a Parr hydrogenation apparatus using palladium on charcoal as catalyst in ethyl acetate. After the uptake of hydrogen had ceased, the solution was filtered and the solvent was removed under reduced pressure. The crude 2-amino-4-ethoxyacetanilide was recrystallized from aqueous ethanol to give white needles: yield 100%; mp 141 °C. 2-Amino-4-ethoxyacetanilide (0.052 mol) was refluxed for 24 hours in 4 M HCl (30 mL). Cooling and neutralization with 10% KOH solution precipitated a solid, which was isolated by filtration and recrystallized from toluene–hexane: yield 76%; mp 137–140 °C. Anal ($C_{10}H_{12}N_2O$) C, H.

5(6)-Ethoxybenzimidazole (79). 2-Amino-4-ethoxyacetanilide (0.052 mol) was refluxed with 20% KOH solution (100 mL) for 2 h until hydrolysis was complete. The product, 4-ethoxy-o-phenylenediamine (0.052 mol) was refluxed with formic acid (20 mL) in 4 M HCl (80 mL) for 4 h and isolated as for compound **80**: yield 78%; mp 116 °C. Anal ($C_9H_{10}N_2O$) C, H.

Quantitative Structure-Activity Relationship by Distance Geometry: Quinazolines as Dihydrofolate Reductase Inhibitors

Arup K. Ghose and Gordon M. Crippen*

Department of Chemistry, Texas A&M University, College Station, Texas 77843. *Received October 13, 1981*

This is a reinvestigation of 68 quinazoline inhibitors of dihydrofolate reductase. As in the earlier study, the binding data fitted to an 11-point model of the site, but improved computer algorithms resulted in a much better overall fit (correlation coefficient 0.95, standard deviation 0.727 kcal) and a more accurate fit for some very loosely bound 2,4-diaminoquinazolines. Removal of two of the site points (numbers 5 and 9) gave an even better fit than the original 11 site points. However, deleting a third one (number 8) worsened the calculated binding energies of the loosely bound 2,4-diaminoquinazolines. The results lead to predictions of chemical modifications of the quinazolines that should improve their biological activity.

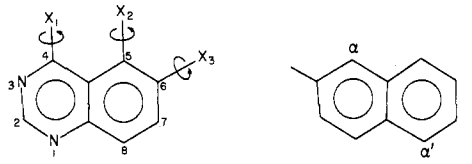
This is the fifth paper in the series[1-4] on the distance geometry approach to quantitative structure–activity relationships. The energetic treatment in this approach to QSAR resembles a linear free energy model or Free–Wilson analysis; the added feature is the geometric constraints during the fitting of the data so that the ultimate outcome is a geometric interpretation of the biological activity. Ideally one should treat all aspects of drug action, i.e., transport of the drug molecule from point of administraiton to the receptor site, in vivo chemical midification along the way, free-energy changes within a flexible drug molecule to adopt the conformation required by the active site, free-energy changes within the site to adapt to the bound molecule, the energetics of the molecule–site interaction, and finally the production of the biological response. Since the first two and the last steps are complex and poorly understood, we make no attempt to take them into account and cosider only in vitro binding assays. We further simplify the situation by assuming the site is relatively rigid and that the intramolecular energy change upon binding is small compared to the interaction energy between site and drug molecule.

Methods

The calculations proceed as follows: (1) The molecules are constructed from the crystallographic data[5] on their constituent fragments. (2) For each molecule, we calculate the matrix of upper and lower bounds on the distances between its atoms over all sterically allowed conformations. In order to take only the physically acceptable conformations of low energy into consideration, slight penetration beyond the van der Walls radii or the radii of closest approach[6,7] may be allowed, although a better approach would be to select only those conformations having energies close to the global minimum. (3) In order to reduce the computer time required for the subsequent steps, some atoms of each molecule must be deleted. (4) Next, a plausible binding mode for each molecule is selected. A binding mode specifies which molecular point should be bound to which site point. Examples are included in Table IV. Often for a data set having varied molecular structures it is preferable to specify initially the binding modes of only those molecules having similar structures and high biological activity. (5) On the basis of the specified binding modes, the site point distance bounds are evaluated, using the earlier reasoning[1] that each intersite point distance should lie in the range common to the corresponding interatomic distances in all the molecules. In many cases no such overlapping range may exist, for instance when the highest of the lower bounds (l) is slightly higher than the least of the upper (u) distance bounds. Under that condition, the two site points may still bind the respective molecular points if it is assumed that the site point distance is intermediate between these two limits and the site flexibility, $\delta \geq (l - u)/2$. However, the binding modes should be chosen such that the value of δ is minimal. (6) The site point distance limits evaluated in this way have been selected quite independently of one another. However, the intersite point distances may be expressed by only $3n - 6$ coordinates, where n is the number of site points, and there are $n(n - 1)/2$ intersite distances. It is therefore obvious that such distances or distance limits cannot hold in three-diminensional space in general. The structures having the closest fit with these distance bounds are determined by the method of Crippen et al.[8] (7) Returning to the problem of step 5, since there is no guarantee in these derived coordinates that the site point distances will be exactly at the mean of the respective two limits, it is necessary to reevaluate the final value of δ corresponding to these coordinates. (8) In case some molecules have ambiguities about the molecular points to be bound with some site points, the site point coordinates, δ, and the intramolecular distance bounds may be used to select which molecular points will best fit these site points. (9) After the site point coordinates, δ, and fixed binding modes from the above algorithm are selected, it is next necessary to check whether these fixed modes are all geometrically acceptable and have no unfavorable forced contacts.[1] This step is very important, since in the next step the various interaction energy parameters are adjusted such that this mode becomes the energetically optimal binding mode, which must also be geometrically feasible. (10) Finally, the interaction parameters are evaluated by quadratic programming,[9]

(1) G. M. Crippen, *J. Med. Chem.*, **22**, 988 (1979).
(2) G. M. Crippen, *J. Med. Chem.*, **23**, 599 (1980).
(3) G. M. Crippen, *J. Med. Chem.*, **24**, 198 (1981).
(4) G. M. Crippen, unpublished results.
(5) O. Kennard, and D. G. Watson, "Molecular Data and Dimensions", Vol. 1-11, Crystallographic Data Center, Cambridge, England, 1970–1980.
(6) N. L. Allinger, *Adv. Phys. Org. Chem.*, **13**, 1 (1976).
(7) A. Bondi, *J. Phys. Chem.*, **68**, 441 (1964).
(8) G. M. Crippen and T. F. Havel, *Acta Crystallogr., Sect. A*, **34**, 282 (1978).

0022-2623/82/1825-0892$01.25/0



I II

Figure 1. Position labeling and principal rotatable bonds for quinazolines (I) and the naphthyl group (II).

which minimizes the sum of the squares of the differences between calculated and observed binding energies subject to the constraints that the fixed modes are energeticaly more favorable than any other geometrically allowed binding mode.

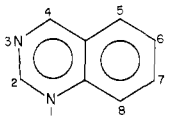
Results and Discussion

We treated the quinazoline data according to the procedure outlined under Methods. **Step 1**: The input data were the binding constants and the structures of the 68 quinazolines given in Table I. These structures were generated from various crystallographic studies as follows: quinazoline,[10] NH_2,[11] SO,[12] S,[13] $C_{10}H_7$,[12] CH_2,[12] COOH,[11] SO_2,[14] CH=CH,[15] glutamic acid,[16] aspartic acid,[17] and CF_3.[18] **Step 2**: The upper and lower bound distance matrix for each molecule was evaluated by rotating the various dihedral angles as shown in Figure 1. When X_2 and X_3 were long chains involving glutamic or aspartic acids attached to a benzene ring, only the dihedral angles between the quinazoline and benzene rings were rotated. In the earlier work[2] we simply measured the upper and lower bounds between atoms and sometimes centers of rings from molecular models. Now we calculate distances between only pairs of atoms from coordinates derived from crystal structures with free rotation around single bonds. The binding data,[19,27] obtained in the form of the concentration of 50% inhibition, were converted[2] to the free energy of binding. **Step 3**: All the hydrogens were deleted, together with most of the atoms of the amino acid residue attached to the benzene rings. **Step 4**: A qualitative consideration of the binding energies of the quinazolines with dihydrofolate reductase suggests that they can be broadly divided into the following three groups: (i) loosely bound "ordinary" quinazolines that are not 2,4-diamino derivatives (compounds **1–35**, except **6**, **10**, **14**, **26**, and **33**), (ii) strongly bound 2,4-diaminoquinazolines (compounds **33** and **36–38**), and (iii) some "odd" 2,4-diaminoquinazolines that are very loosely bound (compounds **6**, **10**, **14**, and **26**). In the earlier QSAR analysis of these 68 quinazolines, Hansch et al.[19] obtained a good overall fit having $r = 0.926$ and $s = 0.672$. These statistics were calculated for 67 molecules, since the activity of one compound (**25**) was mispredicted by more than five times the standard deviation. Also their calculated log $1/C$ values were high for most quinazoline derivatives that were not 2,4-diaminoquinazolines and low for most 2,4-diamino quinazolines. Hansch et al. concluded from their work that (i) position 5 is in a sterically sensitive hydrophobic space and that (ii) the amino groups in the 2 and 4 positions play a very important role and numerous modifications have gained little in activity over that of the parent 2,4-diaminoquinazoline. A logical consequence was to formulate either separate QSARs for ordinary quinazolines and 2,4-diaminoquinazolines or to concentrate the analysis on only the 2,4-diaminoquinazolines, since they have higher biological activity. The separate QSAR approach presumably means two different binding mechanisms, as was modeled earlier[2] using the distance geometry approach. On the other hand, Battershell et al.,[20] with their molecular shape analysis, formulated a QSAR of the 2,4-diaminoquinazolines alone, using the 50% inhibition of rat liver dihydrofolate reductase.

In the distance geometry approach, one constructs a geometry of the receptor site from the drug molecular structure and subsequently evaluates the interaction energy matrix so that the given binding mode for each molecule is its optimal binding mode. In the earlier distance geometry study,[2] the receptor site geometry was deduced in terms of 11 site points. These 11 site points were generated to accommodate two different binding modes of the quinazolines in the same receptor site. It was assumed that the two amino groups of the 2,4-diaminoquinazoline increase the basicity of the ring nitrogens, leading to protonation in solution. Since the type of ring nitrogens is very important for the proposed favorable interaction with site point 1 or, alternatively, site point 2, and since in 2,4-diaminoquinazolines the increased basicity of the ring nitrogens cannot lead to simultaneous protonation of both, it was assumed the N1 should be protonated.[2] Which nitrogen is actually protonated did not affect the ultimate outcome of the study. This protonation ultimately led to a different binding mode of these compounds. The difference amounted to a rotation of the molecule along an axis passing through site points 3 and 4, which bound the second ring at the 6-position and bound the 2-position substituents, respectively. This rotation placed the substituents at positions 4 and 5 in different places, in addition to moving the protonated nitrogen N1. Therefore, three site points were assumed for each of the two rotational states. Three other common site points were assumed, one for the substituents at position 6 and two others for the aromatic substituent. However, some problems resulted from four 2,4-diaminoquinazolines that had unfavorable binding energies. One feature common to these loosely bound 2,4-diaminoquinazolines was that they all had either an SO or SO_2 group as a substitutent of 2,4-diaminoquinazoline. This repulsion returned these molecules to the binding mode of the ordinary quinazolines. Although this trend seemed to be satisfactory, it could not completely explain the binding energies of these odd quinazolines. In summary, the roles of the site points proposed in ref 2 were as follows: point 1 binds N1 of ordinary derivatives; point 2 binds N1 of 2,4-diamino-derivatives, except for the "odd" ones; point 3 binds the second quinazoline ring center; point 4 binds 2-position substituents; point 5 binds 4-position substituents ordinarily; point 6 binds 4-position substituents of 2,4-diamino

(9) J. L. Kuester and J. H. Mize, "Optimization Techniques with Fortran", McGraw-Hill, New York, 1973, pp 106–119.
(10) C. Huiszoon, *Acta Crystallogr, Sect. B*, **32**, 998 (1976).
(11) T. F. Lai and R. E. Marsh, *Acta Crystallogr., Sect. B*, **22**, 885 (1967).
(12) L. Leiserowitz, G. Salem, and M. Weinstein, *Cryst. Struct. Commun.*, **4**, 85 (1975).
(13) S. P. N. Van der Heijden, W. D. Chandler, and B. E. Robertson, *Can. J. Chem.*, **53**, 2102 (1975).
(14) P. B. Rerat, *Acta Crystallogr. Sect. B*, **25**, 1392 (1969).
(15) G. Casalone, A. Gavezzotti, C. Mariani, A. Mugnoli, and M. Simonetta, *Acta Crystallogr., Sect. B*, **26**, 1 (1970).
(16) A. Sequeira, H. Rajagopal, and R. Chidambaram, *Acta Crystallogr., Sect. B*, **28**, 2514 (1972).
(17) S. T. Rao, *Acta Crystallogr., Sect. B*, **29**, 1718 (1973).
(18) D. N. Bol'sutkin, V. M. Gasan, A. I. Prokhvatilov, and A. I. Erenburg, *Acta Crystallogr., Sect. B*, **28**, 3542 (1972).
(19) C. Hansch, J. Y. Fukunaga, P. Y. C. Jow, and J. B. Hynes, *J. Med. Chem.*, **20**, 96 (1977).
(20) C. Battershell, D. Malhotra, and A. J. Hopfinger, *J. Med. Chem.*, **24**, 812 (1981).

Table I. Observed and Calculated Binding Energies of Quinazoline Derivatives to *S. faecium* Dihydrofolate Reductase



no.	group	$-\Delta G_{obsd}$,[a] kcal/mol	$-\Delta G_{cald}$,[b] kcal/mol			
			I	II	III	IV
1	2-H, 4-NH_2, 6-SO_2(2-$C_{10}H_7$)	5.8	6.5	6.7	6.7	6.9
2	2,4-$(SH)_2$, 6-S(2-$C_{10}H_7$)	6.0	6.3	7.0	7.0	6.9
3	2-SH, 4-OH, 6-S(2-$C_{10}H_7$)	6.2	6.3	7.2	7.1	6.9
4	2,4-$(NH_2)_2$, 5-SO_2(2-$C_{10}H_7$)	6.5	8.0	7.7	10.9	7.4
5	2-H, 4-NH_2, 6-S(2-$C_{10}H_7$)	6.5	6.8	6.7	6.3	6.8
6	2-NH_2, 4-OH, 5-CH_3, 6-$NHCH_2$(C_6H_4-4-CO_2H)	6.5	7.9	7.4	7.1	7.4
7	2-OH, 4-SH, 6-S(2-$C_{10}H_7$)	6.8	6.6	6.7	7.0	6.9
8	2,4-$(OH)_2$, 6-S(2-$C_{10}H_7$)	6.9	6.6	6.9	7.0	6.9
9	2-OH, 4-NH_2, 6-S(2-$C_{10}H_7$)	6.9	7.3	7.1	6.6	6.9
10	2,4-$(NH_2)_2$, 5-SO(2-$C_{10}H_7$)	6.9	8.0	7.7	10.9	7.4
11	2-NH_2, 4-OH, 5-CH_3, 6-$NHCH_2$(C_6H_4-4-CO_2Et)	7.1	7.9	7.4	7.1	7.4
12	2-NH_2, 4-OH, 6-$NHCH_2$(C_6H_4-4-CO_2H)	7.2	7.9	7.4	7.1	7.4
13	2-H, 4-NH_2, 6-SO(2-$C_{10}H_7$)	7.2	6.8	7.0	6.8	6.9
14	2,4-$(NH_2)_2$, 5-SO(C_6H_3-3,4-Cl_2)	7.3	8.2	8.2	11.2	7.6
15	2-NH_2, 4-OH, 5-S(2-$C_{10}H_7$)	7.4	8.0	7.9	7.9	7.4
16	2-SH, 4-NH_2, 6-S(2-$C_{10}H_7$)	7.4	7.0	7.4	6.6	6.9
17	2-NH_2, 4-OH, 6-CH_2NH[C_6H_4-4-CONHCH(CO_2Et)$CH_2CH_2CO_2Et$]	7.9	8.9	9.2	9.3	9.0
18	2-NH_2, 4-SH, 6-SO_2(2-$C_{10}H_7$)	8.0	8.9	8.6	9.1	9.2
19	2-NH_2, 4-OH, 6-SO(2-$C_{10}H_7$)	8.2	9.1	9.1	9.2	9.2
20	2-NH_2, 4-OH, 6-SO_2(C_6H_3-3,4-Cl_2)	8.2	9.1	9.3	9.4	9.4
21	2-NH_2, 4-OH, 6-$CH_2N(CH_3)$[C_6H_4-4-CONHCH(CO_2Et)$CH_2CH_2CO_2Et$]	8.3	8.9	9.2	9.3	9.0
22	2-NH_2, 4-OH, 6-$NHCH_2$(C_6H_4-4-CO_2Et)	8.6	7.9	7.4	7.1	7.4
23	2-NH_2, 4-OH, 6-CH_2NH(C_6H_4-4-CO_2H)	8.8	8.9	9.2	9.3	9.0
24	2-NH_2, 4-OH, 6-CH_2NH[C_6H_4-4-CONHCH(CO_2H)$CH_2CH_2CO_2H$]	8.9	8.9	9.2	9.3	9.0
25	2-NH_2, 4-OH, 5-SO_2(2-$C_{10}H_7$)	9.0	9.0	9.0	9.6	8.5
26	2,4-$(NH_2)_2$, 5-SO_2(C_6H_3-3,4-Cl_2)	9.0	8.1	8.2	11.2	7.6
27	2-NH_2, 4-OH, 6-S(C_6H_3-3,4-Cl_2)	9.1	9.4	9.3	9.0	9.3
28	2-NH_2, 4-OH, 5-Cl, 6-$NHCH_2$(C_6H_4-4-CO_2Et)	9.3	9.3	7.4	9.4	8.8
29	2-NH_2, 4-SH, 6-S(2-$C_{10}H_7$)	9.3	9.2	8.6	8.7	9.0
30	2-NH_2, 4-OH, 6-SO_2(2-$C_{10}H_7$)	9.6	8.9	8.8	9.1	9.2
31	2-NH_2, 4-OH, 6-$CH_2N(CH_3)$[C_6H_4-4-CONHCH(CO_2H)$CH_2CH_2CO_2H$]	9.7	8.9	9.2	9.3	9.0
32	2-NH_2, 4-OH, 6-CH_2N(CHO)[C_6H_4-4-CONHCH(CO_2H)$CH_2CH_2CO_2H$]	9.9	8.9	9.2	9.3	9.0
33	2,4-$(NH_2)_2$, 5-S(C_6H_3-3,4-Cl_2)	9.9	11.4	11.3	11.2	11.0
34	2-NH_2, 4-OH, 6-S(2-$C_{10}H_7$)	10.2	9.2	8.8	8.8	9.0
35	2-NH_2, 4-OH, 6-CH_2NH(C_6H_4-4-CO_2Et)	10.6	8.9	9.2	9.3	9.0
36	2,4-$(NH_2)_2$, 5-*trans*-CH=CH(2-$C_{10}H_7$)	10.7	10.1	10.1	10.1	10.3
37	2,4-$(NH_2)_2$, 5-CH_2S(C_6H_4-4-Cl)	10.9	11.7	11.6	11.2	11.1
38	2,4-$(NH_2)_2$, 5-S(2-$C_{10}H_7$)	10.9	11.2	10.8	10.9	10.8
39	2,4-$(NH_2)_2$, 6-$NHCH_2$(C_6H_4-4-CO_2Et)	11.0	11.1	12.5	12.7	11.9
40	2,4-$(NH_2)_2$, 5-*cis*-CH=CH(2-$C_{10}H_7$)	11.1	11.2	10.9	10.9	11.5
41	2,4-$(NH_2)_2$, 6-CH_2NH(C_6H_4-4-CO_2-*n*-Bu)	11.2	12.1	12.1	12.4	12.0
42	2,4-$(NH_2)_2$, 5-CH_2S(2-$C_{10}H_7$)	11.3	11.5	11.1	10.9	10.9
43	2,4-$(NH_2)_2$, 6-CH_2NH(C_6H_4-4-CO_2Et)	11.4	12.1	12.1	12.4	12.0
44	2,4-$(NH_2)_2$, 5-Cl, 6-CH_2NH(C_6H_4-4-CO_2-*n*-Bu)	11.4	12.5	12.2	12.4	12.3
45	2,4-$(NH_2)_2$, 6-CH_2NH[C_6H_4-4-CONHCH(CO_2Et)$CH_2CH_2CO_2Et$]	11.5	12.1	12.1	12.4	12.0
46	2,4-$(NH_2)_2$, 6-CH_2NH[C_6H_4-4-CONHCH(CO_2Et)CH_2CO_2Et]	11.6	12.1	12.1	12.4	12.0
47	2,4-$(NH_2)_2$, 5-CH_3, 6-CH_2NH(C_6H_4-4-CO_2H)	11.8	12.4	12.4	12.4	12.6
48	2,4-$(NH_2)_2$, 5-CH_2CH_2(2-$C_{10}H_7$)	11.9	11.5	11.1	10.9	10.9
49	2,4-$(NH_2)_2$, 6-S(2-$C_{10}H_7$)	12.1	12.4	11.8	11.8	12.0
50	2,4-$(NH_2)_2$, 5-CH_3, 6-CH_2NH(C_6H_4-4-CO_2-*n*-Bu)	12.1	12.4	12.4	12.4	12.6
51	2,4-$(NH_2)_2$, 5-Cl, 6-$NHCH_2$(C_6H_4-4-CO_2Et)	12.2	11.5	12.6	12.7	12.2
52	2,4-$(NH_2)_2$, 6-CH_2NH[C_6H_4-4-CONHCH(CO_2H)CH_2CO_2H]	12.2	12.1	12.1	12.4	12.0
53	2,4-$(NH_2)_2$, 6-S(C_6H_3-3,4-Cl_2)	12.2	12.6	12.2	12.1	12.3
54	2,4-$(NH_2)_2$, 5-CH_3, 6-CH_2NH[C_6H_4-4-CONHCH(CO_2Et)CH_2CO_2Et]	12.2	12.4	12.4	12.4	12.6
55	2,4-$(NH_2)_2$, 5-Cl, 6-CH_2NH(C_6H_4-4-CO_2H)	12.3	12.5	12.2	12.4	12.3
56	2,4-$(NH_2)_2$, 5-CH_3, 6-CH_2NH(C_6H_4-4-CO_2Et)	12.3	12.4	12.4	12.4	12.6
57	2,4-$(NH_2)_2$, 6-SO_2(2-$C_{10}H_7$)	12.4	12.1	11.7	12.2	12.2
58	2,4-$(NH_2)_2$, 5-Cl, 6-CH_2NH[C_6H_4-4-CONHCH(CO_2Et)CH_2CO_2Et]	12.5	12.5	12.2	12.4	12.3
59	2,4-$(NH_2)_2$, 5-Cl, 6-CH_2NH[C_6H_4-4-CONHCH(CO_2H)CH_2CO_2H]	12.6	12.5	12.2	12.4	12.3
60	2,4-$(NH_2)_2$, 5-CH_3, 6-CH_2NH[C_6H_4-4-CONHCH(CO_2H)CH_2CO_2H]	12.7	12.4	12.4	12.4	12.6
61	2,4-$(NH_2)_2$, 5-Cl, 6-CH_2NH(C_6H_4-4-CO_2Et)	12.8	12.5	12.2	12.4	12.3
62	2,4-$(NH_2)_2$, 6-SO(2-$C_{10}H_7$)	12.8	12.3	12.1	12.2	12.2
63	2,4-$(NH_2)_2$, 6-CH_2NH(C_6H_4-4-CO_2H)	13.0	12.1	12.1	12.4	12.0
64	2,4-$(NH_2)_2$, 5-CH_3, 6-$NHCH_2$(C_6H_4-4-CO_2Et)	13.1	11.4	12.8	12.7	12.5
65	2,4-$(NH_2)_2$, 6-CH_2NH[C_6H_4-4-CONHCH(CO_2H)$CH_2CH_2CO_2H$]	13.1	12.1	12.1	12.4	12.0
66	2,4-$(NH_2)_2$, 6-CH_2N(CHO)[C_6H_4-4-CONHCH(CO_2H)$CH_2CH_2CO_2H$]	13.3	12.1	12.1	12.4	12.0
67	2,4-$(NH_2)_2$, 6-S(C_6H_4-3-CF_3)	13.4	13.4	13.4	15.2	13.4
68	2,4-$(NH_2)_2$, 6-SO_2(C_6H_3-3,4-Cl_2)	13.4	12.3	12.2	12.5	12.4

[a] See ref 19 and 27. [b] Free energies of binding calculated for four different versions of the binding site model. See explanation in text under "study I", etc.

Table II. Role of Proposed Site Points to Explain the Binding Energies of the Quinazolines with Dihydrofolate Reductase

point	binds
1	N(1 or 3) unprotonated
2	N(1) protonated
3	C(6)
4	2-position substituent
5	4-position substituent when the quinazoline is not 2,4-diamino
6	4-position substituent when the quinazoline is 2,4-diamino
7	5-position substituent when the quinazoline is not 2,4-diamino
8	5-position substituent when the quinazoline is 2,4-diamino
9	α position of the naphthalene ring (structure II)
10	α' position of the naphthalene ring (structure II) or corresponding substituent of the benzene ring.
11	6-position substituent

derivatives, except for the odd ones; point 7 binds 5-position substituents ordinarily; point 8 binds 5-position substituents of 2,4-diamino derivatives, except for the odd ones; point 9 binds the center of the proximal ring of the naphthalene substituents whether attached to the 5- or 6-position of the quinazoline; point 10 binds the distal half of the naphthalene substituent; and point 11 binds the immediate substituent atom on the 6-position.

Study I. The present work begins with the extra assumption that the protonated nitrogen cannot be attractive to a site point (S_1) that is suitable for a nonprotonated basic nitrogen, and so the odd 2,4-diaminoquinazolines cannot return to the binding mode of the ordinary quinazolines; rather, these molecules will be in a third tilted position in which N1 will come in contact with neither site 2 nor site 1. However, the circle swept out by N1 during the rotation around site points 3 and 4 has a radius of approximately 0.9 Å. This circle is not large enough to keep two nitrogens, N1 and N3, in three different orientations.

Steps 5, 8, and 9. Until now, the amount of rotation about site points 3 and 4 was quite arbitrary. However, to overcome the difficulty mentioned above, it is necessary to assume that the rotation of the 2,4-diaminoquinazolines was 180°. That way, placing the odd 2,4-diaminoquinazolines in a different orientation requires only four different points for the ring nitrogens, but rotation of 180° places N1 very close to the earlier position of N3 and vise versa; therefore, site points 1 and 2 should bind both of these atoms. However, if δ is kept at the reasonably low value of 0.5 Å, it is not possible for site points 1 and 2 to bind both the nitrogens in either of the two rotational states. Since the binding of 2,4-diaminoquinazolines is more important, the binding of N3 with site point 2 was dropped from the proposed binding mode. Also, the most suitable atoms to be bound with the site points 9 and 10 were searched out for molecules like **36** and **40**, which contained an ethylenic group between the quinazoline and naphthalene. For molecule **36**, no suitable atoms were found. It is also geometrically impossible (given δ = 0.5 Å) for site point 11 to bind the 6-substituent while site point 4 binds the 2-substituent when the 2- and 6-substituents are attached simultaneously by either long bonds involving sulfur or short bonds like nitrogen. The site-point roles are listed in Table II.

Step 6. With the geometric constraints derived from these modified binding modes, we determined the site point coordinates presented in Table III.

Table III. Dihydrofolate Reductase Receptor Coordinates

point	coordinates, Å		
	X	*Y*	*Z*
1	3.318	0.039	−1.057
2	3.219	−0.194	1.138
3	−0.679	−1.883	0.490
4	5.100	0.716	0.354
5	0.696	2.157	1.606
6	1.411	0.042	−2.462
7	−0.931	0.404	1.751
8	−0.750	−0.776	−1.759
9	−3.024	0.125	0.123
10	−6.286	1.637	−0.884
11	−2.073	−2.267	0.700

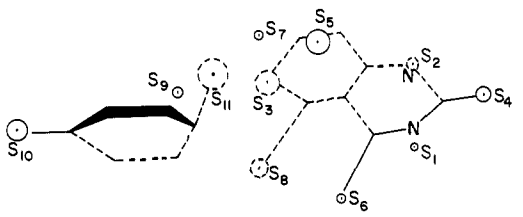


Figure 2. The dihydrofolate binding site geometry. The site points are shown with a strongly bound 2,4-diaminoquinazoline attached. The solid and dotted circles represent above and below the plane, respectively, and their size represents the relative distance from the plane. An "ordinary" quinazoline would bind flipped 180° about the horizontal axis running from S_3 to S_4.

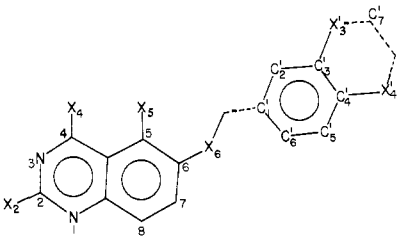
Step 7. These coordinates were found to fit the molecular point distance bounds with a site flexibility (δ) as low as 0.5 Å. Although the relative positions of some of the site points have been changed from our earlier work, their numbering has been kept the same.[2] The pictorial representation of these site points is given in Figure 2.

Step 10. During the energy optimization procedure it was found that the loosely bound "odd" 2,4-diaminoquinazolines could attain a relatively good fit by placing C7 at site point 3, N3 at site point 1, and the amino substituent of position 2 at site point 4, thereby avoiding either the contact of the SO or SO_2 groups with site point 8 or the contact of N1 with site point 1, but still keeping the contact of site points 9 and 10 with the phenyl or naphthyl ring. In this binding mode, their binding energies varied from 8 to 8.2 kcal/mol, while the observed binding energies varied from 6.5 to 9.0 kcal/mol.

The optimized binding modes of the 68 quinazolines are given in Table IV using a general structure (III) for all the molecules. X_i and X'_i represent the atoms directly attached to the indicated ring carbons; e.g., in compound 18, X_2 = N; X_4 = S; X_6 = S (in SO_2); X'_3 and X'_4 are the two α-carbons of the second naphthalene ring. The goodness of fit is given in Table VI under study I.

Examination of the optimized interaction matrix (Table V, study I), obtained from the quadratic programming calculation, suggests that there is a strong attraction of the site point 4 with an amino nitrogen. Furthermore, the strong binding of the 2,4-diaminoquinazolines results from the enhanced attraction of the amino group at position 4 with site point 6, as well as from the simultaneous involvement of both the quinazoline nitrogens with site points 1 and 2, and not just from the attraction of the protonated quinazoline ring nitrogen with the site point 2. The fitting with site point 10 is an important factor for strong binding, but not with site point 9. Since the former site point seems to be very attractive to electronegative atoms, synthetic studies introducing other electronegative

Table IV. Binding of the Quinazolines with the Dihydrofolate Reductase Site Point [a]



III

no.	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	no.	S_1	S_2	S_3	S_4	S_6	S_8	S_9	S_{10}	S_{11}
1	N_1		C_6		X_4				C'_2	X'_4	X_6	**36**	N_3	N_1	C_6	X_2	X_4	X_5			
2	N_1		C_6	X_2	X_4				C'_2	X'_4		**37**	N_3	N_1	C_6	X_2	X_4	X_5	C'_2	X'_4	
3	N_1		C_6	X_2	X_4				C'_2	X'_4		**38**	N_3	N_1	C_6	X_2	X_4	X_5	C'_2	C'_7	
4	N_3		C_7	X_2					C'_6	X'_4		**39**	N_3	N_1	C_6	X_2	X_4		C'_6	X'_4	X_6
5	N_1		C_6		X_4				C'_2	X'_4	X_6	40	N_3	N_1	C_6	X_2	X_4	X_5	C'_2	X'_4	
6	N_1		C_6	X_2	X_4		X_5		C'_2	X'_4		41	N_3	N_1	C_6	X_2	X_4		C'_2	X'_4	X_6
7	N_1		C_6	X_2	X_4				C'_2	X'_4	X_6	42	N_3	N_1	C_6	X_2	X_4	X_5	C'_1	X'_4	
8	N_1		C_6	X_2	X_4				C'_2	X'_4	X_6	43	N_3	N_1	C_6	X_2	X_4		C'_6	X'_4	X_6
9	N_1		C_6	X_2	X_4				C'_2	X'_4	X_6	44	N_3	N_1	C_6	X_2	X_4	X_5	C'_2	X'_4	X_6
10	N_3		C_7	X_2					C'_6	X'_4		45	N_3	N_1	C_6	X_2	X_4		C'_6	X'_4	X_6
11	N_1		C_6	X_2	X_4		X_5		C'_2	X'_4		46	N_3	N_1	C_6	X_2	X_4		C'_6	X'_4	X_6
12	N_1		C_6	X_2	X_4				C'_2	X'_4		47	N_3	N_1	C_6	X_2	X_4	X_5	C'_6	X'_4	X_6
13	N_1		C_6		X_4				C'_2	X'_4	X_6	**48**	N_3	N_1	C_6	X_2	X_4	X_5	C'_2	X'_4	
14	N_3		C_7	X_2					C'_2	X'_4		**49**	N_3	N_1	C_6	X_2	X_4		C'_2	X'_4	X_6
15	N_1		C_6	X_2	X_4		X_5		C'_2	X'_4		**50**	N_3	N_1	C_6	X_2	X_4	X_5	C'_2	X'_4	X_6
16	N_1		C_6	X_2	X_4				C'_2	X'_4		51	N_3	N_1	C_6	X_2	X_4	X_5	C'_6	X'_4	X_6
17	N_1		C_6	X_2	X_4				C'_6	X'_4	X_6	52	N_3	N_1	C_6	X_2	X_4		C'_6	X'_4	X_6
18	N_1		C_6	X_2	X_4				C'_2	X'_4	X_6	**53**	N_3	N_1	C_6	X_2	X_4		C'_2	X'_4	X_6
19	N_1		C_6	X_2	X_4				C'_2	X'_4	X_6	54	N_3	N_1	C_6	X_2	X_4	X_5	C'_6	X'_4	X_6
20	N_1		C_6	X_2	X_4				C'_2	X'_4	X_6	55	N_3	N_1	C_6	X_2	X_4	X_5	C'_6	X'_4	X_6
21	N_1		C_6	X_2	X_4				C'_6	X'_4	X_6	56	N_3	N_1	C_6	X_2	X_4	X_5	C'_6	X'_4	X_6
22	N_1		C_6	X_2	X_4				C'_2	X'_4		57	N_3	N_1	C_6	X_2	X_4		C'_2	X'_4	X_6
23	N_1		C_6	X_2	X_4				C'_6	X'_4	X_6	58	N_3	N_1	C_6	X_2	X_4	X_5	C'_6	X'_4	X_6
24	N_1		C_6	X_2	X_4				C'_6	X'_4	X_6	59	N_3	N_1	C_6	X_2	X_4	X_5	C'_6	X'_4	X_6
25	N_1		C_6	X_2	X_4		X_5		C'_2	X'_4		60	N_3	N_1	C_6	X_2	X_4	X_5	C'_6	X'_4	X_6
26	N_3		C_7	X_2					C'_2	X'_4		61	N_3	N_1	C_6	X_2	X_4	X_5	C'_6	X'_4	X_6
27	N_1		C_6	X_2	X_4				C'_6	X'_4	X_6	62	N_3	N_1	C_6	X_2	X_4		C'_2	X'_4	X_6
28	N_1		C_6	X_2	X_4		X_5		C'_2	X'_4		63	N_3	N_1	C_6	X_2	X_4		C'_6	X'_4	X_6
29	N_1		C_6	X_2	X_4				C'_2	X'_4	X_6	64	N_3	N_1	C_6	X_2	X_4	X_5	C'_6	X'_4	X_6
30	N_1		C_6	X_2	X_4				C'_2	X'_4	X_6	65	N_3	N_1	C_6	X_2	X_4		C'_6	X'_4	X_6
31	N_1		C_6	X_2	X_4				C'_6	X'_4	X_6	66	N_3	N_1	C_6	X_2	X_4		C'_6	X'_4	X_6
32	N_1		C_6	X_2	X_4				C'_6	X'_4	X_6	67 [b]	N_3	N_1	C_6	X_2	X_4		C'_2	X'_3	X_6
33	N_3	N_1	C_6	X_2		X_4		X_5	C'_6	X'_4		68	N_3	N_1	C_6	X_2	X_4		C'_2	X'_4	X_6
34	N_1		C_6	X_2	X_4				C'_2	X'_4	X_6										
35	N_1		C_6	X_2	X_4				C'_6	X'_4	X_6										

[a] X_i and X'_i are the atoms directly attached to the respective carbon atoms. [b] X'_3 for this compound is one fluorine atom.

groups or atoms at the 4-position of the benezene ring are warranted. The involvement of site 9 seems to be almost unnecessary and, therefore, also the presence of the aromatic substituents at the 5- or 6-position. However, we do not recommend the removal of the benzene ring, since any other long aliphatic system will be more flexible, thereby decreasing the probability of the group reaching site point 10. Site point 8 is not very attractive to any of the groups, while site point 7 has strong attraction with electronegative atoms. In 2,4-diaminoquinazolines, that site point may be effective if electronegative atoms are introduced at position 7. The substitution at position 8 to use site 5, however, is not encouraging, since all the interactions here are low. The substitution at position 7 seems to be allowed from the conformational point of view of the substituent at position 6. Since the phenyl ring in a substituent at position 5 can reach site points 9 and 10 without difficulty, any steric hindrance from the substituent at position 7 will not affect the binding of site points 9 and 10.

In the remainder of this work, we explore the effect of removing site points. Of course, the number of ways in which the site points may be removed is enormous. Therefore, we tried deleting only those site points that had low interaction energies with all types of molecular points. Examination of the interaction matrix of the 11 site points (Table V, study I) suggests that the first site point to be removed is site point 9, while site points 8 and 5 are the next most likely candidates. After a site point was removed, the fixed binding modes supplied were still those given in Table IV, except that the column corresponding to the removed site point was deleted.

The comparison of the statistics of these studies, together with those of ref 2, has been sumarized in Table VI. The calculated binding energies of the present studies are given in Table I. The calculated binding energies of the previous calculation may be obtained from ref 2.

Study II. Removal of site point 9 affected the overall fitting of the data only slightly. The correlation coefficient decreased from 0.950 to 0.945, and the standard deviation increased from 0.727 to 0.758. In study I with 11 site points, there were initially 28 adjustable parameters, but during the optimization step, three energetic constraints were generated and eight interaction energies were set to

Table V. Interaction Energy [a] of Dihydrofolate Reductase with the Various Ligand Points in Different Studies

ligand point	ligand type	study	Site Point 1	2	3	4	5	6	7	8	9	10	11
1	C (sp^3)	I							−0.001	−0.305			−1.009
		II							−0.001	−0.251			−1.753
		III							−0.001	-			−2.243
		IV							−0.001	−0.550			−1.598
2	O	I				−0.539	−0.001						
		II				−0.337	−0.197						
		III				−0.219	−0.465						
		IV				−0.112	-						
3	N (basic amino)	I				−3.060	−0.679	−1.857					−0.001
		II				−2.287	−0.410	−1.663					−2.163
		III				−1.959	−0.001	−1.533					−2.554
		IV				−2.285	-	−2.208					−1.487
4	S (S or SH)	I				−1.384	−0.001		−0.001	−0.001			−1.179
		II				−1.568	−0.001		−0.001	−0.001			−0.914
		III				−1.131	−0.449		−0.001	-			−0.881
		IV				−1.790	-		−0.042	−0.430			−1.685
5	Cl	I							−1.380	−0.335		−1.271	
		II							−0.001	−0.121		−1.254	
		III							−2.370	-		−1.110	
		IV							1.366	−0.292		−1.454	
6	F	I										−2.104	
		II										−2.433	
		III										−4.230	
		IV										−2.585	
7	C (sp^2)	I			−2.441					−0.001	−0.001	−1.071	
		II			−2.038					−0.092		−0.781	
		III			−1.858					-		−0.822	
		IV			−1.689					−1.182		−1.240	
8	N (double bonded)	I	−1.416										
		II	−2.600										
		III	−2.776										
		IV	−2.138										
9	N (protonated)	I		−1.350									
		II		−1.469									
		III		−1.997									
		IV		−0.806									
10	S (SO)	I											−1.143
		II											−1.214
		III											−1.301
		IV											−1.802
11	S (SO_2)	I							−1.018				−0.941
		II							−1.099				−0.908
		III							−1.741				−1.252
		IV							−1.134				−1.803
12	C (CO_2H)	I										−1.009	
		II										−0.303	
		III										−0.001	
		IV										−1.311	

[a] All entries are in kcal/mol. Dashes indicate interactions that do not occur in the set of desired binding modes, for a particular study. Those and blank entries are set to a default repulsive value.

Table VI. Necessary Statistics of the Various Studies

study	site points used	no. of constraints [b] (C)	no. of adjustable variables (V)	no. of independent variables $V - C$	correlation coefficients	SD	max error
previous [a]	1-11	11 + 1	20	8	0.847	1.3	4.0
I	1-11	3 + 8	28	17	0.950	0.727	1.7
II	1-8, 10, 11	13 + 5	27	9	0.945	0.758	1.9
III	1-7, 10, 11	9 + 4	23	10	0.874	1.175	4.5
IV	1-4, 6-8, 10, 11	13 + 1	24	10	0.955	0.69	1.6

[a] Reference 2. [b] The first term is the number of explicit linear inequality constraints active at the optimum of the quadratic programming procedure; the second term is the number of energetic variables driven to their upper limit of −0.001 kcal at the optimum.

their limit of −0.001, so there were only 28 − 8 − 3 = 17 independent variables. The most interesting feature of the removal of site point 9 was the change in the number of independent variables. Here there were initially 27 adjustable parameters, but the optimization generated 13 energy constraints and set 5 interactions to their limiting value, leaving only 9 independent variables. Comparison of the present interaction energy matrix (Table V, study II) with the earlier one (Table V, study I), suggests that the higher interaction energies in study I still have high

values in the second study, although their relative magnitudes have been changed. Once the removal of site point 9 was found to have little effect on the fitting problem, we never reintroduced it.

Study III. Removal of site point 8 definitely worsened the overall fit, as judged by the correlation coefficient and standard deviation (Table VI). The reason, however, was quite localized. In our intial hypothesis we assumed that one of the two site points responsible for the low binding energy of four loosely bound 2,4-diaminoquinazolines was site point 8, due to its repulsion with the SO or SO_2 groups at the 5-position of the quinazoline ring. As soon as we removed this site point, there was nothing to keep these molecules from binding in a mode similar to that of the strongly bound 2,4-diaminoquinazolines. Therefore, the calculated binding energies of these four "odd" 2,4-diaminoquinazolines became much more favorable and made the main contributions to the worsening of the correlation coefficient to 0.874 and standard deviation to 1.175 kcal. During this study there were initially 23 adjustable variables; however, in the optimization algorithm, 9 energy constraints were generated, and 4 variables were set to their limiting values, leaving only 10 independent parameters. The interaction matrix obtained in this study (see Table V, study III) generally resembles that obtained from the 11 site points. Once it was realized that the removal of site point 8 was so deleterious, it was restored for all that follows.

Study IV. Removal of site point 5 produced a very interesting result: here the fitting was even better than that with 11 site points, with a correlation coefficient of 0.955 and a standard deviation of 0.69 kcal! Examination of the interaction matrix (see Table V, study IV) suggested that the alternate binding mode of the 2,4-diaminoquinazoline resulted not from any strong attraction of site point 2 to the protonated N1 of the quinazoline ring but from a repulsion between the site point 2 and basic ring nitrogen N1 in the ordinary quinazolines. Otherwise, to gain the favorable interaction between site point 6 and a 4-position basic amino group, a 2-nonamino-4-amino substituted compound should have the binding mode of a 2,4-diamino derivative. If this conclusion is true, a question arises about the nature of this attraction or repulsion to site point 2. The attraction may be from a hydrogen bond involving the protonated ring nitrogen or from a simple electrostatic interaction. If it is from a simple electrostatic interaction, the methylquinazolinium system should also be active. Our suggestion above in study I about introducing halogens at the 7-position of the quinazoline ring to increase its activity remains unchanged by these results.

Conclusions

It is interesting to compare our results with that of Hansch et al.[19] Although their method is apparently easier to apply and uses fewer parameters to fit the data, our approach can lead to novel conclusions that are unobtainable by other methods. In particular, since the data set included no derivatives with 7- or 8-position substitutions, a Hansch analysis can tell nothing about the receptor site in those areas. However, our model of alternate binding modes for the quinazolines suggests that position 7 for a bound 2,4-diaminoquinazoline should be close to site point 7, which is attractive to electronegative groups. Thus, we are able to suggest testing 7-Cl or 7-CF_3 derivatives. Their conclusion that the 5-position of the quinazoline ring is in a hydrophobic region is in at least partial agreement with our result that site point 8 is weakly attractive to saturated carbon and chlorine atoms. However, we do not agree with their conclusion that the 6-position of quinazoline is open to the solvent. That region corresponds to our site point 11, which has strong interactions with several types of groups.

It is not really legitimate to compare our results with those of Battershell et al.[20] because their restricted data set covered only 2,4-diaminoquinazolines, and our picture of the site would have been quite different if we had done the same. Furthermore, they used binding data for dihydrofolate reductase from a different species, and species specificity is well known to be important for this enzyme. For the same reason, we are unable to make any but the vaguest commentary on the resemblance of our site model to the X-ray crystallographic results. Our data come from the *S. faecium* enzyme, and the only available crystal data are for the *L. casei* and *E. coli* enzymes.

One analogue having the maximum error in both study I and study IV is compound **35**. If we focus our attention on the set of compounds **17**, **21**, **23**, **24**, **31**, and **35**, it is clear that their observed binding energies vary from −7.9 to −10.6, while the changes that have been made in these structures are all beyond our site points. Therefore, there is little chance that this model will be able to account for the changes in the binding energy of these compounds. Of course, one way in which a substituent beyond the reach of a site point can affect the binding energy is by restricting the rotation of other groups and thereby preventing these from reaching the required site points. Unfortunately, in these cases the substituents are far from the important rotatable bonds and therefore have no such effect. Consequently, the optimization step could only assign intermediate values for these interactions, which gave larger deviations for the terminal molecules **17** and **35**. Some other sets of compounds having relatively large errors can be explained in the same way. Of course, the problem may be solved by introducing site points for these long substituents. However, we chose to better our understanding of the receptor by focussing on only the most important site points. If we draw a physical picture of the biological phenomena responsible for the dihydrofolate reductase inhibitor, we will admit that the environment of these long substituents in the aqueous phase cannot be completely equivalent to that at the receptor site. That change in environment should have some effect on the binding energies of these molecules even if there is no specific interaction with a site point. The average interaction of these unimportant substituents may be treated with the help of Hansch's π function.

In the last few years, various groups[20-25] have been trying to develop some method to relate structural features of drug molecules to their activity, rather than using exclusively broad physicochemical parameters[26] of the compounds. Our approach differs fundamentally from these in a number of aspects. (i) The methods still generally

(21) E. C. Olson and R. E. Christoffersen, Eds., "Computer Assisted Drug Design", American Chemical Society, Washington, DC, 1979; *ACS Symp. Ser.*, **no. 112** (1979).
(22) Z. Simon, "Quantum Biochemistry and Specific Interactions", Abacus Press, Kent, England, 1976.
(23) A. J. Stuper, W. E. Brugger, and P. C. Jurs, "Computer Assisted Studies of Chemical Structure and Biological Function", Wiley, New York, 1979.
(24) A. J. Hopfinger, *J. Am. Chem. Soc.*, **102**, 7196 (1980).
(25) G. L. Olson, H. C. Cheung, K. D. Morgon, J. F. Blount, L. Tadaro, L. Berger, A. B. Davidson, and E. Boff, *J. Med. Chem.*, **24**, 1026 (1981).
(26) Y. C. Martin, "Quantitative Drug Design", G. L. Grunewald, Ed., Marcel Dekker, New York, 1978.
(27) J. B. Hynes, W. T. Ashton, D. Bryansmith, and J. H. Freisheim, *J. Med. Chem.*, 17, 1023 (1974).

focus on the comparison of chemically similar analogues, where it is clear that a substantial subset of the atoms of one drug molecule match corresponding subsets in the other molecule. Distance geometry, of course, can treat such a situation, but it can also deal with chemically very dissimilar drugs. (ii) There is usually a tacit assumption that all the analogues bind in the same orientation at the receptor site, such that their corresponding atoms always occupy the same positions. In reality, however, drug molecules bind in whatever orientation and internal conformation will minimize the free energy of the drug–receptor–solvent system. The distance geometry calculation directly simulates this search for the most favorable binding mode, and rather similar compounds may bind quite differently. (iii) Other structural methods choose a particular "active conformation" for each analogue and base their relevant geometric and steric parameters on it alone. Our approach more realistically permits a flexible drug molecule to adopt whichever energetically reasonable conformation gives the best calculated binding, given the proposed site. (iv) Most methods presume that differences in binding are (to paraphrase Hopfinger[24] in his discussion of his molecular shape analysis) a smoothly varying function of differences between analogues and, indeed, should be a linear combination of molecular differences. Granted, this is often the case; however distance geometry can also model instances where a small alteraton in chemical structure gives rise to a large difference in activity. (v) Ordinarily, the drug molecules are the focus of attention, and the receptor site is described only secondarily in terms of the environment of bound ligands. Our approach, instead, devotes primary attention to building a tangible model of the site in terms of Cartesian coordinates of the site points and empirically determined contributions to the free energy of binding from the interaction between groups on ligands and site points.

Clearly, each QSAR method has at least some sets of binding data for which it works well. We claim the distance geometry approach will account for the observations on any sort of binding study, although perhaps requiring more computational effort and adjustable parameters than other methods. In addition, we claim our method will give good results on more difficult data sets, where drugs are structurally diverse, where critical steric effects cause large differences in binding for small structural changes, and where different binding modes are implicated.

Acknowledgment. This work was supported by National Institute of Arthritis, Metabolism, and Digestive Diseases, NIH, Grant 1-R01-AM28140-01.

Structure–Antitumor Activity Relationships of 9-Anilinoacridines Using Pattern Recognition

Douglas R. Henry,† Peter C. Jurs,*,† and William A. Denny‡

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, and Cancer Chemotherapy Research Laboratory, University of Auckland, School of Medicine, Auckland, New Zealand.
Received September 28, 1981

A pattern-recognition analysis using the ADAPT system was performed on a set of 9-anilinoacridine antitumor agents, to determine whether computer-generated descriptors could be used to separate active from inactive compounds. A training set of 213 compounds was chosen by random computer selection from a list of 776 structures. Maximal increase in life span at the LD_{10} dosage, a response which is difficult to model using traditional Hansch analysis, was used as the measure of biological activity. A set of 18 molecular descriptors, including fragment, substructure environment, and physicochemical property descriptors (molar refraction, partial electronic charge) was identified which could correctly classify 94% of the compounds in the training set (97% of active and 85% of inactive compounds). Eight of the inactive compounds that were misclassified contained amino substituents, suggesting a role for ionization. The weight vector that was obtained from the training set was applied to a prediction set of 50 compounds that were not included in the original analysis and to a set of 69 structures drawn from the recent literature. The prediction set results, ranging from 73 to 86% correct, were lower than those of the training set, but they clearly indicate that pattern-recognition techniques can be useful in the screening of proposed or already existing agents and especially useful for the identification of active compounds.

Since the turn of the century, derivatives of acridine have been used as therapeutic agents, primarily for the control of malaria (quinacrine) and bacterial infections (proflavine and acriflavine). It has been established that the primary binding site for these compounds in vivo is DNA, by the intercalation mode. It is widely and reasonably assumed that the observed biological effects result from this tight binding, although the detailed mechanism of action remains unknown.[1,2] A similar attachment, leading to the insertion or deletion of bases, has been proposed to explain the mutagenic and carcinogenic potential of acridine compounds.[3] The mutagenic activity may result from stabilization of imperfect pairing caused by single-strand slippages of the DNA.[4,5]

The antitumor potential of acridine derivatives has also been recognized for some time. This activity may result from the fragmentation of DNA, which has recently been shown to occur in tumor cells.[6] Despite much research in this area, from the earliest studies of antitumor, acridines[7,8] until the late 1960's, no definite structure–activity

† The Pennsylvania State University.
‡ University of Auckland.

(1) Adrien Albert, "The Acridines", Edwin Arnold, Ltd., London, 1966.
(2) D J. Patel, in "Drug–DNA Interactions in Solution: Acridine Mutagen, Anthracycline Antitumor, and Peptide Antibiotic Complexes", American Chemical Society, Washington, DC, 1980; *ACS Symp. Ser.*, **no. 142**, 219 (1980).
(3) Albert Lehninger, "Biochemistry", 2nd ed., Worth, New York, 1975, p 880.
(4) John R. Roth, *Annu. Rev. Genet.*, 8, 319 (1974).
(5) J. P. Schreiber and M. P. Daune, *J. Mol. Biol.*, **83**, 487 (1974).
(6) N. B. Furlong, J. Sato, and T. Brown, *Cancer Res.*, **38**, 1329 (1978).

0022-2623/82/1825-0899$01.25/0