probe, is a superior replacement for V_0 in this study must be viewed with reserved mechanistic interpretation.

However, in conjunction with the QSAR given by eq 9 it has been possible to generate the following satellite information on the benzylpyrimidines, which can be used in a design mode: (1) the active conformation with respect to θ_1 and θ_2; (2) identification of preferred binding to a site having a net positive charge; and (3) realization that a H^+ test probe yields relative intermolecular energetics that strongly correlate with biological activity. The "trick" of replacing $\Delta P_{\alpha,\beta}$ with $\Delta P_{\alpha,\beta}(C, 2)$ is critical to the practical and reliable usage of MSA based upon molecular potential energy fields. The calculation of potential energy fields in the 30- to 40-Å range is quite time consuming and of limited physical meaning. The form of the potentials at such distances are unknown, and, again, dielectric effects are expected to alter (decrease) field strength. The $\Delta P_{\alpha,\beta}(C, 2)$ for C in the range just beyond intermolecular steric contacts, C = 8 to 15 Å for the benzylpyrimidines, may, in fact, be the descriptors reflecting actual binding. These distances correspond to direct engagement with the receptor site. In any event, evaluating the general applicability of $\Delta P_{\alpha,\beta}(C, 2)$ as the comparative shape descriptor in future studies is a high priority in our work, as is understanding what it measures.

Acknowledgment. The author acknowledges the helpful discussions with R. A. Pearlstein and D. Malhotra of Case Western Reserve University during the course of this study.

Registry No. 1, 71525-05-8; 2, 69945-50-2; 3, 69945-51-3; 4, 46726-70-9; 5, 20285-70-5; 6, 49561-94-6; 7, 59481-28-6; 8, 69945-52-4; 9, 69945-53-5; 10, 18588-43-7; 11, 738-70-5; 12, 5355-16-8; 13, 69945-54-6; 14, 69945-55-7; 15, 836-06-6; 16, 7319-45-1; 17, 69945-56-8; 18, 69945-57-9; 19, 69945-58-0; 20, 50823-94-4; 21, 69945-59-1; 22, 50823-96-6; 23, 69945-60-4; DHFR, 9002-03-3.

Combined Distance Geometry Analysis of Dihydrofolate Reductase Inhibition by Quinazolines and Triazines

Arup K. Ghose and Gordon M. Crippen*

Department of Chemistry, Texas A&M University, College Station, Texas 77843. *Received July 30, 1982*

Guided by the success of distance geometry in explaining the inhibition of dihydrofolate reductase by 68 quinazolines, we have made a combined analysis on the inhibition of rat liver dihydrofolate reductase by 33 triazines and 15 quinazolines. The model gave a fit having the correlation coefficient 0.892 and root mean square (rms) deviation 0.596 in log $(1/C_{50})$ units. The model was applied to predict the biological activity of 91 compounds. The predicted values showed an rms deviation of 0.907 and a correlation coefficient of 0.790. The present study suggested the synthesis of some triazines as possible potent dihydrofolate inhibitors. The site geometry was compared with the crystal structure of a triazine bound to chicken liver dihydrofolate reductase, and a good correlation has been found.

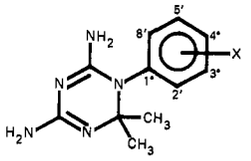
The medicinal importance of dihydrofolate reductase has stimulated a great deal of work in several fields, and a large number of molecules have been evaluated[1] as its inhibitors; the species specificity[2] of the enzyme inhibitors has been examined; crystallographers have not only studied the structures of the inhibitors[3] but also the structure of inhibitor–enzyme complexes.[4,5] Such extensive data have attracted quantitative structure–activity relationship (QSAR) workers. Hansch et al.[6-8] made an extensive analysis using their physicochemical parameter dependent QSAR. On the other hand, the three-dimensional structure directed QSAR has been investigated by Crippen et al.,[9-11] Hopfinger et al.[12,13] and Simon et al.[14]

Guided by the success[10,11] of the distance geometry approach with even as few as six site points to explain the inhibition of *Streptococcus faecium* dihydrofolate reductase by 68 quinazolines, we have undertaken in this work a combined QSAR analysis of the triazines and quinazolines. Although in our previous studies we took the inhibition data of 68 quinazolines against dihydrofolate reductase from *S. faecium*, in the present study we selected the dihydrofolate reductase from rat liver, since a large variety of inhibitors have been investigated for this mammalian enzyme. We took all 33 3′- and 4′-substituted triazines having the general structure I (shown in Table I), reported by Dietrich et al.[15] and Hansch et al.[16] In order to keep the data set to a reasonable size and hold

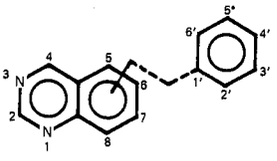
(1) B. R. Baker, *Ann. N.Y. Acad. Sci.*, **186**, 214 (1971).
(2) B. R. Baker, "Design of Active Site Directed Irreversible Enzyme Inhibitors", Wiley, New York, 1967, pp 192–266.
(3) A. Camerman, H. W. Smith, and N. Camerman, *Acta Crystallogr., Sect. B*, **35**, 2113 (1979).
(4) D. A. Matthews, R. A. Alden, J. T. Bolin, S. T. Freer, R. Hamlin, N. Xuong, J. Kraut, M. P. M. Williams, and K. Hoogsteen, *Science*, **197**, 452 (1977).
(5) D. A. Matthews, R. A. Alden, S. T. Freer, N. Xuong, and J. Kraut, *J. Biol. Chem.*, **254**, 4144 (1979).
(6) S. W. Dietrich, R. N. Smith, S. Brendler, and C. Hansch, *Arch. Biochem. Biophys.*, **194**, 612 (1979).
(7) S. W. Dietrich, J. M. Blaney, M. A. Reynolds, P. Y. C. Jow, and C. Hansch, *J. Med. Chem.*, **23**, 1205 (1980).
(8) C. Hansch, J. Y. Fukunaga, P. Y. C. Jow, and J. B. Hynes, *J. Med. Chem.*, **20**, 96 (1977).
(9) G. M. Crippen, *J. Med. Chem.*, **22**, 988 (1979).
(10) A. K. Ghose and G. M. Crippen, *J. Med. Chem.*, **25**, 892 (1982).
(11) A. K. Ghose and G. M. Crippen, "Proceedings of the 4th European Symposium on Chemical Structure and Biological Activity: Quantitative Approaches", Elsevier, Amsterdam, 1982, in press.
(12) A. J. Hopfinger, *J. Am. Chem. Soc.*, **102**, 7196 (1980).
(13) C. Battershell, D. Malhotra, and A. J. Hopfinger, *J. Med. Chem.*, **24**, 812 (1981).
(14) Z. Simon, I. Badilescu, and T. Racovital, *J. Theor. Biol.*, **66**, 485 (1977).
(15) S. W. Dietrich, R. N. Smith, J. Y. Fukunaga, M. Olney, and C. Hansch, *Arch. Biochem. Biophys.*, **194**, 600 (1979).
(16) C. Hansch, S. W. Dietrich, and J. Y. Fukunaga, *J. Med. Chem.*, **24**, 544 (1981).

0022-2623/83/1826-0996$01.50/0 © 1983 American Chemical Society

Table I. The Molecular Structure and the Observed and Calculated Inhibition Data of the Various Triazenes and Quinazolines Studied



I



II

no.	substituent	$\log (1/C)_{obsd}$	$\log (1/C)_{calcd}$	$\Delta_{obsd-calcd}$
	Triazines (I)			
1	4'-CO_2CH_3	4.26	5.76	−1.50
2	3'-SO_2NH_2	4.41	6.30	−1.89
3	4'-$CO_2C_2H_5$	4.44	5.76	−1.32
4	4'-SO_2NH_2	4.54	5.32	−0.78
5	4'-$COCH_3$	5.25	5.76	−0.51
6	3'-$COCH_3$	5.35	5.68	−0.33
7	3'-$CO_2C_2H_5$	5.69	5.64	−0.95
8	H	5.99	5.32	0.67
9	4'-CF_3	6.06	6.36	−0.30
10	4'-Br	6.24	5.32	0.92
11	3'-OCH_3	6.26	6.02	0.24
12	4'-OCH_3	6.26	5.76	0.50
13	4'-I	6.28	5.32	0.96
14	4'-$C(CH_3)_3$	6.40	5.76	0.64
15	4'-CH_3	6.41	5.76	0.65
16	3'-F	6.42	6.42	0.00
17	4'-F	6.67	6.36	0.31
18	3'-CN	6.68	6.77	−0.14
19	3'-[$O(CH_2)_4O(C_6H_4$-3-CF_3)]	6.79	7.36	−0.57
20	3'-CH_3	6.81	6.29	0.52
21	3'-[$OCH_2(C_6H_3$-3,4-Cl_2)]	6.82	7.54	−0.72
22	3'-$C(CH_3)_3$	6.89	6.64	0.25
23	3'-Cl	6.93	6.70	0.23
24	3'-$OCH_2C_6H_5$	6.94	6.98	−0.04
25	3'-$O(CH_2)_3CH_3$	7.01	6.03	0.98
26	3'-Br	7.06	7.06	0.00
27	3'-I	7.09	7.09	0.00
28	3'-CF_3	7.10	6.99	0.11
29	4'-$(CH_2)_3CH_3$	7.14	6.73	0.41
30	3'-$O(CH_2)_2OC_6H_5$	7.18	7.33	−0.15
31	3'-$O(CH_2)_4OC_6H_5$	7.20	7.33	−0.13
32	4'-[$OCH_2(C_6H_3$-3,4-Cl_2)]	7.22	7.47	−0.25
33	4'-$OCH_2C_6H_5$	7.27	7.06	0.22
	Quinazolines (II)			
34	2,4-H_2	2.26	2.07	0.19
35	2-SH, 4-NH_2	3.72	3.66	0.06
36	2,4-$(OH)_2$	3.89	3.80	0.09
37	2,4-$(NH_2)_2$, 5-[$SO_2(C_6H_3$-3,4-Cl_2)]	4.24	5.05	0.81
38	2,4,6-$(NH_2)_3$	4.57	4.78	−0.22
39	2,4-$(NH_2)_2$	4.66	4.78	−0.13
40	2,4-$(NH_2)_2$, 6-CN	4.92	4.78	0.14
41	2-OH, 4-NH_2, 6-[$S(2$-$C_{10}H_7$)]	5.77	5.77	0.00
42	2,4-$(NH_2)_2$, 5-Cl, 6-[$NHCO(CH_2)_2(C_6H_4$-4-Cl)]	7.44	7.47	0.03
43	2,4-$(NH_2)_2$, 6-[$NHCOCH_2(C_6H_3$-3,4-Cl_2)]	7.51	7.47	0.04
44	2-NH_2, 4-SH, 6-[$SO_2(2$-$C_{10}H_7$)]	7.60	7.66	−0.06
45	2,4-$(NH_2)_2$, 5-CH_3, 6-[$NHCOCH_2(C_6H_3$-3,4-Cl_2)]	7.77	6.93	0.84
46	2-NH_2, 4-OH, 6-[$SO_2(2$-$C_{10}H_7$)]	7.96	7.80	0.16
47	2,4-$(NH_2)_2$, 6-[$SO_2(C_6H_3$-3,4-Cl_2)]	8.05	8.17	−0.12
48	2,4-$(NH_2)_2$, 6-[$SO_2(2$-$C_{10}H_7$)]	8.40	8.37	0.03

down computing costs, we selected only 15 out of 104 quinazolines from the data reported by Fukunaga et al.[17] Our selection attempted to cover the maximum variety of structures and maximum range of inhibition. Another reason for selecting only a few of the quinazolines was that if a site model and interaction matrix having a good fit of the biological data could be developed, the rest of the molecules could be used to test the predictive power of the method.

Method

The method of calculation used in this study was similar to the steps reported earlier.[10] The only difference was in the optimization step as reported in ref 11.

Results and Discussion

Table I displays the structures of the various triazines and quinazolines, together with the observed and calculated inhibition data. It can be shown [7,18] that for a rapid equilibrium bireactant system, Figure 1, where the in-

(17) J. Y. Fukunaga, C. Hansch, and E. E. Steller, *J. Med. Chem.*, 19, 605 (1976).

(18) I. H. Segel, "Enzyme Kinetics", Wiley, New York, 1975, pp 273–291.

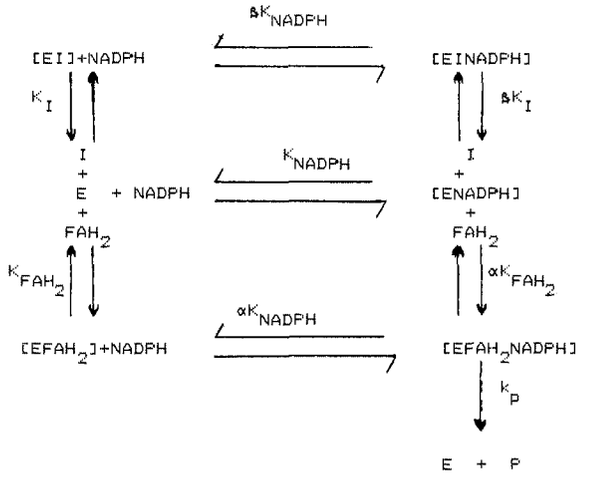


Figure 1. A schematic representation of the NADPH-dependent reduction of dihydrofolate (FAH_2) to tetrahydrofolate, assuming the inhibitor–enzyme cofactor complex is catalytically inactive and the cofactor is saturating.

hibitor–enzyme cofactor complex is catalytically inactive and the cofactor is saturating, as is true for NADPH-dependent reduction of dihydrofolate to tetrahydrofolate by dihydrofolate reductase,

$$V_I/V_0 = K_I\,(\text{app})/(K_I\,(\text{app}) + [I]) \qquad (1)$$

where V_I and V_0 are the velocities of reaction in the presence and absence of inhibitor respectively, and

$$K_I\,(\text{app}) = \beta K_I[1 + ([FAH_2]/\alpha K_{FAH_2})] \qquad (2)$$

From eq 1 it is obvious that the concentration of the inhibitor for 50% inhibition, $[I_{50}]$, is K_I (app). However, for a particular dihydrofolate concentration, K_I (app) will be a true measure of the relative K_I values of the various inhibitors only if β is constant for all inhibitors, which may or may not be true. In the present study, we simply used the $-\log I_{50}$ values without attempting to convert them to free energies of binding because we have no data on the β's, the binding of the inhibitors to the NADPH–enzyme complex is physiologically more relevant than binding to the free enzyme, and comparison with other QSAR studies is facilitated, since they also used $\log I_{50}$.

The molecules were constructed from the crystallographic data[3,10] on their constituent fragments. The upper and lower distance limits between all atom pairs in the triazines were evaluated by rotating the N_5–C'_1 (III) torsion angle and the various torsion angles in the substituents attached to the phenyl ring. The limits for the quinazolines were deduced by rotating the C_5–X_5 torsion angle and the other torsion angles between C_6 and C'_1 (IV). In all subsequent computations, the hydrogen atoms and some other atoms that were structurally equivalent to another atom were deleted from the structure in order to keep the calculation to a reasonable size.

The initial set of given binding modes originated from our earlier studies on the inhibition of *S. faecium* dihydroflate reductase by 68 quinazolines and from the similarities in the structures of the various quinazolines and triazines. Furthermore, we used only the site points corresponding to the strongly bound 2,4-diaminoquinazolines, together with a few others to bind the triazines' substituents. The correspondence between the present site points and the earlier site points is shown in Table II. The protonation of a ring nitrogen was assumed for all the triazines, as well as for the 2,4-diaminoquinazolines, except those having SO or SO_2 substituents at the 5-position. In the earlier studies[11] it was concluded

Table II. Site Point Description

site	correspondence with previous site points[a]	binds
1	2	N_1 of the quinazoline or triazine ring
2	4	the amino or other substituents at the 2-position of the quinazoline and triazine rings
3	6	the same as site 2 for the substituents of position 4
4	8	the 6'-C of the phenyl group in triazines or the 5-substituents of the quinazolines
5	9	4'-substituents in the triazines or the C'_6 of the quinazolines
6	10	4'-substituents of the quinazolines or the long substituents of the triazines
7	11	the 3'-position of the triazines or the 6-substituents of the quinazolines
8		the 2nd atom of the 3'-substituents of the triazines
9		the 1st atom of the 3'-substituents of the triazines

[a] Reference 11.

Table III. Coordinates of the Site Points[a]

site point	x	y	z
1	−4.419	−0.804	0.329
2	−6.114	0.877	0.641
3	−1.908	2.270	−1.160
4	−0.628	−1.629	−2.040
5	2.968	−0.239	−1.279
6	5.469	2.723	−0.402
7	1.203	−1.574	0.182
8	2.060	−0.267	1.938
9	1.368	−1.357	1.790

[a] In angstroms.

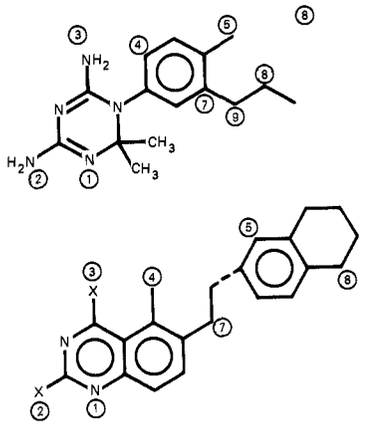
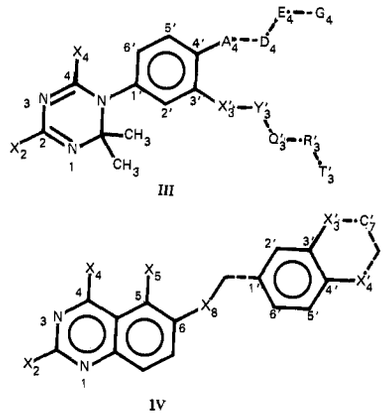


Figure 2. A two-dimensional representation of the site points indicating the hypothesized binding of the triazine and quinazoline molecules.

that the very low activity of the 5-sulfinyl- or 5-sulfonyl-substituted 2,4-diaminoquinazolines might be due to the electron-withdrawing effect of these groups from the ring nitrogens rather than due to steric requirements at the 5-position. Consequently, for this work, these quinazolines and the other quinazolines without 2,4-diamino substitution were assumed to have the same binding mode as that of 2,4-diaminoquinazolines but with a different interaction between site point 1 and the ring nitrogen (N_1).

The site-point geometry is given by the coordinates of the 9 site points in Table III; a schematic representation

Chart I[a]



III

IV

[a] For structure III: Although the numbering used here does not correspond to the official numbering, it is used to make the similarities in the structures of triazine and quinazoline more obvious.

of the site points relative to the triazines and quinazoline molecules is given in Figure 2. The site geometry can accomodate the input binding modes if a site flexibility[10] of 0.53 Å is assumed.

In the initially supplied binding modes and the best fitted binding modes of the triazines, the triazine ring and its 1-phenyl ring always maintain the same binding mode (see Table IV). The binding modes differ only in attachments of the various substituents with the appropriate site points. The one exception is molecule **19**, which shows an inverted, coiled conformation when bound to the receptor. However the physical significance of this is not very obvious. Consideration of the binding modes of substituents suggests that site point 5 is possibly a hydrophobic pocket with the ability to bind fluorine atoms. This site point does not bind the other halogen substituents at the 4′-position for purely geometric reasons. A much higher site flexibility would be necessary if these halogen atoms are assumed to bind with this site point. Site point 6 was originally assumed for X'_4 substituents of the quinazolines (IV, Chart I) and, indeed, this site point remained seldom used by the triazines. The exceptions are the chlorine at the para position of the second benzene ring in molecule **21** and the meta and para carbons of the second benzene ring in molecules **30** and **33**, respectively.

Site points 9 and 8 were assumed to bind the first and second atom of the 3′-substituents. If such atoms are available, they consistently bind there, except for the very long substituents where the second aromatic ring can occupy sites 5 and 7. A different binding mode is observed for these molecules, e.g., **21**, 24, **31**, and **32**.

The quinazolines, on the other hand, also show some consistency in the binding of the quinazoline ring and the substituents attached directly to it. The consistency, however, fails for the first few molecules having very few substituents. These molecules undergo some sort of rotation or translation in order to bind more favorably. Although the pyrimidine part of the quinazoline ring and the substituents directly attached to this ring maintain the binding mode, the benzene part shows some changes. In many systems the C_6 binds with site 4. At first glance it seems impossible, but it is reasonable if we remember that the site points are not perfectly rigid, thereby allowing an inplane counterclockwise rotation of the molecule. Such movement also often allows the X'_4 or X'_3 substituents to bind at site 9.

Since the site points are generated on the basis of the molecular structures in the data set, some deformation in the site geometry is expected as molecules having different structures are added to the data set. In order to compare the present structure with that of our previous studies,[10,11] the two sets of intersite point distances are presented in Table V. The comparison clearly shows that although site points 1–3 and 7 have almost the same relative arrangements, site points 4–6 suffered some deformations. This deformation is a consequence of the steric requirement that the phenyl and the triazine rings are not coplanar.

The apparent interaction matrix, Table VI, shows that the interaction of the protonated ring nitrogen with site point 1 is very strong, as we concluded in our previous studies.[10,11] Any comparison of the present interaction matrix with the previous one[11] should be made only qualitatively. Since even if we assume that β in eq 2 is constant for all the inhibitors and, therefore, that K_I (app) is a true measure of the relative stability of the enzyme–inhibitor complex, the apparent binding energy $\Delta G' = -RT \ln K_I$ (app) and not simply $-\ln K_I$ (app) as is used in the present study. In our previous studies, a strong interaction of the saturated carbon with site 7 (previously site 11) was observed, but the present study does not use this interaction, because there is no example of such a 6-substituted quinazoline in this data set (Table I). The moderately high interaction of sp^2 carbon with site 6 is in accord with the previous studies.[10] Comparing the relative interaction of a site point with various molecular points, we see that site point 1 seems to be a very good hydrogen-bond acceptor but a rather poor electron-pair acceptor. The interactions of site point 2 seem to arise from hydrogen bonding. In fact, the higher interaction with the amino group compared to that with hydroxyl group suggests that this site point may be the hydrogen donor.[19] On the other hand, the interaction of site point 3 seems to arise from its hydrogen-bond acceptance from the X_4 group, although the high interaction with the SH group suggests that some van der Waals type interaction may also be involved. Site point 4 has little contribution toward the total binding energies of the molecules. The comparative interactions of site point 5 with saturated carbon, unsaturated carbon, and fluorine are somewhat confusing. The moderately high interaction of site point 6 with chlorine and unsaturated carbon may arise from van der Waals interactions. On the other hand, some sort of dipolar interaction seems to be responsible for the interactions with site points 7 and 9. Of course, one should be wary of overinterpreting the entries in Table VI, since they are simply empirically adjusted parameters in an idealized model of the binding site.

The interaction matrix suggests some triazines that may be promising as potent dihydrofolate reductase inhibitors. Referring to structure III, these molecules correspond to $X_2 = NH_2$; $X_4 =$ OH, SH, and NH_2; $X'_3 =$ halogen; and $A_4 =$ long chain with halogen-substituted aromatic ring. Unlike the quinazolines, 2,4-diamino substitution may not be essential for good binding of a triazine. In the nonaromatic triazine system, one NH_2 and one OH or one SH may be sufficient to increase the basicity of N_1 so that it becomes protonated in solution. The purpose of the 3′-halo is to have the interaction with site 9, and the long chain with the halo-substituted aryl group is to be involved with site 6. The interest in triazines began with the isolation of 2,4-diamino-1-(4′-chlorophenyl)triazine as the active metabolite of the antimalarial drug proguanil.[20,21] In the

(19) L. N. Ferguson, "The Modern Structural Theory of Organic Chemistry", Prentice Hall, Englewood Cliffs, NJ, 1963, p 129.
(20) H. C. Carrington, A. F. Crowther, D. G. Davey, A. A. Levi, and F. L. Rose, *Nature (London)*, 168, 1080 (1951).
(21) A. F. Crowther and A. A. Levi, *Br. J. Pharmacol.*, 8, 93 (1953).

Table IV. The Initially Supplied and the Best Fitted Binding Modes[a]

no.	sup/obsd (I/II)	site points 1	2	3	4	5	6	7	8	9
1	I	N_1	X_2	X_4	C'_6	A_4		C'_3		
	II	N_1	X_2	X_4	C'_6	E_4		C'_3		
2	I	N_1	X_2	X_4	C'_6			C'_3		
	II	N_1	X_2	X_4				X'_3		Y'_3 (O)
3	I	N_1	X_2	X_4	C'_6	A_4		C'_3	G_4	
	II	N_1	X_2	X_4	C'_6	E_4		C'_3		
4	I	N_1	X_2	X_4	C'_6	A_4		C'_3		
	II	N_1	X_2	X_4	C'_6			C'_3		
5	I	N_1	X_2	X_4	C'_6	A_4		C'_3		
	II	N_1	X_2	X_4	C'_6	D_4		C'_3		
6	I	N_1	X_2	X_4	C'_6			C'_3	Y'_3	X'_3
	II	N_1	X_2	X_4		Y'_3		C'_3		
7	I	N_1	X_2	X_4	C'_6			C'_3	Y'_3	X'_3
	II	N_1	X_2	X_4	C'_6			C'_3	Q'_3	R'_3
8	I	N_1	X_2	X_4	C'_6			C'_3		
	II	N_1	X_2	X_4	C'_6			C'_3		
9	I	N_1	X_2	X_4	C'_6	A_4		C'_3		
	II	N_1	X_2	X_4	C'_6	D_4		C'_3		
10	I	N_1	X_2	X_4	C'_6			C'_3		
	II	N_1	X_2	X_4	C'_6			C'_3		
11	I	N_1	X_2	X_4	C'_6			C'_3	Y'_3	X'_3
	II	N_1	X_2	X_4	C'_6			C'_3	Y'_3	X'_3
12	I	N_1	X_2	X_4	C'_6	A_4		C'_3		
	II	N_1	X_2	X_4	C'_6	D_4		C'_3		
13	I	N_1	X_2	X_4	C'_6			C'_3		
	II	N_1	X_2	X_4	C'_6			C'_3		
14	I	N_1	X_2	X_4	C'_6	A_4		C'_3		
	II	N_1	X_2	X_4	C'_6	A_4		C'_3		
15	I	N_1	X_2	X_4	C'_6	A_4		C'_3		
	II	N_1	X_2	X_4	C'_6	A_4		C'_3		
16	I	N_1	X_2	X_4	C'_6			C'_3		X'_3
	II	N_1	X_2	X_4	C'_6			C'_3		X'_3
17	I	N_1	X_2	X_4	C'_6	A_4		C'_3		
	II	N_1	X_2	X_4	C'_6	A_4		C'_3		
18	I	N_1	X_2	X_4	C'_6			C'_3		
	II	N_1	X_2	X_4	C'_6			C'_3		X'_3
19[b]	I	N_1	X_2	X_4	C'_6		F	C'_3	Y'_3	X'_3
	II	N_1	N_5	T'_3					F	C
20	I	N_1	X_2	X_4	C'_6			C'_3		X'_3
	II	N_1	X_2	X_4	C'_6			C'_3		X'_3
21[c]	I	N_1	X_2	X_4	C'_6			C'_3	Y'_3	X'_3
	II	N_1	X_2	X_4	C'_6	C (o)	Cl (p)	C'_3		X'_3
22	I	N_1	X_2	X_4	C'_6			C'_3	Y'_3	X'_3
	II	N_1	X_2	X_4	C'_6			C'_3	Y'_3	X'_3
23	I	N_1	X_2	X_4	C'_6			C'_3		X'_3
	II	N_1	X_2	X_4	C'_6			C'_3		X'_3
24[d]	I	N_1	X_2	X_4	C'_6			C'_3	Y'_3	X'_3
	II	N_1	X_2	X_4		C (m)		C(1)		Y'_3
25	I	N_1	X_2	X_4	C'_6			C'_3	Y'_3	X'_3
	II	N_1	X_2	X_4					Y'_3	Q3
26	I	N_1	X_2	X_4	C'_6			C'_3		X'_3
	II	N_1	X_2	X_4	C'_6			C'_3		X'_3
27	I	N_1	X_2	X_4	C'_6			C'_3		X'_3
	II	N_1	X_2	X_4	C'_6			C'_3		X'_3
28[e]	I	N_1	X_2	X_4	C'_6			C'_3	F	X'_3
	II	N_1	X_2	X_4	C'_6			C'_3	F	F
29	I	N_1	X_2	X_4	C'_6	A_4		C'_3		
	II	N_1	X_2	X_4	C'_6	A_4		C'_3		G4
30[d]	I	N_1	X_2	X_4	C'_6		C (m)	C'_3	Y'_3	X'_3
	II	N_1	X_2	X_4			C (m)		Q'_3	Y'_3
31[d]	I	N_1	X_2	X_4	C'_6		C (m)	C'_3	Y'_3	X'_3
	II	N_1	X_2	X_4		C (m)		C (o)	Q'_3	R'_3
32[d]	I	N_1	X_2	X_4	C'_6	A_4	C (m)	C'_3		
	II	N_1	X_2	X_4	C'_6	C(1)		C (m)		Cl (m)
33	I	N_1	X_2	X_4	C'_6	A_4	C (p)	C'_3		
	II	N_1	X_2	X_4	C'_6	D_4	C (p)	C'_3		
34	I	N_1								
	II					C_5	C_2			
35	I	N_1	X_2	X_4						
	II			X_2		C_5		C_8		
36	I	N_1	X_2	X_4						
	II			X_2		C_5		C_8		
37	I	N_1	X_2	X_4						
	II	N_3		$O(SO_2)$	C_2			C'_3		X'_4

Table IV (*Continued*)

no.	sup/obsd (I/II)	site points 1	2	3	4	5	6	7	8	9
38	I	N_1	X_2	X_4				X_6		
	II	N_1	X_2	X_4	C_6					
39	I	N_1	X_2	X_4						
	II	N_1	X_2	X_4	C_6					
40	I	N_1	X_2	X_4				X_6		
	II	N_1	X_2	X_4	C_6					
41	I	N_1	X_2	X_4		C'_2		X_6		
	II	N_1	X_2	X_4		C'_2	X'_4	X'_6		
42	I	N_1	X_2	X_4		C'_2	X'_4	X_6		
	II	N_1	X_2	X_4	C_6	C'_2		C'_4		X'_4
43	I	N_1	X_2	X_4		C'_2	X'_4	X_6		
	II	N_1	X_2	X_4	C_6	C'_2		C'_4		X'_4
44	I	N_1	X_2	X_4		C'_2	X'_4	X_6		
	II	N_1	X_2	X_4		C'_2	X'_4	X_6		$O(SO_2)$
45	I	N_1	X_2	X_4			X'_4	X_6		
	II	N_1	X_2	X_4	C_6	C'_1				X'_3
46	I	N_1	X_2	X_4		C'_2	X'_4	X_6		
	II	N_1	X_2	X_4		C'_2	X'_4	X_6		$O(SO_2)$
47	I	N_1	X_2	X_4		C'_6	X'_4	X_6		
	II	N_1	X_2	X_4		C'_6	X'_4	X_6		$O(SO_2)$
48	I	N_1	X_2	X_4		C'_2	X'_4	X_6		
	II	N_1	X_2	X_4		C'_2	X'_4	X_6		$O(SO_2)$

[a] Atom labels of the triazines correspond to structure III and those of the quinazolines to structure IV. [b] F and C are the fluorine and carbon atoms in CF_3. [c] C(o) and Cl(p) are the ortho carbon and *p*-chloro substituent of the second phenyl ring. [d] C(m) and C(1) are the meta and first carbon of the second phenyl ring respectively. [e] F is the fluorine atom in CF_3.

Table V. Comparison of Site Distances[a]

site	1	2	3	4	5	6	7	8	9
1		2.4 (2.2)	4.2 (4.0)	4.5 (4.9)	7.6 (6.3)	10.5 (9.9)	5.7 (5.7)	6.7	6.0
2	2.4 (2.2)		4.8 (4.7)	6.6 (6.4)	9.3 (8.1)	11.8 (11.5)	7.7 (7.8)	8.4	7.9
3	4.2 (4.0)	4.8 (4.7)		4.2 (2.4)	5.5 (5.1)	7.4 (8.0)	5.1 (5.2)	5.6	5.7
4	4.5 (4.9)	6.6 (6.4)	4.2 (2.4)		3.9 (3.1)	7.7 (6.1)	2.9 (3.2)	5.0	4.3
5	7.6 (6.3)	9.3 (8.1)	5.5 (5.1)	3.9 (3.1)		4.0 (3.7)	2.7 (2.6)	3.3	3.6
6	10.5 (9.9)	11.8 (11.5)	7.4 (8.0)	7.7 (6.1)	4.0 (3.7)		6.1 (6.0)	5.1	6.2
7	5.7 (5.7)	7.7 (7.8)	5.1 (5.2)	2.9 (3.2)	2.7 (2.6)	6.1 (6.0)		2.4	1.6
8	6.7	8.4	5.6	5.0	3.3	5.1	2.4		1.3
9	6.0	7.9	5.7	4.3	3.6	6.2	1.6	1.3	

[a] All entries are in angstroms. The values in parentheses are the distances in the previous study (ref 10 and 11).

Table VI. Apparent Interaction Energy of the Dihydrofolate Reductase Site Points with the Various Molecular Points

ligand point type	site points 1	2	3	4	5	6	7	8	9
1. C (sp^3)				0.001[b]	0.437		0.001[b]	0.349	0.969
2. O		0.648	2.498						0.349
3. N (basic amino)		1.074	1.386						
4. S (S or SH)			2.354	0.001[b]			1.097		
5. Cl				0.001[b]		1.098	0.001[b]		1.376
6. F					1.044			0.567	1.009
7. C (sp^2)				0.075	0.770	1.298	0.535		
8. N (double bonded)	0.570								
9. N (protonated)	2.250								
10. S (SO)									
11. S (SO_2)							1.244		
12. C (CO)									
13. Br						0.001[b]	0.001[b]		1.739
14. I									1.769
15. C (sp)									1.452

[a] All entries are in units of log ($1/C_{50}$). [b] Converted to slightly attractive value during the study of predictive power of the method.

earlier synthetic efforts[22,23] some variations in the substitution were made by converting one of the amino groups to some substituted anilino or methylamino groups. However, we have not found any subsequent record in the literature of a substitution such as we suggest here. The treatment of the 2,4-diaminotriazines with alkali often leads to rearrangements,[23] thereby indicating some sort of instability in the compounds. Also, there may be problems in synthesizing the compounds having the predicted structures. It is therefore difficult to comment on the practicality of the prediction.

Statistics on the present study are given in Table VII. Table I shows that, in general, the difference between the calculated and observed log ($1/C$) is well below 0.5. Both correlation coefficient and the stanard deviation improve

(22) G. E. Foley, *Proc. Soc. Exp. Biol. Med.*, 83, 733 (1953).
(23) E. J. Modest, *J. Org. Chem.*, 21, 1 (1956).

Table VII. Necessary Statistics of the Study

study	site point used	no. of compds	no. of variables	correl coef	SD	max error
original data	9	48	25	0.892	0.596	1.89
without compds 1-3	9	45	25	0.935	0.460	0.98
data for prediction	9	91	0	0.79	0.907	2.57

markedly if the first three triazines are removed from the data set (see Table VII). It is difficult to comment on the degrees of freedom of our calculation. It does not correspond to the number of site points or their coordinates because these were not varied to improve the fit. Neither does it correspond to the number of used interaction parameters. In fact, energy parameters were evaluated by using the quadratic programming under several inequality constraints, where each was generated during the optimization algorithm to keep a binding mode energetically more favorable than any other for a particular molecule. Another difficulty in counting the degrees of freedom is that in the present algorithm some changes in the binding mode were allowed. Other QSAR methods suffer from similar questions about the degrees of freedom. For example, in molecular shape analysis, since both the choice of idealized conformation and the interacting conformations are dictated by a regression analysis,[13] their shape descriptor parameter, S_0, has more degrees of freedom than are immediately apparent. See Appendix for a detailed analysis of the relationship between numbers of parameters and data fit in distance geometry models.

Although each QSAR method differs from the others in its basic hypotheses, their ultimate results merit some comparison. The physicochemical parameter dependent QSAR has two eras, namely, QSAR in 1970 and 1980.[24] In the 1970's, QSAR workers tried to correlate the overall molecular properties to the biological activity. However, in more recent years, it is generally believed that the biological activity does not necessarily depend on the overall properties but rather on the properties of some specific parts of the molecule. This factoring of the molecular properties is a powerful tool for mapping the different parts of the receptor. Since most of these analyses considered molecules of very similar and often rigid structures, they indirectly assumed that the site geometry was constant. The distance geometry[9-11] approach is one step forward from these QSAR's, in that it treats the site geometry rather explicitly. It not only determines which of the interactions are geometrically allowed, but it also considers the other geometrically feasible binding modes. Consequently, the actual interaction should be the energetically most favorable one. Another interesting three-dimensional structure directed QSAR is the molecular shape analysis.[12] While it assumes such factoring of the molecular properties, it suggests that the common overlap region of the interacting shape of the ligand molecule with an idealized structure has a direct relationship to the biological activity of the molecule.

The present distance geometry model for dehydrofolate reductase inhibition is unique in its ability to fit three different sets of molecules (3′- and 4′-substituted phenyltriazines and quinazolines) in the same model, and successfully predicts the biological activity of 91 compounds.

The earlier QSAR by Dietrich et al.[7] and Hansch et al.[16] suggested that both the 3′ and 4′ substituents in triazines

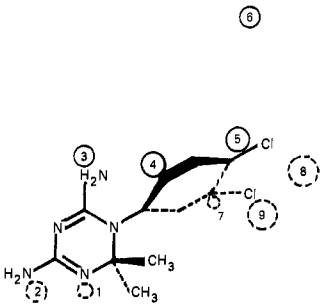


Figure 3. A three-dimensional view of the site points. The relative size of the circles represent the depth of the point from the plane of the paper. The dashed circles lie below the plane, and the solid circles lie above the plane.

Table VIII. Comparison of the Atomic Distances in the Active Conformation of 3′,4′-Dichlorotriazine Systems[a] with the Present Site Distances[b]

site 1	2	3	4	7	9
1	2.3 (2.4)	4.0 (4.2)	4.5 (4.5)	5.4 (5.7)	6.1 (6.0)
2		4.6 (4.8)	6.4 (6.6)	7.2 (7.7)	7.8 (7.9)
3			3.6 (4.2)	4.5 (5.1)	5.6 (5.7)
4				2.7 (2.9)	4.4 (4.3)
5					1.7 (1.6)
9					

[a] Reference 25. [b] The values within parentheses represent the site distances. All the values are in angstroms.

interact at the hydrophobic pockets of rat liver DHFR. In contrast, the present study suggests the existence of some dipolar interaction with the 3′-substituents, while the interaction of the 4′-substituent is not very clear, suggesting a dual nature of the site point. The interaction of site 5 with the 4′-substituent of triazines posed a geometrical problem too. A large site flexibility would have been necessary to bind the atoms with long bonds, like Cl, Br, and I. On the other hand, 4′-bromo- and 4′-iodotriazines had only slightly higher inhibition constants than the unsubstituted compound. Therefore, we chose to position site point 5 such that these atoms with long bonds avoided specific interaction with this site point.

One simplifying assumption of distance geometry analysis is that the differences in internal conformational energy between possible conformers of the drug molecule at the binding site are small compared to the binding energy. Molecular shape analysis[12] confirms our assumption in this case, since it found that the conformational energy did not make meaningful contributions to the correlation. Since neither Hansch et al.[7,16] nor Hopfinger et al.[12,13,25] made the combined analysis of the triazines and quinazolines, it is not possible to comment on the overall fit values. In fact, Hopfinger[25] suggested several alternate equations to explain the rat liver dihydrofolate reductase inhibition by 3′-substituted phenyltriazines. However, an excellent match was found between the geometry of our site points and their "active conformation" of the triazines. According to our model, six site points (1–4, 7, and 9) will

(24) Y. C. Martin, *J. Med. Chem.*, 24, 229 (1981).

(25) A. J. Hopfinger, *Arch. Biochem. Biophys.*, 206, 153 (1981).

Table IX. Comparison of the Site Distances with the Dihydrofolate Reductase Bound Phenyltriazine Atomic Distances[a]

site 1	2	3	4	5	7
1	2.3 (2.4)	4.0 (4.2)	4.4 (4.5)	7.4 (7.6)	5.5 (5.7)
2		4.6 (4.8)	6.4 (6.6)	9.3 (9.3)	7.2 (7.7)
3			3.8 (4.2)	6.2 (5.5)	4.2 (5.1)
4				3.7 (3.9)	2.3 (2.9)
5					2.3 (2.7)
7					

[a] Values within parentheses represent the site distances. See Figure 3 to identify the molecular points bound to the site points. All the values are in angstroms.

bind the 3',4'-dichlorophenyltriazine derivative Figure 3. All these site-point distances are within the range of site flexibility from the corresponding molecular points in the Hopfinger active conformation (see Table VIII).

Although the X-ray data on mammalian dihydrofolate reductase is not available, the structures are known for the complexes of methotrexate and NADPH with dihydrofolate reductase from *Escherichia coli*[4] and *Lactobacillus casei*[26] and of phenyltriazine and NADPH with chicken liver dihydrofolate reductase.[27] Since the sensitivity of vertebrate dihydrofolate reductase is different from those of the bacterial enzyme, let us compare our site model with the vertebrate (chicken liver) dihydrofolate reductase. There are two aspects of the comparison: (1) the geometrical aspects and (2) the nature of interaction.

Volz et al.[27] reported the principal torsion angle of the dihydrofolate reductase bound 2,4-diamino-5,6-dihydro-6,6-dimethyl-5-(4'-methoxyphenyl)-*s*-triazine as 88°, which is very close to the observed X-ray structure,[28] 82°, and that reported by Hopfinger[12] for the minimum-energy conformation, 70°. We compared our site distances with the respective molecular point distances of the phenyltriazine and found an excellent agreement between the two (Table IX). In fact, it is now quite certain that the phenyl ring is almost perpendicular to the triazine ring.

Volz et al.[27] showed that the phenyltriazine binds at the active site with its triazine ring occupying a position analogous to that found for the pyrimidine portion of the methotrexate when the latter is bound to dihydrofolate reductase from *E. coli* and *L. casei*. Cocco et al.[29] suggested that the pK_a of the ring nitrogen (N1) in enzyme-bound methotrexate is increased to a value greater than 10, which is similar to the pK_a (~11) of N1 in the enzyme-bound triazine system as reported by Baker et al.[30] It is therefore obvious that the triazine ring is protonated in the enzyme. In fact, these are the two basic hypotheses of our work. In addition, Volz et al. showed that the negatively charged carboxylate group of Glu-30 in the chicken enzyme makes hydrogen bonds to both the protonated ring nitrogen (N1) and the 2-amino group. The hydroxyl group of Thr-136 hydrogen bonds to a carboxylate oxygen of Glu-30 and to the 2-amino group of the inhibitor. The involvement of both the hydrogens of the 2-amino group in hydrogen bonding is probably responsible for its higher interaction with the site 2 compared to that of the hydroxyl group.

Hydrogen bonds are donated by the 4-amino group of the triazine to the carbonyl of Ile-7 and Val-115. The relative values of the interaction of site point 3 may be explained if only one carbonyl group takes the major part in the interaction, thereby reflecting the acidity of the hydrogen involved.

Since the crystallographic study[27] is not very explicit about the interaction of the substituted phenyl group, except suggesting the van der Waals contacts with side chains of Leu-22, Phe-34, Thr-56, Ser-59, Ile-60, Val-115, and the nicotinamide ring of NADPH, it is not possible to compare this portion of our model with the crystallographic result.

It should, however, be mentioned that the actual site of Volz et al.[27] has several other amino acid residues that do not correspond to our model. In fact, the enzyme's active site is a highly complex molecular system with a large number of atoms. One can hardly expect an exact correlation with our simplified model.

In order to test the predictive power of the model, we examined the 91 compounds (88 quinazolines and 3 others)[15-17] not included in our data set. However, some additional interaction parameters not used in the original study were necessary to test these compounds. All these interactions were arbitrarily assigned a very slightly attractive value of 0.001. The predictive power of the model has been summarized in Table X and XI. A careful review of Table X shows that the site point 4 and chlorine interaction, the one unused parameter in the original study, may have values as low as 0.00, comparing compounds **134** and **135**, or as high as 1.40, comparing compounds **79** and **97**. Since the objective was to test the predictive power of the present model, we did not try to adjust these unused interactions for better fit. Even then the overall fit produced a correlation coefficient of 0.790 and an rms deviation of 0.907, which are not much worse than those of the original data set. Since the additional interactions contributed almost nothing to the overall binding energy of the compounds of Table X, it was expected that the binding energies of the molecules having these interactions should be underestimated. This was found to be true in many cases. Compounds **61**, **86**, **88**, **94**, **98**, **114**, **115**, **117**, **123**, **125–127**, **130**, and **136**, all having Cl to site point 4 interaction, have been underestimated in the binding energy calculation. This feature indicates that Cl has a considerable interaction with site point 4. Such conclusion, however, does not totally hold for the site point 4 to $C(sp^3)$ interaction, since here the positive and negative deviations occur almost evenly.

Consideration of binding modes of these predicted compounds from Table XI suggests that the quinazoline ring and the substituents attached to it have a tendency to maintain a fixed binding mode, so that only the long substituents attached at the 5- or 6-position of the quinazoline ring differ in the general pattern of attachment to the receptor site.

The strong interaction of the OH groups wit site point 3 causes the poorly substituted molecule **52** to bind at the receptor in an odd orientation. In contrast, substitution of the weakly bound 4-NH_2 by the 4-SH group in molecule **73** swings the preference back to the usual binding mode. Although the poorly substituted compounds bind in a variety of (energetically optimal) orientations, the calculated binding energies are not overestimated. The suggests that such rotations or translations may occur at the real receptor site, a feature that has been established in *L. casei* dihydrofolate reductase either from proton nuclear mag-

(26) D. A. Matthews, R. 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. Xuong, and J. Kraut, *J. Biol. Chem.*, **253**, 6946 (1978).
(27) K. W. Volz, D. A. Matthews, R. A. Alden, S. T. Freer, C. Hansch, B. T. Kaufman, and J. Kraut, *J. Biol. Chem.*, **257**, 2528 (1982).
(28) W. E. Hunt, C. H. Schwalbe, K. Bird, and P. D. Mallinson, *Biochem. J.*, **187**, 533 (1980).
(29) L. Cocco, C. Temple, J. A. Montgomery, R. E. London, and R. L. Blakley, *Biochem. Biophys. Res. Commun.*, **100**, 413 (1981).
(30) B. R. Baker and J. H. Jordaan, *J. Pharm. Sci.*, **54**, 1740 (1965).

Table X. Test of Predictive Power of the Present Site and the Interaction Matrix

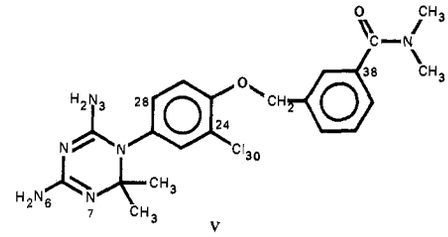
no.	substituents[a]	log (1/C) obsd[b]	log (1/C) calcd	$\Delta_{calcd-obsd}$
49	2-H, 4-NH_2	3.09	2.16	−0.93
50	2-H, 4-OH	3.62	3.27	−0.35
51	2-NH_2, 4-OH, 5,6,7,8-H_4	3.96	4.14	0.18
52	2-OH, 4-NH_2	4.02	3.80	−0.22
53	2-NH_2, 4-OH, 6-Me, 5,6,7,8-H_4	4.26	5.11	0.85
54	2,4-$(NH_2)_2$, 5-[SO(C_6H_3-3,4-Cl_2)]	4.26	4.97	0.71
55	2-AcNH, 4-OH	4.28	4.14	−0.14
56	2-NH_2, 4-OH	4.31	4.14	−0.17
57	2-NH_2, 4-OH, 7-CF_3	4.35	4.14	−0.21
58	2,4-$(NH_2)_2$, 5-[SO(2-$C_{10}H_7$)]	4.35	5.17	0.82
59	2-NH_2, 4-SH	4.37	4.00	−0.37
60	2-NH_2, 4-OH, 6-CH_3	4.44	3.80	−0.64
61	2-NH_2, 4-OH, 5-Cl	4.66	4.14	−0.52
62	2-H, 4-NH_2, 6-[S(2-$C_{10}H_7$)]	4.68	5.12	0.44
63	2-NH_2, 4-OH, 5-CH_3	4.80	4.14	−0.66
64	2,4-$(NH_2)_2$, 5-[SO_2(2-$C_{10}H_7$)]	4.82	5.17	0.35
65	2-H, 4-NH_2, 6-[SO(2-$C_{10}H_7$)]	4.82	4.64	−0.18
66	2-NH_2, 4-OH, 5-[SO_2(2-$C_{10}H_7$)]	4.85	6.28	1.36
67	2,4-$(NH_2)_2$, 6-CHO	5.00	5.06	0.06
68	2-SH, 4-OH, 6-[S(2-$C_{10}H_7$)]	5.00	6.23	1.23
69	2,4-$(NH_2)_2$, 6-CH_2NH_2	5.03	5.68	0.65
70	2,4-$(SH)_2$, 6-[S(2-$C_{10}H_7$)]	5.05	3.24	−1.81
71	2,4-$(OH)_2$, 6-[S(2-$C_{10}H_7$)]	5.07	6.88	1.81
72	2,4-$(NH_2)_2$, 5,6,7,8-H_4	5.13	4.71	−0.42
73	2-OH, 4-SH, 6-[S(2-$C_{10}H_7$)]	5.24	6.74	1.50
74	2,4-$(NH_2)_2$, 5-[*trans*-CH=CH(2-$C_{10}H_7$)]	5.28	6.78	1.50
75	2,4-$(NH_2)_2$, 6-Cl	5.40	6.09	0.69
76	2-H, 4-NH_2, 6-[SO_2(2-$C_{10}H_7$)]	5.47	5.62	0.15
77	2-SH, 4-NH_2, 6-[S(2-$C_{10}H_7$)]	5.52	5.12	−0.40
78	2-NH_2, 4-OH, 5-CH_3, 6-CO-Glu$(Et)_2$	5.57	5.88	0.31
79	2,4-$(NH_2)_2$, 6-Br	5.60	6.45	0.85
80	2,4-$(NH_2)_2$, 6-CH_3	5.66	5.68	0.02
81	2,4-$(NH_2)_2$, 6-CH_3, 5,6,7,8-H_4	5.70	5.68	−0.02
82	2,4-$(NH_2)_2$, 6-[$CH_2NHCOCH_2$(C_6H_3-3,4-Cl_2)]	5.82	7.55	1.73
83	2,4-$(NH_2)_2$, 5-CH_3	6.09	4.71	−1.38
84	2-AcNH, 4-OH, 6-[SO_2(2-$C_{10}H_7$)]	6.11	7.80	1.69
85	2-NH_2, 4-OH, 5-[S(2-$C_{10}H_7$)]	6.15	6.24	0.14
86	2,4,6-$(NH_2)_3$, 5-Cl	6.18	4.71	−1.47
87	2-NH_2, 4-OH, 6-CO-Glu$(Et)_2$	6.24	5.88	0.36
88	2,4-$(NH_2)_2$, 5-Cl	6.25	4.71	−1.54
89	2-NH_2, 4-OH, 5-CH_3, 6-CO-Glu	6.41	5.88	−0.53
90	2,4-$(NH_2)_2$, 5-[*cis*-CH=CH(2-$C_{10}H_7$)]	6.52	6.09	−0.43
91	2,4-$(NH_2)_2$, 6-[CH_2NHCO(C_6H_4-3-CF_3)]	6.59	7.46	0.87
92	2-NH_2, 4-SH, 6-[S(2-$C_{10}H_7$)]	6.64	7.16	0.52
93	2-NH_2, 4-OH, 6-CO-Glu	6.68	5.88	−0.80
94	2,4-$(NH_2)_2$, 5,6-Cl_2	6.70	6.09	−0.61
95	2,4-$(NH_2)_2$, 5-[S(C_6H_3-3,4-Cl_2)]	6.80	6.86	0.06
96	2-NH_2, 4-OH, 6-[S(C_6H_3-3,4-Cl_2)]	6.92	7.11	0.19
97	2,4-$(NH_2)_2$, 5-Cl, 6-Br	7.00	6.45	−0.55
98	2,4-$(NH_2)_2$, 6-[CH_2NHCO(2-$C_{10}H_7$)]	7.00	7.75	0.75
99	2,4-$(NH_2)_2$, 6-[CH_2NHCO(C_6H_3-3,4-Cl_2)]	7.00	7.55	0.55
100	2,4-$(NH_2)_2$, 6-[$NHCOCH_2$(C_6H_4-3-CF_3)]	7.02	7.46	0.44
101	2-NH_2, 4-OH, 6-[SO_2(C_6H_3-3,4-Cl_2)]	7.12	7.60	0.48
102	2,4-$(NH_2)_2$, 5-CH_3, 6-[$NHCOCH_2$(C_6H_4-3-CF_3)]	7.23	7.46	0.23
103	2,4-$(NH_2)_2$, 6-[$NHCO(CH_2)_2$(C_6H_4-4-Cl)]	7.33	7.55	0.22
104	2-NH_2, 4-OH, 6-[S(2-$C_{10}H_7$)]	7.35	7.31	−0.04
105	2-NH_2, 4-OH, 6-[SO(2-$C_{10}H_7$)]	7.38	5.64	−1.74
106	2,4-$(NH_2)_2$, 6-[S(C_6H_4-3-CF_3)]	7.44	6.65	−0.79
107	2,4-$(NH_2)_2$, 5-CH_3, 6-[$NHCOCH_2$(C_6H_4-3-Br)]	7.47	7.75	0.28
108	2,4-$(NH_2)_2$, 5-CH_3, 6-[$NHCO(CH_2)_2$(C_6H_4-4-Cl)]	7.48	7.55	0.07
109	2,4-$(NH_2)_2$, 5-CH_3, 6-[$NHCOCH_2$(C_6H_4-4-Cl)]	7.48	7.39	−0.09
110	2,4-$(NH_2)_2$, 5-CH_3, 6-[$NHCOCH_2$(C_6H_4-4-Br)]	7.49	6.98	−0.51
111	2,4-$(NH_2)_2$, 6-CO-Glu$(Et)_2$	7.54	6.45	−1.09
112	2,4-$(NH_2)_2$, 6-CO-Glu	7.59	6.45	−1.14
113	2,4-$(NH_2)_2$, 5-Cl, 6-[$NHCOCH_2$(C_6H_3-3,4-Cl_2)]	7.62	7.39	−0.23
114	2,4-$(NH_2)_2$, 5-Cl, 6-[$NHCOCH_2O$(C_6H_3-3,4-Cl_2)]	7.64	7.55	−0.09
115	2,4-$(NH_2)_2$, 6-[$NHCOCH_2$(C_6H_4-4-Cl)]	7.64	7.39	−0.25
116	2,4-$(NH_2)_2$, 5-Cl, 6-[$NHCOCH_2$(C_6H_4-3-CF_3)]	7.64	7.55	−0.09
117	2,4-$(NH_2)_2$, 5-CH_3, 6-[$NHCOCH[CH(CH_3)_2]S$(2-$C_{10}H_7$)]	7.68	9.19	1.51
118	2,4-$(NH_2)_2$, 5-CH_3, 6-CO-Glu$(Et)_2$	7.70	6.01	−1.69
119	2,4-$(NH_2)_2$, 5-[CH_2S(C_6H_4-4-Cl)]	7.70	6.93	−0.77
120	2,4-$(NH_2)_2$, 5-[CH_2CH_2(2-$C_{10}H_7$)]	7.70	6.85	−0.85
121	2,4-$(NH_2)_2$, 6-[$NHCH_2$(C_6H_3-3,4-Cl_2)]	7.74	7.39	−0.35
122	2,4-$(NH_2)_2$, 5-Cl, 6-[$NHCOCH(CH_3)C_6H_5$]	7.77	6.98	−0.79
123	2,4-$(NH_2)_2$, 5-CH_3, 6-[$NHCOCH_2O$(2-$C_{10}H_7$)]	7.82	7.75	−0.07

Table X (*Continued*)

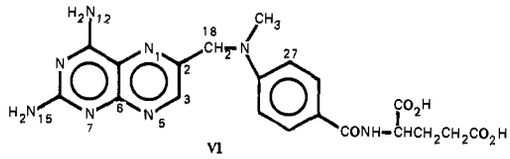
no.	substituents[a]	log (1/C) obsd[b]	log (1/C) calcd	$\Delta_{calcd-obsd}$
124	2,4-$(NH_2)_2$, 5-Cl, 6-[$NHCOCH_2O(2\text{-}C_{10}H_7)$]	7.85	7.75	−0.10
125	2,4-$(NH_2)_2$, 5-Cl, 6-[$NHCOCH_2(C_6H_4\text{-}4\text{-}Br)$]	7.89	6.98	−0.91
126	2,4-$(NH_2)_2$, 5-Cl, 6-[$NHCOCH_2(C_6H_4\text{-}4\text{-}Cl)$]	7.92	7.39	−0.53
127	2,4-$(NH_2)_2$, 6-[$S(C_6H_3\text{-}3,4\text{-}Cl_2)$]	7.96	7.67	−0.29
128	2,4-$(NH_2)_2$, 5-[$CH_2S(2\text{-}C_{10}H_7)$]	8.00	6.85	−1.15
129	2,4-$(NH_2)_2$, 5-Cl, 6-[$NHCOCH_2(C_6H_4\text{-}3\text{-}Br)$]	8.12	7.75	−0.37
130	2,4-$(NH_2)_2$, 6-[$S(2\text{-}C_{10}H_7)$]	8.15	7.87	−0.27
131	2,4-$(NH_2)_2$, 6-[$SO(2\text{-}C_{10}H_7)$]	8.15	7.13	−1.02
132	2,4-$(NH_2)_2$, 5-CH_3, 6-CO-Glu	8.25	6.45	−1.80
133	2,4-$(NH_2)_2$, 5-CH_3, 6-[$NHCOCH_2S(2\text{-}C_{10}H_7)$]	8.27	8.84	0.57
134	2,4-$(NH_2)_2$, 6-[$NHCOCH_2(2\text{-}C_{10}H_7)$]	8.30	6.98	−1.32
135	2,4-$(NH_2)_2$, 5-Cl, 6-[$NHCOCH_2(2\text{-}C_{10}H_7)$]	8.30	6.98	−1.32
136	2,4-$(NH_2)_2$, 5-CH_3, 6-[$NHCOCH_2(2\text{-}C_{10}H_7)$]	8.38	6.98	−1.40
137	(Baker's antifolate)	7.13	7.99	0.86
138	(methotrexate)	8.13	6.45	−1.68
139	(trimethoprim)	4.22	6.79	2.57

[a] Quinazoline derivatives, except for compounds 137–139. CO-Glu = $CH_2NHC_6H_4$-4-$CONHCH(COOH)CH_2CH_2COOH$.
[b] Data from ref 15–17.

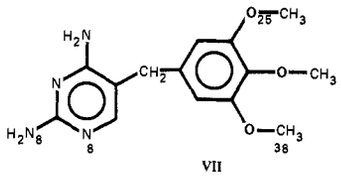
Chart II



V



VI



VII

netic resonance studies[31] or from ^{31}P nuclear magnetic resonance studies.[32]

Site point 7 was originally constructed for X_6 substituents. While in some molecules this is true, in many molecules the X_6 substituent goes to site point 9, especially when the 6-substituent is weakly attractive to site 7 but strongly to site 9. In fact, these two site points are not very far apart, thereby posing no geometrical problem in the interchangable use by X_6 substituents.

The calculated binding energies of Baker's antifolate, methotrexate, and trimethoprim merit some discussion. The calculated binding energy for Baker's antifolate (V, Chart II) corresponds to the observed value within the rms deviation. Its binding mode does not use the dimethyl amide group, and, in fact, the original data showed that compound 81 with no such group binds equally well at the receptor site. On the other hand, the binding energy of methotrexate has been substantially underestimated, probably because methotrexate has a ring system that was not present in the original data set from which the site geometry and interaction matrix were generated. This ring system may have some special interaction with the site that has not been included in our model and would give rise to the discrepancy.

Trimethoprim, on the other hand, shows the calculated binding energy to be much higher than the observed value. Although it is a pyrimidine derivative, the reason for the discrepancy seems to lie elsewhere. It has three substituents attached to the benzene ring, but no such substituent was included in the original triazines. Therefore, a possible repulsive site point for 5′-substituents has been overlooked. That would account for trimethoprim's weak observed binding energy. A ^1H and ^{31}P NMR study[33] of a ternary complex between *L. casei* dihydrofolate reductase, the coenzyme $NADP^+$, and trimethoprim suggests that unlike the ternary mixture with methotrexate, it exists in two interconvertable conformations as is evident from the difference in resonance at His-28. Although the explicit forms of these conformations are not very obvious, it may be that trimethoprim does not fit the normal enzyme as methotrexate does, and the conformational changes necessary for the binding process may be responsible for the low binding energy. On the other hand, Baker et al.[34] found from X-ray crystallography on the binary complex of trimethoprim and *E. coli* (strain RT 500) dihydrofolate reductase that the 2,4-diaminopyrimidine ring binds in the same way as is observed in methotrexate,[4] although the difference Fourier electron density between methotrexate and trimethoprim complexes showed that the diaminopyrimidine rings are not in identical environments in the two systems. At present, we are reluctant to conclude anything from these observations for the following reasons: (1) Besides trimethoprim, we did not use any other pyrimidine derivative in the data set. (2) The present model could not explain very well the binding energy of trimethoprim. (3) The work of Baker et al. is on binary complexes and with the *E. coli* enzyme for which trimethoprim is a very good inhibitor, whereas our model is based on rat liver DHFR for which trimethoprim is a weak

(31) E. I. Hyde, B. Birdsall, G. C. K. Roberts, J. Feeney, and A. S. V. Burgen, *Biochemistry*, **20**, 1717 (1981).
(32) E. I. Hyde, B. Birdsall, G. C. K. Roberts, J. Feeney, and A. S. V. Burgen, *Biochemistry*, **19**, 3738 (1980).
(33) A. Gronenborn, B. Birdsall, E. Hyde, G. Roberts, J. Feeney, and A. Burgen, *Mol. Pharmacol.*, **20**, 145 (1981).
(34) D. J. Baker, C. R. Beddell, J. N. Champness, P. J. Goodford, F. E. A. Norrington, D. R. Smith, and D. K. Stammers, *FEBS Lett.*, **126**, 49 (1981).

Table XI. Binding Modes of the Tested Molecules[a]

	site points								
no.	1	2	3	4	5	6	7	8	9
49			X_4		C_7				
50			X_4		C_7				
51	N_1	X_2	X_4	C_6					
52			X_2		C_5		C_8		
53	N_1	X_2	X_4						X_6
54	N_1	X_2	X_4	C_6	C'_2	X'_4			
55	N_1	X_2	X_4						
56	N_1	X_2	X_4						
57	N_1	X_2	X_4						
58	N_1	X_2	X_4	C_6	C'_2	X'_4			
59	N_1	X_2	X_4						
60			X_4		C_5		C_8		
61	N_1	X_2	X_4						
62	N_1		X_4		C'_2	X'_4	X_6		
63	N_1	X_2	X_4						
64	N_1	X_2	X_4	C_6	C'_2	X'_4			
65				C'_2	C_8	C_2			
66	N_1	X_2	X_4	C_6	C'_2	X'_4			
67	N_1	X_2	X_4						
68	N_1		X_4		C'_2	X'_4	X_6		
69	N_1	X_2	X_4						X_6
70				C_5	C'_2	X'_4	X_6		
71	N_1	X_2	X_4		C'_2	X'_4	X_6		
72	N_1	X_2	X_4	C_6					
73	N_1	X_2	X_4		C'_2	X'_4	X_6		
74	N_1	X_2	X_4		C'_2	X'_4			
75	N_1	X_2	X_4						X_6
76	N_1		X_4		C'_2	X'_4	S		$O(SO_2)$
77	N_1		X_4		C'_2	X'_4	X_6		
78	N_1	X_2	X_4		C'_2				X_6
79	N_1	X_2	X_4						X_6
80	N_1	X_2	X_4						X_6
81	N_1	X_2	X_4						X_6
82	N_1	X_2	X_4		C'_2	X'_3			X_6
83	N_1	X_2	X_4						
84	N_1	X_2	X_4		C'_2	X'_4	X_6		$O(SO_2)$
85	N_1	X_2	X_4	C_6	C'_2	X'_4			
86	N_1	X_2	X_4						
87	N_1	X_2	X_4		C'_2				X_6
88	N_1	X_2	X_4						
89	N_1	X_2	X_4		C'_2				X_6
90	N_1	X_2	X_4	C_6	C'_2		C'_6		
91	N_1	X_2	X_4		C'_5		C'_2	X'_3	F
92	N_1	X_2	X_4		C'_2	X'_4	X_6		
93	N_1	X_2	X_4		C'_2				X_6
94	N_1	X_2	X_4						X_6
95	N_1	X_2	X_4		C'_5				X'_3
96	N_1	X_2	X_4		C'_2	X'_4	X_6		
97	N_1	X_2	X_4						X_6
98	N_1	X_2	X_4		C'_2	X'_4			X_6
99	N_1	X_2	X_4		C'_1	X'_3			X_6
100	N_1	X_2	X_4		C'_1		C'_4	X'_3	F
101	N_1	X_2	X_4		C'_2	X'_4	S		$O(SO_2)$
102	N_1	X_2	X_4		C'_1		C'_4	X'_3	F
103	N_1	X_2	X_4		C'_2	X'_4	CH_2[b]		CH_2[c]
104	N_1	X_2	X_4		C'_2	X'_4	X_6		
105		X_2	X_4		C'_2	X'_4			
106	N_1	X_2	X_4	C_6	C'_2				
107	N_1	X_2	X_4		C'_1		C'_3		X'_3
108	N_1	X_2	X_4		C'_2	X'_4	CH_2[b]		CH_2[c]
109	N_1	X_2	X_4		C'_2		C'_4		X'_4
110	N_1	X_2	X_4				C'_4		X'_4
111	N_1	X_2	X_4		C'_2				X_6
112	N_1	X_2	X_4		C'_2				X_6
113	N_1	X_2	X_4		C'_1		C'_3		X'_3
114	N_1	X_2	X_4		C'_5	X'_3			CH_2
115	N_1	X_2	X_4		C'_2		C'_4		X'_4
116	N_1	X_2	X_4		C'_1		C'_3	F	X'_3
117	N_1	X_2	X_4		C'_2	X'_4	S	C[d]	C[e]
118	N_1	X_2	X_4		C'_1	C'_3			
119	N_1	X_2	X_4	C_6	C'_2				X'_4
120	N_1	X_2	X_4	C_6	C'_2	X'_4			
121	N_1	X_2	X_4		C'_1		C'_3		X'_3
122	N_1	X_2	X_4		C'_3		C'_1		CH
123	N_1	X_2	X_4		C'_2	X'_4			CH_2

Table XI (*Continued*)

	site points								
no.	1	2	3	4	5	6	7	8	9
124	N_1	X_2	X_4		C'_2	X'_4			CH_2
125	N_1	X_2	X_4				C'_4		X'_4
126	N_1	X_2	X_4		C'_2		C'_4		X'_4
127	N_1	X_2	X_4		C'_2	X'_4	X_6		
128	N_1	X_2	X_4	C_6	C'_2	X'_4			
129	N_1	X_2	X_4		C'_1		C'_3		X'_3
130	N_1	X_2	X_4		C'_2	X'_4	X_6		
131	N_1	X_2	X_4		C'_2	X'_4			O(SO)
132	N_1	X_2	X_4		C'_2				X_6
133	N_1	X_2	X_4		C'_2	X'_4	S		CH_2
134	N_1	X_2	X_4			X'_4			CH_2
135	N_1	X_2	X_4			X'_4			CH_2
136	N_1	X_2	X_4			X'_4			CH_2
137[f]	N_7	N_6	N_3	C_{28}		C_{38}	C_{24}		Cl_{30}
138[g]	N_7	N_{15}	N_{12}		C_{27}				C_{18}
139[h]	N_6	N_8	O_{25}						C_{36}

[a] The binding mode is given in reference to structure IV. [b] Second CH_2 group of the chain linking the quinazoline and the benzene ring. [c] First CH_2 group of the same chain. [d] The middle carbon of the isopropyl group. [e] The CH carbon of the chain. [f] Atom numbering of Baker's antifolate is given in structure V. [g] Atom numbering of methotrexate is given in structure VI. [h] Atom numbering of trimethoprim is given in structure VII.

Table XII. Hypothetical Molecules in Two Dimensions

name structure ΔG_{obsd}	A H2 / \ C1—N3 −5		B N2 / \ C1—H3 −3	
geometrically allowed modes for sites[a]	no. of modes	modes	no. of modes	modes
s1	4	0; C1; H2; N3	4	0; C1; H3; N2
s1 s2	7	0 0; C1 0; H2 0; N3 0; 0 C1; 0 H2; 0 N3	7	0 0; C1 0; H3 0; N2 0; 0 C1; 0 H3; 0 N2
s1—s2	13	0 0; C1 0; H2 0; N3 0; 0 C1; 0 H2; 0 N3; C1 H2; C1 N3; H2 N3; H2 C1; N3 C1; N3 H2	13	0 0; C1 0; H3 0; N2 0; 0 C1; 0 H3; 0 N2; C1 H3; C1 N2; H3 N2; H3 C1; N2 C1; N2 H3
s3 s1 s2	13	0 0 0; C1 0 0; H2 0 0; N3 0 0; 0 C1 0; 0 H2 0; 0 N3 0; 0 0 C1; 0 0 H2; 0 0 N3	13	0 0 0; C1 0 0; H3 0 0; N2 0 0; 0 C1 0; 0 H3 0; 0 N2 0; 0 0 C1; 0 0 H3; 0 0 N2
s3 s1—s2	16	0 0 0; C1 0 0; H2 0 0; N3 0 0; 0 C1 0; 0 H2 0; 0 N3 0; 0 0 C1; 0 0 H2; 0 0 N3; C1 H2 0; C1 N3 0; H2 N3 0; H2 C1 0; N3 C1 0; N3 H2 0	16	0 0 0; C1 0 0; H3 0 0; N2 0 0; 0 C1 0; 0 H3 0; 0 N2 0; 0 0 C1; 0 0 H3; 0 0 N2; C1 H3 0; C1 N2 0; H3 N2 0; H3 C1 0; N2 C1 0; N2 H3 0
s3 / \ s1—s2	16	0 0 0; C1 0 0; H2 0 0; N3 0 0; 0 C1 0; 0 H2 0; 0 N3 0; 0 0 C1; 0 0 H2; 0 0 N3; C1 H2 0; C1 N3 H2; H2 C1 N3; H2 N3 0; N3 C1 0; N3 H2 C1	16	0 0 0; C1 0 0; H3 0 0; N2 0 0; 0 C1 0; 0 H3 0; 0 N2 0; 0 0 C1; 0 0 H3; 0 0 N2; C1 H3 N2; C1 N2 0; H3 C1 0; H3 N2 C1; N2 C1 H3; N2 H3 0

[a] Pictorial representation of classes of site configurations discussed in text; single point, two points close together, two points far apart, three points all close, three points with two far apart, and three points all well separated.

inhibitor. (4) Our inhibition data are for the ternary system including NADPH. (5) Baker et al. have yet to give any explanation for species specificity of methotrexate and trimethoprim. (6) We are currently developing a common model for a large variety of DHFR inhibitors, in an effort to throw light on the problem in the near future.

In our earlier studies[10,11] we assumed that the 2,4-diaminoquinazolines have different binding modes compared to the monoamino or nonamino derivatives. The fact, many authors suggested from the spectroscopic evidence that methotrexate, folate, and dihydrofolate are bound somewhat differently at the active site. Matthews et al.[26] determined the dihydrofolate reductase–methotrexate–NADPH complex structure by X-ray crystallography. Although the geometrical relationship between the nicotinamide ring of NADPH and the pteridine ring of methotrexate is consistent with the known A side reaction stereochemistry, Fontecilla-Camps et al.[35] later determined

(35) J. C. Fontecilla-Camps, C. E. Bugg, C. Temple, Jr., J. D. Rose, J. A. Montgomery, and R. L. Kisliuk, *J. Am. Chem. Soc.*, 101, 6114 (1979).

the absolute configuration of tetrahydrofolate and showed that the configuration at the asymmetric C6 cannot be explained with Matthews' model. Afterwords, Charlton et al.[36] showed by NMR that the reduction at C7 also takes place from the same side as that of asymmetric C6. They therefore concluded that there is no major difference in the oxidized and reduced pteridine ring of folate when bound to the enzyme. However, since some earlier works[37-39] suggested that methotrexate and dihydrofolate compete for the same binding site, Matthews et al.[26] already suggested a possible alternate binding mode for the folate in which the pteridine ring is rotated by 180° around C2–C18 (VI). This mode of course explains the stereochemistry of reduction in folates. Following the same type of arguments, we[10,11] also succeeded in explaining the high activity of 2,4-diaminoquinazolines over the mono- or nonamino derivatives toward *S. faecium* dihydrofolate reductase. The alternate binding mode was attributed to the protonation of the ring nitrogen (N1) in the 2,4-diamino compounds. At that time we faced the trouble of explaining four loosely bound 2,4-diaminoquinazolines, all of which had 5-sulfinyl or 5-sulfonyl substituents. We concluded that their low activity can be explained either by some steric repulsion of these bulky substituents with the receptor site or by their electron withdrawing via resonance from the ring nitrogens. There is another way that these substituents may decrease the basicity of N1: the steric inhibition of resonance by disallowing[13] the 4-amino group to be planar. At this point we became suspicious of the generally accepted theory that some inhibitors are protonated at N1 leading to the alternate binding mode and stronger binding. In fact, we assumed in our work that the quinazolines with two amino groups became protonated in solution, whereas the general belief is that in dihydrofolate, the N1 ring nitrogen is not protonated, although it is flanked by two amino groups. On the other hand, Fukunaga et al.[17] explained the data set without going to these complications. All these ideas led us to ask whether a unique binding model can explain the data of all the inhibitors. However, it should be stated that this work in no way casts doubt on the different binding modes of methotrexate and dihydrofolate. Instead, it is just to show that the activity of the inhibitors may be explained by using a single binding mode for the triazine and quinazoline systems.

Conclusion

From the present as well as our previous studies it is obvious that the distance geometry approach is not only capable of fitting the biological data but also giving a geometric interpretation of the biological data, which can suggest the structural modifications that should be made in future drug research. One advantage of the simplified site geometry and the interaction matrix over the actual X-ray crystallographic data is that even when the X-ray structure of the enzyme-bound inhibitor is known, it is very difficult to interpret which part of the enzyme is the most important in the binding. The simplified distance geometry model of the site focusses on only the important parts of the site geometry with the help of the interaction matrix, thereby making the interpretation much easier.

Acknowledgment. This work was supported by the National Institute of Arthritis, Metabolism, and Digestive Diseases (Grant 1-R01-AM28140-01). The authors express their gratitude to Dr. E. F. Meyer for the use of his computer graphic system and to G. Cole for his sincere assistance. We thank Dr. B. Roth for helpful discussions.

Appendix

Workers in the QSAR field are so accustomed to linear regression analyses that all studies of binding data are judged in terms of the number of compounds in the study, the number of adjustable parameters in the model, and the resultant correlation coefficient, ρ. By these standards, the distance geometry approach makes a poor showing, since it generally requires a relatively large number of site points and adjustable energy parameters compared to the number of compounds and the final ρ. A distance geometry study of medicinal interest on a large set of complicated molecules consists of guessing the number of site points, guessing a set of binding modes, calculating coordinates for the site points, and then adjusting the energy parameters for a least-squares fit to the binding data subject to the constraint that the proposed modes have lower calculated energy than any other geometrically allowed mode for the molecule. Unfortunately it is not feasible to assess the quality of a distance geometry study on such a set of molecules, because that would involve finding all possible site models and showing that the one selected by the investigator is relatively "good". The best we can do in practice is to present one possible model of the binding site and demonstrate that it is in agreement with the observed binding free energies of the ligands and their structures. Therefore, we now present an exhaustive distance geometry analysis of an extremely simple problem, which illustrates that for a satisfactory explanation of the observed binding it may be necessary to invoke more site points than ligands, and even having many more adjustable energy parameters than ligands will not guarantee a solution! The advantage of this extremely simple data set is that we can easily enumerate all possible binding modes (which may run into the hundreds for realistic binding studies) and explicitly examine all possible combinations of proposed modes to see which ones correspond to solutions and, if so, how good the fit is.

Suppose we have only two drug molecules, called A and B, which are planar, equilateral triangles with a bond length of 1 unit, as shown in Table XII. The three atoms in each are numbered clockwise as 1, 2, or 3, and their types are given as C, H, and N. Suppose further that the molecules exist in only two spatial dimensions, so that they are the two-dimensional analogues of an enantiomeric pair. They are distinct because they cannot employ the third dimension to flip over. This example could as well be done with three dimensions and nonplanar molecules of four or more atoms, but it would be harder to visualize and much more lengthy to calculate. When we write a mode for molecule A interacting with a three-point site, such as "A:N3 C1 0", this means that atom N3 is contact with site point s1, C1 is in contact with s2, and no atom is in contact with s3. The energetics of the various binding modes are calculated from the interaction energy parameters, which we denote by, for example, $e_{C,s1}$. The first subscript refers to the atom type, and the second subscript refers to the site point. The distance geometry model insists that for given e's, each molecule will bind in the geometrically allowed mode that gives the lowest calculated binding energy, which is just the sum over the contacts in the mode

(36) P. A. Charlton, D. W. Young, B. Birdsall, J. Feeney, and G. C. K. Roberts, *J. Chem. Soc., Chem. Commun.*, 922 (1979).
(37) F. M. Huennekens, G. P. Mell, N. G. L. Harding, L. E. Gundersen, and J. H. Freisheim, "Chemistry and Biology of Pteridines", K. Iwai, M. Akino, M. Gato, and Y. Iwanami, Eds., International Printing C. Ltd., Tokyo, 1976, pp 329–350.
(38) M. N. Williams, M. Poe, N. J. Greenfield, J. M. Hirshfield, and K. Hoogsteen, *J. Biol. Chem.*, 248, 6375 (1973).
(39) E. J. Pastore, L. T. Plante, J. M. Wright, R. L. Kisliuk, and N. O. Kaplan, *Biochem. Biophys. Res. Commun.*, 68, 471 (1976).

of the corresponding e's. "No contact" contributes zero to the calculated binding energy, favorable contacts have a negative contribution, and unfavorable ones are positive. The site structure determines the list of geometrically allowed modes, and the interaction energy parameters determine which mode is the preferred one for each molecule. We will now explicitly enumerate all possible models of a binding site that would explain the ΔG's of binding given in Table XII.

The simplest site would consist of a single site point, denoted by s1. Its position is space is irrelevant, so there is only one possible configuration for this site. There are only three energy parameters (as in Table VI), since there are three atom types and one site point. Let $e = \min(e_{C,s1}, e_{H,s1}, e_{N,s1})$. If $e > 0$ (i.e., there is no favorable interaction with any atom), then neither molecule will contact the site, and the calculated binding energy for both will be zero. Hence, σ (the rms deviation of the fit) = 4.1, and ρ (the correlation coefficient) = 0. Alternatively, if $e < 0$, then both molecules will want to bind the single the single most favorably interacting atom with the one site point. If $e = e_{N,s1}$, then the energetically preferred mode for A would be A:N3, but since B:N2 is also geometrically allowed, that will be the preferred mode for B. Whichever energy parameter is least, it can be set to -4 by least-squares fitting of the calculated to the observed binding energy, with the other two e's > -4. Then, $\sigma = 1$, but still $\rho = 0$ because the same energy will always be calculated for both molecules. Whether the minimal energy parameter is greater or less than zero, it is clear that although each molecule has available to it the four binding modes given in the first line of Table XII, only 4 of the 16 possible combinations of modes correspond to the energetically best binding modes for either a trivial (completely repulsive) or least-squares fit to the observed binding energies. These are A:0 and B:0; A:C1 and B:C1; A:H2 and B:H3; and A:N3 and B:N2. These combinations of modes are the four solutions referred to in the first line of Table XIII. For example, the combination of modes "A:N3 and B:N2" gives a least-squares fit to the data when $e_{N,s1} = -4$ and the other e's > -4. The choice "A:N3 and B:C1" is not a solution, because in order for A to prefer binding N3 to s1, $e_{N,s1}$ must be the least of all the e's, but then B would prefer also binding its N2 to s1, since that mode is available to it.

The next most complicated possible site consists of two points, s1 and s2. Now we must consider the site configuration, which was trivial for a single site point. We must examine all possible site geometries in a thorough, unbiased fashion. Simply dividing the possibilities into appropriate classes and counting only one example from each class would not correctly simulate a random search over all site configurations. Without loss of generality, assume that s1 is located at the origin and s2 lies on the positive x-axis. Suppose the site flexibility parameter $\delta = 0.1$ unit, so that the mode A:C1 H2 is geometrically allowed if the distance between s1 and s2 is in the range 0.9 to 1.1, since the C1–H2 distance is 1 unit. Considering that the greatest interatomic distance in either A or B is only 1 unit, we can restrict the choices of intersite point distances to the set {0.1, 0.3, 0.5, 0.7, 0.9}. Any values between these five choices are indistinguishable by the distance geometry algorithm. If we take the s1–s2 distance to be 0.1, 0.3, 0.5 or 0.7, then either molecule can make, at most, a single contact at a time, and the situation is equivalent to the case of one site point (see 2 line of Tables XII and XIII). There is only a single s1–s2 distance, namely, 0.9, where a pair of contacts can be made. As shown in line 3 of Table XII, each molecule can engage in no contacts, any one of the three atoms in contact with either of the site points, or any pair of atoms in contact with both. Although there are $13 \times 13 = 169$ possible combinations of modes, only 13 of these lead to solutions, namely, when both A and B have the same types of atoms in contact, e.g., "A:H2 0 and B:H3 0" or "A:C1 H2 and B:C1 H3". The reason is the same as for one site point: if the energy parameters make a particular mode for A the most favorable and all others worse, then B has the corresponding preferred mode (for which the calculated interaction energy is the same sum of energy parameters), and B has as alternate modes only those that correspond to the worse ones of A. The optimal fit amounts to adjusting one parameter, but since the calculated binding energies of both molecules must be the same for any of these solutions, we have as before $\sigma = 1$ and $\rho = 0$. There is a trivial case of no contacts for both A and B having $\sigma = 4.1$, as before. Observe that there is no solution with ρ not equal to 0, even though we are using as many site points as molecules. In spite of there being $2 \times 3 = 6$ energy parameters, the binding modes for solutions are such that only one parameter can be adjusted to fit the observed binding energies.

The last possibility we will consider is that of three site points. Calculating the number of site configurations becomes more difficult with three site points. As before, s1 is at the origin, s2 lies on the positive x-axis with x values chosen from the set {0.1, 0.3, 0.5, 0.7, 0.9}, and now s3 lies on the first quadrant of the xy-plane with x and y coordinates chosen from the same set. The first case is where the three points all lie within less than 0.9 unit from each other, some 60 possible configurations. The outcome is the same as for a single site point. The second case is where the s1–s2 distance is 0.9, but the s1–s3 and s2–s3 distances are less, altogether 15 configurations. Only contact pairs can be made, and the outcome is the same as for two site points. The only different case is the single configuration with s1 at (0, 0), s2 at (0.9, 0) and s3 at (0.5, 0.9). Only then is it possible to achieve three contacts simultaneously. As shown in Table XII, only with the triple contacts do the two molecules behave differently. As before, the trivial noncontact case is a solution, and having analogous single contacts for both yields solutions. Each of the $6 \times 6 = 36$ possible combinations of double or triple contact modes yields an interesting solution. For example, if we choose "A:C1 N3 H2 and B:C1 N2 0", the least-squares fit requires that $e_{C,s1} + e_{N,s2} = -3$, $e_{H,s3} = -2$, and $0 > e_{C,s1} > -3$, effectively fixing two parameters. Then $\sigma = 0$ and $\rho = 1$. Other choices of these last 36 sets of binding modes yield the same fit but different values of the energy parameters.

Clearly, any number of site points could be used, but going beyond three would introduce utterly useless ones, since the best three-point site given above would always appear as subsets of the whole site.

Table XIII summarizes the exhaustive analysis. If we choose site point configurations and mode combinations at random, there are 10477 sensible choices, of which only 919 lead to solutions by proper adjustment of the energy parameters. Even give that a solution has been found, by far the most likely outcome is that $\rho = 0$. In all, the chance of finding a site geometry and combination of binding modes that perfectly fits the data is only $36/10477 = 0.003$. Note that this can be achieved only when there are three site points, which is more than the number of ligands in the study. We conclude that the number of site points is determined more by the structural complexity of the ligands than by their number. Finding a solution of any sort for a distance geometry model of a site is nontrivial and

Table XIII. Summary of Solutions

site[a]	site configns	mode combinations	solutions with σ = 4.1	1	0
s1	1	16	1	3	0
s1 s2	4	49	4	12	0
s1—s2	1	169	1	12	0
s3 s1 s2	60	100	60	540	0
s3 s1—s2	15	256	15	225	0
s3 / \ s1——s2	1	256	1	9	36

[a] See footnote *a* of Table XII.

involves a carefully coordinated choice of site point coordinates and proposed binding modes. The odds are that the initial choice of modes will not lead to a solution, so that some provision for altering the initial choice is necessary.[11] Note that the mere availability of nine energy parameters to fit two binding energies in the case of three site points was not sufficient; correct binding modes had to be proposed. Distance geometry binding studies cannot be judged by the same criteria as those used for more conventional QSAR methods.

Registry No. 1, 85304-87-6; 2, 70579-32-7; 3, 17740-29-3; 4, 90-08-4; 5, 85304-88-7; 6, 70579-34-9; 7, 70650-62-3; 8, 4022-58-6; 9, 47071-11-4; 10, 3567-84-8; 11, 17711-73-8; 12, 21316-30-3; 13, 46781-41-3; 14, 4653-75-2; 15, 15233-37-1; 16, 3850-94-0; 17, 1542-59-2; 18, 70743-55-4; 19, 70579-41-8; 20, 4038-60-2; 21, 70579-39-4; 22, 70579-36-1; 23, 13351-02-5; 24, 70579-38-3; 25, 70606-63-2; 26, 24849-96-5; 27, 51012-14-7; 28, 1492-81-5; 29, 4653-73-0; 30, 19161-84-3; 31, 70579-31-6; 32, 85304-89-8; 33, 17944-10-4; 34, 253-82-7; 35, 50440-89-6; 36, 86-96-4; 37, 50828-14-3; 38, 13741-90-7; 39, 1899-48-5; 40, 18917-68-5; 41, 52979-06-3; 42, 55096-56-5; 43, 55096-39-4; 44, 52979-00-7; 45, 55096-41-8; 46, 52979-11-0; 47, 51123-28-5; 48, 51123-83-2; 49, 15018-66-3; 50, 491-36-1; 51, 33081-07-1; 52, 50440-88-5; 53, 50440-83-0; 54, 50828-13-2; 55, 50440-87-4; 56, 20198-19-0; 57, 50440-86-3; 58, 50828-19-8; 59, 49873-59-8; 60, 50440-82-9; 61, 50440-85-2; 62, 52979-15-4; 63, 50440-84-1; 64, 50828-20-1; 65, 53159-19-6; 66, 50828-21-2; 67, 27023-77-4; 68, 52979-04-1; 69, 38944-10-4; 70, 52979-02-9; 71, 52979-05-2; 72, 1899-40-7; 73, 52979-03-0; 74, 50828-17-6; 75, 18671-95-9; 76, 53159-19-6; 77, 52978-97-9; L-78, 58724-36-0; 79, 50440-75-0; 80, 1955-61-9; 81, 1899-41-8; 82, 55096-64-5; 83, 27018-14-0; 84, 52979-13-2; 85, 50930-12-6; 86, 17511-20-5; L-87, 27069-81-4; 88, 17511-21-6; L-89, 58724-37-1; 90, 50828-08-5; 91, 55096-66-7; 92, 52978-99-1; L-93, 5854-11-5; 94, 50546-08-2; 95, 50828-12-1; 96, 52979-09-6; 97, 41934-85-4; 98, 55096-67-8; 99, 55096-15-6; 100, 55096-42-9; 101, 53667-27-9; 102, 55096-44-1; 103, 55096-55-4; 104, 52979-08-5; 105, 52979-10-9; 106, 51123-99-0; 107, 55096-54-3; 108, 55096-57-6; 109, 55096-50-9; 110, 55096-52-1; L-111, 58724-38-2; L-112, 18921-68-1; 113, 55096-40-7; 114, 55096-62-3; 115, 55096-48-5; 116, 55096-43-0; 117, 55096-61-2; L-118, 24205-32-1; 119, 50828-16-5; 120, 50828-09-6; 121, 13794-65-5; 122, 55096-63-4; 123, 55096-59-8; 124, 55096-58-7; 125, 55096-51-0; 126, 55096-49-6; 127, 51124-31-3; 128, 43170-98-5; 129, 55096-53-2; 130, 51124-09-5; 131, 51123-77-4; L-132, 32043-09-7; 133, 55096-60-1; 134, 55096-45-2; 135, 55096-46-3; 136, 55096-47-4; dihydrofolate reductase, 9002-03-3.

Substituent Effects in Cephalosporins as Assessed by Molecular Orbital Calculations, Nuclear Magnetic Resonance, and Kinetics

Donald B. Boyd

Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, Indiana 46285. *Received October 28, 1982*

For cephalosporins with different side chains at position 3, the quantum mechanically computed charge distribution in the β-lactam carbonyl group can be correlated with observables, such as carbon-13 chemical-shift differences at C_3 and C_4 of the dihydrothiazine ring and alkaline rates of hydrolysis of the β-lactam. The relationship of these properties and the theoretical transition-state energy (TSE) corroborate the fact that chemical reactivity is one important determinant affecting inhibitory activity of cephalosporins against peptidoglycan-regulating enzymes.

Several physicochemical properties of cephalosporins have now been identified that can be related to antibacterial activity. These properties reflect the effects on the rest of the molecule that derive from the different substituents at position 3 of the 3-cephem nucleus. For instance, molecular orbital calculations can be used to evaluate the ease of approach of a nucleophile to a 3-substituted 3-cephem in a model reaction.[1-4] The calculations yield a transition-state energy (TSE), which is defined as the change in the CNDO/2 total energy of the 3-cephem–OH^- complex (formed by placing OH^- 1.5 Å from the α face of the β-lactam carbonyl carbon) with respect to the sum of the energies of the separated 3-cephem and OH^- reactants.[3] TSE values for cephems with 13 different R groups are given in Table I. Depending on the 3-substituent, those cephalosporins with a more favorable energy of interaction in the complex tend to exhibit better in vitro Gram-negative activity.

Another property more recently found to correlate with minimum inhibitory concentrations (MICs) of cephalosporins is the difference in ^{13}C chemical shifts for carbons 3 and 4 of the dihydrothiazine ring of the 3-cephem nucleus.[5,6] This correlation was discovered by making note of the fact that cephalothin and cephaloridine had been reported to have larger $\Delta\delta(4-3)$ quantities than cephalexin.[7] Not only does $\Delta\delta(4-3)$ correlate with MICs, but it also correlates linearly with TSEs and inductive substituent constants σ_I for the 3-position side chains.[6]

Also, it is known that the TSEs and antibacterial activities each correlate linearly with alkaline hydrolysis rates for a series of 7-(thien-2-ylacetyl)cephalosporins.[8] The

(1) Boyd, D. B.; Hermann, R. B.; Presti, D. E.; Marsh, M. M. *J. Med. Chem.* **1975**, *18*, 408.
(2) Boyd, D. B.; Lunn, W. H. W. *J. Antibiot.* **1979**, *32*, 855.
(3) Boyd, D. B.; Herron, D. K.; Lunn, W. H. W.; Spitzer, W. A. *J. Am. Chem. Soc.* **1980**, *102*, 1812.
(4) Boyd, D. B. *Ann. N.Y. Acad. Sci.* **1981**, *367*, 531.
(5) Tori, K.; Nishikawa, J.; Takeuchi, Y. *Tetrahedron Lett.* **1981**, *22*, 2793.
(6) Nishikawa, J.; Tori, K. *J. Antibiot.* **1981**, *34*, 1641.
(7) Paschal, J. W.; Dorman, D. E.; Srinivasan, P. R.; Lichter, R. L. *J. Org. Chem.* **1978**, *43*, 2013.

0022-2623/83/1826-1010$01.50/0 © 1983 American Chemical Society