minimum (a) varying $\omega_3$(C4–N5–C′1–C1′), (b) varying $\omega_4$(N5–C′1–C1′–C2′). All other dihedral angles are held fixed at the value of its global minimum-energy conformation.

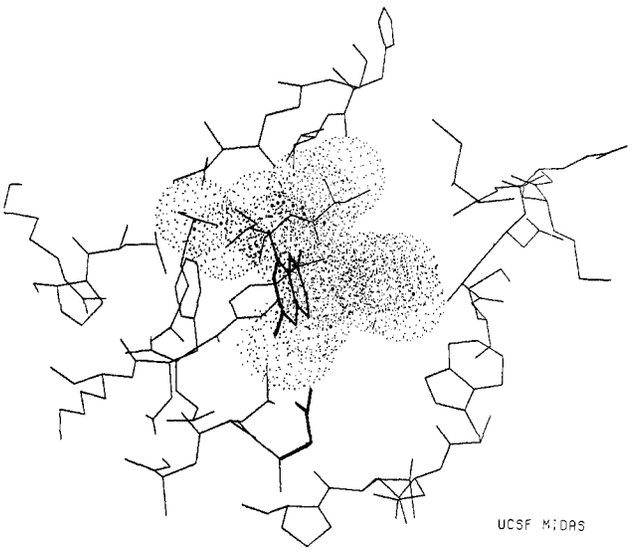



**Figure 10.** The present site model superimposed over the crystallographic site of the *E. coli* DHFR and methotrexate complex.[25] The thin lines are the active site residues of DHFR; the pteridine ring of methotrexate is seen edge on in heavy lines; site pockets appear as stippled spheres of a 0.7-Å radius. The amino acid side chains completely envelop the binding site but always leave room for the site pockets (apparent overlaps in this view are actually in front or behind the pockets). Site pockets 1 and 2 appear at the bottom of the picture near the glutamic acid side chain responsible for their hydrogen-bonding character. Pockets 8 and 10 extend far to the right, consitituting regions unused by methotrexate, but required by pyrimidines in our calculations. Hydrophobic amino acid side chains cluster around the hydrophobic pockets 13 and 14, as seen in the upper left wing of site spheres. Illustration is courtesy of the UCSF Computer Graphics Laboratory.[32]

ables present in any biological system are separated. The study of the receptor itself rather than the entire biological system eliminates a large number of variables from the problem. Even then the flexibility of the receptor and the ligand molecules and the rigid rotations and translations of the ligand relative to the receptor amounts to too many variables to solve the problem. Mapping of the receptor site is therefore possible only if some simplifying assumptions are made. Hansch's analysis[12] was the first fruitful approach to solve the problem, although it disregards the flexibility of the ligand molecules, rigid rotations or translations of the ligand molecule relative to the receptor site, and the quantitative spatial requirements for the biological activity. Such approximations were reasonable in this pioneering work and also unavoidable due to the cost of computer time during that period. The perturbation of the receptor itself by the ligand molecule may be another important aspect of biological activity. However, consideration of the conformational changes in the receptor by the influence of the ligand molecule is still a problem for medicinal chemists since they usually have little information about the specific structure of the receptor site.

It is of interest to compare our method with various other existing QSAR techniques. Our method is concep-

**Table X.** Molecular Structure and Observed[a] and Calculated[b] *E. coli* DHFR Inhibition Data[c] of the Compounds Used To Test the Predictive Power of the Model

| no. | groups | log $1/C_{obsd}$ | log $1/C_{calcd}$ | Δ(calcd−obsd) |
|---|---|---|---|---|
| | | Pyrimidines (IV) | | |
| 40 | 3′-I | 7.23 | 9.33 (15.09) | 2.10 (7.86) |
| 41 | 3′-$NO_2$, 4′-$NHCOCH_3$ | 6.97 | 7.76 (8.79) | 0.79 (1.82) |
| 42 | 4′-F | 6.35 | 6.93 (7.29) | 0.58 (0.94) |
| 43 | 3′-F | 6.23 | 5.55 (6.22) | −0.68 (−0.01) |
| 44 | 3′-$OCH_2C_6H_5$ | 6.99 | 7.64 (8.07) | 0.65 (1.08) |
| | | Triazines (V) | | |
| 45 | *n*-$C_3H_7$ | 5.33 | 4.89 (4.92) | −0.44 (−0.41) |
| 46 | *m*-$C_6H_4NO_2$ | 5.51 | 5.08 (5.20) | −0.43 (−0.31) |
| 47 | *p*-$C_6H_4CN$ | 4.04 | 3.59 (3.63) | −0.45 (−0.41) |
| 48 | *m*-$C_6H_4CH_2C_6H_4$ | 6.55 | 7.49 (7.73) | 0.94 (1.18) |
| 49 | *p*-$C_6H_4Cl$ | 6.40 | 6.89 (6.54) | 0.49 (0.14) |

[a] From ref 8 and 9. [b] Values within parentheses represents the calculated values if no alternative binding modes are kept in the model. [c] *C* represents the concentration of the inhibitor causing 50% inhibition.

**Table XI.** The Best Fitted Binding Modes of the Test Compounds[a]

| | site points | | | | | | | | | | | site points | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| no. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | no. | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 40 | H1 | H2 | H4 | C1′ | C3′ | C5′ | | A | D | | 40 | | | | | C7 | | | | |
| 41 | C1′ | C3′ | C5′ | A | | D | E | | | | 41 | F | | | | C7 | | | | |
| 42 | H1 | H2 | H4 | H[b] | | | | | | | 42 | | | | D | | C7 | | C6′ | |
| 43 | H1 | H2 | H4 | | | | | | | | 43 | | A | | | C7 | | | | |
| 44 | H1 | H2 | H4 | C1′ | C3′ | C5′ | F | | D | | 44 | A | | | C[c] | C7 | | | | |
| 45 | H1 | H2 | H4 | H[d] | | | | | | | 45 | | | | | C′1 | H[e] | | | |
| 46 | H1 | H2 | H4 | | | | | | | | 46 | | | | | C1′ | C3′ | C5′ | A | B |
| 47 | H1 | H2 | H4 | | | | | | | | 47 | | | | | C1′ | C3′ | C5′ | A | B |
| 48 | H1 | H2 | H4 | | | | | | | | 48 | | | C[f] | C[g] | C1′ | C3′ | C5′ | A | B |
| 49 | H1 | H2 | H4 | | | | | | | | 46 | | | | | C1′ | C3′ | C5′ | A | B |

[a] For atom symbols, see structure I for the pyrimidines (40–44) and structures II and III for the triazines (45–49). [b] One hydrogen of the methylene group. [c] Para carbon of the second benzene ring. [d] Hydrogen of the first methylene group. [e] Hydrogen of the terminal methyl group. [f] 3′-Carbon of the second phenyl ring. [g] 6′-Carbon of the second phenyl.

tually very similar to Hansch type analysis[12] with various added features. Those are as follows: (i) it precisely locates the ligand atoms in the three-dimensional space relative to the receptor site, allowing one to analyze ligand molecules of several different classes in a single model, if they interact at the same receptor site; (ii) it considers the alternate binding modes arising from the internal rotations around various rotatable bonds as well as from rigid rotations and translations, which often allows one to decide which part of the site is sterically blocked (in that sense it is a dynamic three-dimensional QSAR); (iii) the present approach may lead to a more refined mapping of the receptor site since the ligand molecule can be subdivided at the atomic level. Since our method can analyze ligand molecules of different classes in a single receptor model, it can give very novel suggestions for future synthesis. For example, it can design molecules which will fruitfully use all the site pockets which have been explored.

Hopfinger's molecular shape analysis[26] is very similar to Hansch's approach, with the addition of one more parameter related to the three-dimensional structure of the drug molecule, namely, the common overlap steric volume with an idealized molecule, the shape reference. He in fact quantitifed Marshall's[27] qualitative ideas of pharmacophore mapping. However, we think any single molecule may not represent the idealized structure, and also the entire structure of even the most active compound may not be equally important. In the present approach we subdivide the important part of the drug molecules both geometrically and parametrically, which allows us to explore the nature of the various regions of the receptor site rather explicitly. However, one major difference between the two methods is the way of considering the overlap. Hopfinger measured the exact overlap volume with the shape reference molecule, which corresponds to our site pockets. On the contrary we take the overlap as "all or none". Both approaches have at least one favorable and one unfavorable point. Molecular interaction is a continuous process, therefore, our "all or none" representation is something of an oversimplification of the actual state. However, the receptor site is not perfectly rigid, and the receptor atoms can always adjust slightly to accommodate the incoming ligand molecule. It is extremely difficult to evaluate the exact overlap under that situation, especially since: (i) we do not know the energetics of the changes within the receptor, and (ii) continuous search of the conformational space of the ligand molecule is computationally too lengthy. "All or none" representation is therefore a useful approximation. Physically this means that up to a certain limit, the ligand and receptor atoms will adjust themselves to get the maximum interaction. In Hopfinger's more recent work,[28] the shape descriptor has been redefined with use of a molecular mechanics potential field, which he finds to be an improvement over the earlier model.

The approach of Simon et al.[29,30] conceptually resembles ours to some extent, but they excluded the possibility of alternate binding modes. Furthermore, it suffers from the

(26) Hopfinger, A. J. *J. Am. Chem. Soc.* **1980**, *102*, 7196.
(27) Marshall, G. R.; Barry, C. D.; Bosshard, A. E.; Dammkoehler, R. A.; Dunn, D. A. In "Computer Assisted Drug Design"; Olson, E. C., Christoffersen, R. E., Eds; American Chemical Society: Washington, DC, 1979; ACS Symp. Series No. 112, p 205.
(28) Hopfinger, A. J. *J. Med. Chem.* **1983**, *26*, 990.
(29) Simon, Z.; Badileuscu, I.; Racovitan, T. *J. Theor. Biol.* **1977**, *66*, 485.
(30) Simon, Z.; Dragomir, N.; Plauchithiu, M. G.; Holban, S.; Glatt, H.; Kerek, F. *Eur. J. Med. Chem.* **1980**, *15*, 521.

parametric rigidity we find in Hopfinger's approach, since in the process of superposition of the drug molecule over the hypermolecule, it evaluates only one parameter (MTD). It is less sophisticated since it does not undertake the conformational features of the drug molecule or even its three-dimensional structure.

## Conclusion

The proportionality constants of the various physicochemical parameters (Table VIII) suggests that the protonated heterocyclic ring contributes approximately 4.62 in log $1/I_{50}$ units of activity. The compounds having activity lower than that suffer from soft repulsive interaction from the substituents with the receptor. Any positive value of the proportionality constants of hydrophobicity suggests that the pocket is hydrophobic. It is therefore obvious that site pockets of types 2 and 9 are hydrophobic in nature. Their relative magnitude suggests that in order to explain the biological activity of triazine (**39**) it is necessary to assume a site pocket which is nearly threefold[31] more hydrophobic than the pockets holding the phenyl ring of the pyrimidines.

When both the hydrophobic factor and dispersive forces are involved, the relative magnitude of the proportionality constants does not indicate anything about their relative importance unless the hydrophobic parameter and the molar refractivity of the concerned atom are comparable. For some atoms the magnitude of molar refractivity is nearly tenfold higher than its hydrophobic parameter, while in some other cases it is less than half of its hydrophobic parameter. However, using the proportionality constants (Table VIII), one can easily compile the interactions of various atoms with different types of site pockets and the relative importance of hydrophobic parameter and dispersive forces in each interaction. This information can be used for designing new drugs. Of course, during that step one should be aware of other factors, like atomic volume. Prediction may fail if the new atom and corresponding old atom differ greatly in volume. Valency is another factor to be given due consideration. For example, a monovalent atom may not be replaced by polyvalent atoms simply because we do not know the effects of these additional atoms. Proportionality constants derived from a few types of atoms may not hold for atoms which differ markedly in their physicochemical properties.

In our approach it is always advisable to construct large numbers of site pockets to position the molecule precisely at the receptor site and for the exact mapping of the receptor. However, the number of site types should be kept low to hold the number of adjustable parameters to a minimum. Even then the number of adjustable parameters, either visible or hidden, ought to be higher in three-dimensional QSAR than that in two-dimensional QSAR. The constrained solution poses a difficulty[3] in checking the statistical significance of the model. It is therefore advisable to check the merit of the model by allowing it to predict the biological activity of some molecules not included in the training set. If one wants to explore all sorts of binding modes arising from the free rotations and translations of the molecule relative to the receptor, the number of geometrically feasible binding modes soon goes beyond manageable limit with the increased number of site pockets. However, if the rotations and translations are restricted to a limited number, the number of site pockets may be increased greatly without any difficulty.

The hypothetical site pockets developed here agree quite well with the crystallographic receptor site in terms of their accessibility and nature of bonding. The greatest advantage of the present method is that X-ray data and three-dimensional molecular graphics can be used directly in this quantitative model building process. However, we must remember that drug design is an extremely difficult problem. The success of an equation of a model is not the ultimate answer. It needs a systematic collaboration between experimental and theoretical medicinal chemistry groups.

**Acknowledgment**. This work was supported by grants from the National Institutes of Health (1-R01-AM28140-03 and 5-R01-GM30561-02), the National Science Foundation (PCM-8314998), and the Robert A. Welch Foundation. Acknowledgment is also made to the donors of the Petrolium Research Fund, administered by the American Chemical Society, for partial support of this research. We thank Dr. Robert Langridge and co-workers for the use of the Computer Graphics Laboratory, University of California, San Francisco (NIH RR-1081).

**Registry No.** IV [3′,5′-$(CH_2OH)_2$], 77113-54-3; IV, 7319-45-1; IV (4′-$NO_2$), 69945-52-4; IV (3′-$CH_2OH$), 77113-56-5; IV (4′-$NH_2$), 69945-50-2; IV (4′-Cl), 18588-43-7; IV [3′,4′-$(OH)_2$], 71525-05-8; IV (3′-OH), 77113-55-4; IV (4′-$CH_3$), 46726-70-9; IV (4′-$OCF_3$), 49561-94-6; IV (3′-$CH_2OCH_3$), 77113-57-6; IV (3′-Cl), 69945-58-0; IV (3′-$CH_3$), 69945-56-8; IV [4′-$N(CH_3)_2$], 69945-51-3; IV (4′-$OCH_3$), 20285-70-5; IV (4′-Br), 69945-55-7; IV (4′-$NHCOCH_3$), 69945-53-5; IV (3′-$OSO_2CH_3$), 77113-58-7; IV (3′-$OCH_3$), 59481-28-6; IV (3′-Br), 69945-59-1; IV (3′-$CF_3$), 50823-94-4; IV (3′-$CF_3$, 4′-$OCH_3$), 50823-96-6; IV [3′,4′-$(OCH_3)_2$], 5355-16-8; IV [3′,5′-$(OCH_3)_2$], 20344-69-8; IV [3′,4′,5′-$(OCH_3)_3$], 738-70-5; IV (3′-I), 30077-60-2; IV (3′-$NO_2$, 4′-$NHCOCH_3$), 69945-54-6; IV (4′-F), 836-06-6; IV (3′-F), 69945-57-9; IV (3′-$OCH_2C_6H_5$), 69945-60-4; V (R = $C_6H_4$-4′-COOH), 14484-50-5; V (R = $C_6H_4$-3′-COOH), 17740-28-2; V (R = $CH_3$), 3977-24-0; V (R = $C_6H_4$-4′-$COOC_2H_5$, 17740-29-3; V (R = $CH_2C_6H_5$), 94295-02-0; V (R = $C_6H_4$-4′-$C_6H_5$), 4653-78-5; V (R = Pr), 4038-62-4; V (R = $C_6H_4$-3′-$OCH_3$), 17711-73-8; V [R = $(CH_2^-)_2C_6H_5$], 94295-03-1; V [R = $C_6H_5$], 4022-58-6; V (R = $C_6H_4$-3′-$CF_3$), 1492-81-5; V (R = $C_6H_4$-3′-Cl), 13351-02-5; V (R = $C_6H_3$-3′,4′-$Cl_2$), 13344-99-5; V [R = $(CH_2)_3C_6H_5$), 640-02-8; V (R = m-$C_6H_4NO_2$), 17711-74-9; V (R = p-$C_6H_4CN$), 17711-68-1; V (R = m-$C_6H_4CH_2C_6H_4$), 4038-61-3; V (R = p-$C_6H_4Cl$), 516-21-2; DHFR, 9002-03-3.

(31) To represent the phenyl ring in the triazine (**39**) we used only two points, but for that in the pyrimidines we used three points. If we take this into account, site type 9 becomes threefold more hydrophobic and not fourfold, as the entries in Table VIII would indicate.

(32) Huang, C.; Gallo, L.; Ferrin, T.; Langridge, R. L. "MIDAS, Molecular Interactive Display and Simulation, Users Guide"; Computer Graphics Laboratory, University of California, San Francisco, 1983.