Distance Geometry Approach to Rationalizing Binding Data[1]

Gordon M. Crippen

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143.
Received January 12, 1979

A new method is presented for calculating a type of quantitative structure–activity relationship, given experimental data on the binding affinity of a series of ligands to a receptor site on a protein. All ligands are presumed to have known chemical structure but may be conformationally flexible, and all are presumed to bind to the same, single, fairly rigid site on the (pure) receptor protein molecule. Given the experimentally determined free energies of binding of the ligand molecules, possible binding sites are deduced in terms of geometry and the chemical character of the various parts of the site. A test of the method is given for a series of chymotrypsin inhibitors and for a series of dihydrofolate reductase inhibitors. The proposed dihydrofolate reductase site suggests that a quinazoline inhibitor may rock between two different binding modes depending on the pK of the ring N(1).

Deducing quantitative structure–activity relationships (QSAR) for drugs has long been recognized as an important step in rational drug design. Possibly the best known method is that of Hansch[2] and co-workers, where some measure of activity for a series of drugs is empirically correlated with their physical and chemical properties. Ordinarily, little information is obtained concerning the size and shape of the binding site. Recently, Simon et al.[3] have shown a way to calculate the shape of the receptor site from binding data on a series of sterically dissimilar ligands having similar polar groups. The derived picture of the site contains only information on steric accessibility but nothing about the chemical nature of parts of the site, such as hydrogen-bonding regions or hydrophobic pockets. Geometric information has also been obtained using the pharmacophoric pattern search method of Gund et al.[4] A major difficulty with their approach is that the conformation of the drug molecule in the binding site is assumed to be the conformation found in the crystal structure of the pure drug. This is a poor assumption for flexible ligands where the free energy of binding is large in magnitude compared to the differences in free energy between various conformations of the ligand molecule.

The present work is an attempt to overcome some of the shortcomings outlined above. We make the following assumptions: (i) binding is observed to occur on a single site of a pure receptor protein; (ii) each ligand has a well-determined chemical structure and stereochemistry but may be flexible due to rotation about single bonds; (iii) no chemical modification of the ligands occurs during the binding experiment, although the conformation of the ligand may change upon binding to fit the binding site; (iv) the free energy of such a conformational change is small compared to the free energy of binding; (v) the experimentally determined free energy of binding is given and is approximately the sum of the "interaction energies" for all "contacts" between parts of the ligand molecule and parts of the receptor site; (vi) the site itself may be slightly flexible, although no major conformational changes are permitted, and the energetic cost of any deformation is negligible.

The goals of the procedure outlined in this paper are twofold: first, we intend to aid in the deduction of the nature of the binding site from experimental binding data. We avoid a purely empirical correlation of binding affinity to properties of the ligands by producing a physically reasonable picture of the geometry of the site and plausible deductions as to the chemical nature of various parts of the site. The mathematical machinery described in the next sextion and at greater length under the Appendix is to be viewed as a tool for proposing and testing hypotheses about the actual size, shape, and binding interactions of the site. Once some understanding has been gained about the binding site, our second goal ultimately is to use the deductions to predict the binding of molecules outside of the original data set in hopes of proposing better inhibitors or drugs. Not only can one calculate the binding energy of a new ligand, given the deduced site, but the *mode* of binding is also predicted. Thus, although conformationally variable ligands are handled in these calculations, the predicted binding mode can be used to suggest conformationally restricted analogues having an improved binding energy and specificity of binding.

Methods

Because our approach to QSAR is so unlike those of other workers in the field, it will be helpful to first outline the basic concepts. The basis of our method is that each ligand molecule is represented as a collection of points in space. For example, by looking ahead to Figure 1 we see the decomposition of *m*,*m*-$(CH_3)_2C_6H_3OCH_2COCH_3$ into five points. The advantages for this representation will become clear shortly. A very precise description of a molecule in these terms would be to position a point at the nucleus of every atom; however, such precision may not be necessary, and it does lead to more expensive calculations later on. At the other extreme, one may choose to approximate whole groups of atoms, for example, an entire benzene ring, as a single point located, perhaps, at its center of mass. This is an adequate description if the binding of the ligand molecules is rather nonspecific, such as having only a large hemispherical pocket where a benzene ring may be positioned equally well in a variety of orientations. On the other hand, suppose the hydrophobic pocket is shaped like a narrow slot so that the ring may now be inserted in only one orientation. Then it would be necessary to represent the benzene ring as at least three points, perhaps positioned at C(1), C(3), and C(5). In general, larger numbers of points will be required to describe a ligand molecule when the fit in the site is more specific and intricate. With the present algorithms, one discovers that the site is rather specific only by noting that simple descriptions of the ligand molecules fail to account for the binding data.

For some choice of representing the ligands as points corresponding to atoms or groups of atoms, we describe the conformation of a molecule in terms of distances among its points. A table of distances between all possible pairs of points in one ligand will be referred to as a distance matrix. The entry in the ith row of the jth column is the distance between points i and j. If a ligand is conformationally rigid, the distance matrix completely specifies the relative locations of all its points. However in general, most drug molecules can easily undergo internal rotations that change their conformations and, hence, alter the corresponding distance matrix. We handle such flexibility by specifying not just a single distance matrix, but a matrix of lower bounds on the distances and a matrix of upper bounds. For example, if we consider *n*-butane to be four points, one at each carbon atom, then the C(1)–C(4) distance reaches its upper bound in the trans configuration, and the lower bound in the cis. For all other distances, the upper bound equals the lower bound by virtue of fixed bond lengths and angles. Table II compactly shows the upper and lower bound distance matrices for *m*,*m*-$(CH_3)_2C_6H_3OCH_2COCH_3$. Given an atomic model of a ligand molecule, one can rotate about single bonds appropriately and measure off upper and lower bounds on the interatomic distances. With recent advances in distance

0022-2623/79/1822-0988$01.00/0

geometry techniques,[6-8] it is also possible to calculate the coordinates of the points given the upper and lower bound distance matrices. Our method of representing even conformationally variable drug molecules by simply an upper and lower bound distance matrix is an approximation, since sometimes there could be correlated flexibility, as in the case of cyclohexane, where each carbon–carbon distance across the ring has a large range, but one cannot choose a value within that range independently for all pairs. For simplicity in this work, we have chosen to neglect this difficulty.

One similarly proposes a binding site in terms of a number of "site points" whose relative positions are specified by a distance matrix. As in the case of the ligand points, the more detail that is required for the site, the more site points must be chosen. Whereas each ligand point represents some atom or group of atoms, the site points may be called either "empty" or "filled". An empty site point is a vacant place positioned where a ligand point may lie when binding takes place. For example, the simplest sort of hydrophobic pocket would consist of one such empty site point, and in binding, a phenyl group from the ligand molecule might coincide with that point. Similarly, hydrogen bonding or ion-pair sites may be represented by other empty site points. A filled site, however, indicates the position of some steric blocking group, and no ligand point may coincide with it during binding. In contrast to the ligands, the geometry of the site is represented by only a single distance matrix for the site points, instead of upper and lower bound matrices (see, for example, Table III). Thus, the site is assumed to be rather rigid, with only some small variation allowed in each interpoint distance.

Representing both site and ligands as sets of points with conformations given by distance matrices has advantages in the calculation of binding which are crucial to the success of the method. Most importantly, a distance matrix is invariant under translation and rotation, so the elaborate rigid body translation and rotation calculations involved in the usual docking studies (see, for instance, ref 5) are totally avoided. Therefore, a possible binding mode of some ligand amounts to simply a list of which ligand points coincide with which empty site points. Filled site points must be avoided; at most, one ligand point may occupy an empty site point; and the geometry of the ligand points involved in binding must match that of their corresponding site points. Details of the binding calculation are given under the Appendix.

The calculated free energy of binding is obtained in a simplified all-or-nothing fashion by adding up the contribution from each contact between a ligand point and a site point. The individual interaction energy contributions are specified in a proposed energy table, where each row corresponds to a *type*, t_l, of ligand point (e.g., methyl, phenyl, carbonyl) and each column is for a *type*, t_s, of site point (e.g., hydrophobic pocket, hydrogen bond acceptor); see, for example, Table IV. In general, there will be none, one, or several ligand points of a given type in one ligand; the same type of ligand point may appear in more than one molecule; and more than one site point may be of the same type. Each point–point interaction energy is taken to be the ΔG for the process: solvated ligand point + solvated site point → occupied site point. Thus, solvation, enthalpy, and entropy are all included. Although the table of energies is arbitrarily proposed by the investigator, one must be realistic about the choices. A hydrophobic pocket should be attractive ($\Delta G < 0$) to a phenyl group but should not be also strongly attractive to some ionized ligand group. Some filled site point may be repulsive to a bulky *tert*-butyl group while being mildly attractive to a methyl but not the other way around. Careful choice of ligand point types and their corresponding rows in the energy table can be used to represent the influence of a group on another group in the same ligand. For instance, a phenolic hydroxyl should be designated as different from an aliphatic hydroxyl to represent the difference in pK. We will see another example of this in the section on dihydrofolate reductase.

The overall procedure is then as follows: (i) One obtains from experiment the ΔG values of binding of a series of ligand molecules of known chemical structure. The nature of the binding site is otherwise unknown. (ii) Choose the precision of representation of the ligands, and picture each as a (small) number of points. (iii) Calculate or measure from models the upper and lower bound distance matrices for each ligand. (iv) If the number of points in each ligand is quite small, it is possible to automatically search over all possible sites and energy tables for the simplest site that agrees with the observed binding energies. The details of the algorithm for doing this are given under the Appendix. (v) In the usual case of more complicated ligands, one must propose a binding site and then test its accuracy. Generally, if the ligands contain many points, and hence much detail, the site must consist of many points also. The size and shape of the site must be specified by either coordinates of the site points or a distance matrix. Ordinarily one begins by arranging some empty site points to match some common feature of the ligand molecules' structure. Then filled site points are added around the core of the site where needed to force steric repulsion of some of the ligands. In order to ensure reasonable binding, appropriate rough interaction energies are chosen. The computer algorithm described under the Appendix can then be used to objectively calculate a predicted ΔG of binding for each ligand in the data set. This is automatically done by testing each geometrically permissible mode of binding a given ligand to the proposed site and selecting the mode with the lowest (most favorable) calculated binding energy. If the fit to the experimental ΔG values is not satisfactory, the energy table or the number and geometry of the site points may be altered, or both. Poor agreement with experiment may even be due to a choice of ligand points which does not adequately represent a significant feature of their structure. This cut-and-try procedure is clearly subjective, in that the investigator builds his preconceptions into the proposed site. However, the calculated binding energies and modes of binding are produced entirely objectively by a computer program, once the site and energy tables have been selected. It is quite possible that there might be several different, perhaps even simpler, sites that would account for the data equally well, but at least the consequences of the proposed site are objectively tested.

The exhaustive algorithm for deducing the site geometry and the interaction energy matrix and the interactive binding algorithm are both described in full detail under the Appendix. The Results section contains two applications of the method: one a simple series of chymotrypsin inhibitors and the other a lengthier, more difficult series of dihydrofolate reductase inhibitors.

Results

Chymotrypsin Inhibitors. As a simple, yet realistic example, we tested our algorithms on a selected set of eight inhibitors of α-chymotrypsin, according to the data given by Baker and Hurlbut.[9] Their binding data were given in the form of I_{50}, the millimolar concentration of an inhibitor required to produce 50% inhibition, but they can be converted, at least approximately, to ΔG values of binding, assuming Michaelis–Menten kinetics:

$$\Delta G_{\text{bind}} = +RT \ln\left(\frac{K_m[I_{50}]}{K_m+[S]}\right)$$

Here K_m is the Michaelis constant and [S] is the substrate concentration used in the binding assay. The argument to the logarithm is the equilibrium constant for the *dissociation* of enzyme and competitive inhibitor, whereas ΔG_{bind} is the *association* free energy. The compounds are substituted phenoxyacetones, $RC_6H_5OCH_2COCH_3$, except for the first, which is simply phenylacetone. Table I lists the inhibitors, their experimental binding data, and the calculated binding free energies. We arbitrarily chose to represent each ligand in terms of at most five types of points: t_{l1} = $-CH_2COCH_3$ centered on the carbonyl carbon, t_{l2} = $-O-$, t_{l3} = phenyl centered on the middle of the ring, t_{l4} = $-CH_3$ centered on the carbon, and t_{l5} = $-Cl$. For example, Figure 1 shows how inhibitor no. 8 is represented by five points having types t_{l1}, t_{l2}, t_{l3}, t_{l4}, and t_{l4}; the geometry is shown in Table II, where the upper triangle contains the upper distance bounds, and the lower triangle is the lower bounds.

Fitting the first seven inhibitors is remarkably easy, using the exhaustive enumeration of possibilities algorithm

Table I. Binding of Phenoxyacetone Derivatives to α-Chymotrypsin

$RC_6H_4OCH_2COCH_3$

inhibitor no.	R	ΔG_{obsd},[a] kcal	ΔG_{calcd}, kcal
1[b]		−2.5	−2.5
2	H-	−2.8	−2.8
3	p-CH_3-	−2.8	−2.8
4	p-Cl-	−3.5	−3.7
5	m-Cl-	−4.2	−4.6
6	m-CH_3-	−2.8	−2.5
7	p-CH_3O-	−2.6	−2.5
8	m,m-$(CH_3)_2$-	−0.2	−0.2

[a] See ref 9. [b] Phenylacetone.

Table II. Representation of Chymotrypsin Inhibitor 8 as Five Points of Four Types.[a]

	point no.	point type, point no. t_{l1}, 1	t_{l2}, 2	t_{l3}, 3	t_{l4}, 4	t_{l4} 5
$-CH_2COCH_3$	1	0.0	2.6	5.5	7.8	7.8
-O-	2	2.6	0.0	3.0	5.2	5.2
C_6H_3	3	3.5	3.0	0.0	3.1	3.1
$-CH_3$	4	3.8	5.2	3.1	0.0	5.4
$-CH_3$	5	3.8	5.2	3.1	5.4	0.0

[a] The upper triangle (all entries above and to the right of the diagonal line of zeros) is upper bounds on interpoint distances. The lower triangle (below and to the left of the diagonal) is lower bounds. Values are in Angstroms.

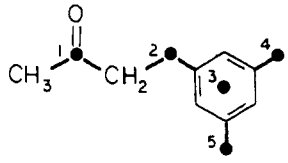


Figure 1. Chymotrypsin inhibitor no. 8, m,m-dimethylphenoxyacetone, decomposed into ligand points. The ligand points are assumed to lie at the heavy dots as indicated. Note that point 1 coincides with the carbonyl C, point 2 with the ether O, and points 4 and 5 with the m-methyls. Point 3 coincides with no atom but rather lies at the center of the benzene ring.

(the first algorithm described under the Appendix). When the permitted site flexibility is 1 Å and the required accuracy of the calculated binding energies is ±1.0 kcal, the first two-site-point solution consists of two points both of one type separated by 2.6 Å. The interaction energy with ligand points types t_{l1}, t_{l4}, and t_{l5} is −3.0 kcal, and with types t_{l2} and t_{l3} is −0.5 kcal. The observed binding energies are accounted for, up to some reasonable accuracy, but otherwise this result is an example of an empirical fit which is not physically very meaningful. Accounting for the rather unfavorable binding energy of inhibitor no. 8 requires many more site points, as we shall see, and the exhaustive approach becomes infeasible. However, that expanded site picture is easier to interpret as a description of the real binding site.

Comparing inhibitors 2, 6, and 8, it is clear that the binding site can accommodate one m-methyl substituent on the phenyl ring, but a second m-methyl is apparently sterically disallowed. In a situation like this, the use of forced contacts is crucial. Briefly, we mean that, under certain arrangements of site points, formation of certain triples, pairs, or individual contacts can force by triangulation the formation of additional contacts that may even be energetically unfavorable. The logic involved is explained under the Appendix. Referring to Table III and

Table III. Final Proposed Chymotrypsin Binding Site, Consisting of Eight Points of Five Types.[a]

point no.	point type, point no. t_{s1}, 1	t_{s2}, 2	t_{s3}, 3	t_{s4}, 4	t_{s5}, 5	t_{s5}, 6	t_{s5}, 7	t_{s5}, 8
1	0.0	2.6	5.5	7.8	7.8	7.8	7.8	11.0
2	2.6	0.0	3.0	5.2	5.2	5.2	5.2	6.0
3	5.5	3.0	0.0	3.1	3.1	3.1	3.1	3.1
4	7.8	5.2	3.1	0.0	5.4	3.8	3.8	3.0
5	7.8	5.2	3.1	5.4	0.0	3.8	3.8	3.0
6	7.8	5.2	3.1	3.8	3.8	0.0	5.4	3.0
7	7.8	5.2	3.1	3.8	3.8	5.4	0.0	3.0
8	11.0	6.0	3.1	3.0	3.0	3.0	3.0	0.0

[a] The symmetric matrix of interpoint distances in Angstroms is shown, rather than upper and lower distance bounds.

Table IV. Final Proposed Chymotrypsin Interaction Energy Table[a] for the Five Ligand Point Types and the Five Site Point Types

ligand point types	site point types t_{s1}	t_{s2}	t_{s3}	t_{s4}	t_{s5}
t_{l1}	−0.01	0.1	10.0	1.0	1.0
t_{l2}	1.0	−0.2	10.0	1.0	0.05
t_{l3}	1.0	1.0	−2.6	1.0	1.0
t_{l4}	1.0	1.0	−0.1	0.15	10.0
t_{l5}	1.0	1.0	10.0	−1.05	1.0

[a] In kcal.

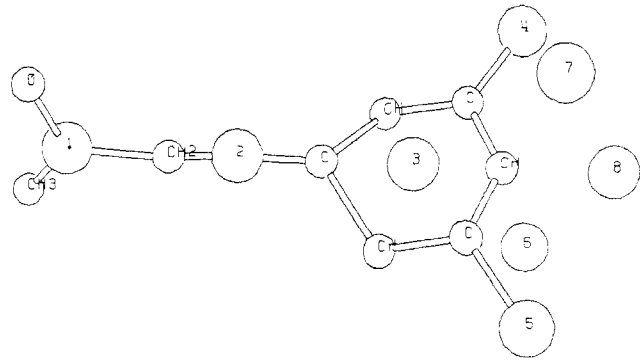


Figure 2. Proposed chymotrypsin binding-site geometry with inhibitor no. 8 in place in a sterically disallowed fashion. Large spheres are the locations of site points, numbered as in Table III, and the small spheres are the nonhydrogen atoms of the ligand connected by bonds. Note that site point 1 coincides with the carbonyl carbon, as does site point 2 with the ether oxygen, no. 4 with one methyl, and no. 5 with the other methyl group.

Figure 2, if we propose a binding site consisting only of site points 1 through 3, there are energetically favorable and geometrically allowed sites for the acetone group, the ether oxygen, and the phenyl ring but no binding site for the first methyl. This is provided by site point no. 4, so that ligand 6 binds slightly more firmly than ligand 2, but ligand 8 still binds equally well. Simply adding repulsive site point no. 5 (see interaction energy, Table IV) does not solve the problem, because, although proposing contacts with ligand points 2–4 of inhibitor no. 8 forces ligand point 5 to come into contact with site point 5 due to a triple point forced contact (see the Appendix for explanation), the binding algorithm instead prefers to *not* propose a contact with methyl group 4, thus allowing the phenyl ring to rotate out of the way of the unfavorable site point 5, resulting in a much too favorable binding energy. Adding repulsive site points 6 and 7 creates a four-point ring for the m-methyls to contend with, so that even proposing favorable contacts with ligand points 2 and 3 forces one

methyl in contact with site point 4 (the energetically most favorable possibility) as a double point forced contact, and then the unfavorable methyl contact results as a triple point forced contact. However, the energetically most favorable binding mode in this case now involves contacts between site point 3 and ligand point 3, but none with ligand point 2, thus allowing the phenyl ring to tip up out of the plane of site points 3–5, and avoiding the repulsive sites for the methyls. This is at last countered by adding in site point 8, so that there are now six site points (2 and 4–8) which are 3 Å from the site for the phenyl. Thus, whenever a contact between site point 3 and ligand point 3 is proposed, there must be a single point forced contact with ligand point 2 (the energetically most favorable one being with site point 2), which in turn requires contacts with ligand points 4 and 5. The energetically optimal binding mode for inhibitor 8 is now ligand point 1 in contact with site point 2 and ligand point 4 in contact with site point 4, which results in a sufficiently poor calculated binding energy.

In fact, the final proposed binding site geometry displayed in Table III and approximately the energies given in Table IV were arrived at using the interactive approach, in this very sequence of events. However, the fit of the calculated binding energies to the observed ones is not optimal at this point. Given the calculated modes of binding, only some of the entries of the interaction matrix directly contribute to the calculated binding energies, and the adjustment of these used entries is simply a linear least-squares refinement, since the calculated binding energy is the sum of the contribution from each contact. In this case there are seven parameters to be adjusted, namely, all those entries in Table IV which are neither 1.0 or 10.0. It turns out that there are only six linearly independent parameters in the problem, so the 1,1 entry is arbitrarily fixed at the barely favorable level of −0.01 kcal. The other six variable entries then follow as shown in Table IV, and the resultant root mean square deviation between calculated and observed binding energies (see Table I) is only 0.19 kcal. The energy refinement in this example is particularly easy to deal with, because substituting the least-squares values for the rough energies of interaction does not change the calculated preferred binding modes. The general problem is much more subtle in that one must minimize the sum over the ligands of the squares of the deviations between calculated and observed energies for the energetically optimal (but still geometrically allowed) binding modes, which in turn depend on the interaction energies. Thus, the general problem must be a sort of constrained nonlinear optimization, for which there exist only locally convergent algorithms, rather than being possible to calculate the unique global solution in the linear least-squares case. In other words, we are not aware of any general procedure to guarantee that the best possible site has been found.

Figure 2 shows the geometry of the proposed binding site, with site points shown as large spheres and labeled as in Table III. Inhibitor no. 8 is displayed as small spheres connected by rods. The positioning of the ligand is a sterically disallowed one, where the two phenyl ring methyls are in contact with site points 4 and 5. Comparing the illustration with the energies in Table IV, it is clear that site point no. 1 very weakly binds the acetone group of the inhibitors, while being at least mildly repulsive to the other sorts of ligand points, perhaps because point no. 1 is a hydrogen-bond donor to the carbonyl oxygen. Point no. 2 can be interpreted as a polar pocket large enough to accommodate the ether linkage but is unfavorable as a binding site for the acetone group, perhaps for steric reasons. Site point no. 3 is a strongly hydrophobic region for binding the phenyl group. It constitutes the center of a structured pocket surrounded by sterically repulsive regions (no.'s 5–8), one small pocket for a methyl group (no. 4) and of course the ether oxygen binding site (no. 2). Site point 4, being of site type 4, is apparently energetically attractive to a chloro substituent (ligand point type 5) on the ring, while being faintly repulsive to a methyl group, although it will accommodate one. The above site point descriptions are summarized in Table V. As detailed as this site description is, it is difficult to tell at present just how correct it is. Although the high-resolution X-ray crystal structure of α-chymotrypsin is known, there has been no X-ray study with phenoxyacetone inhibitors bound to the enzyme. Tulinsky and co-workers have examined the binding of toluenesulfonamide, pipsylamide, and phenylethaneboronic acid,[10] while Steitz et al.[11] have reported the binding geometry for formyl-L-tryptophan, formyl-L-phenylalanine, dioxane, β-(*p*-iodophenyl)-propionate, and related compounds. Unfortunately, the conclusion is that, although there is a generally well-defined region that binds the aromatic ring, the orientation of the ligand depends greatly upon the nature of the substituents. Thus, without knowing experimentally where the acetone moiety binds, we cannot tell whether the rest of our proposed active site is consistent with the X-ray data. We have at least been able to fit our proposed site and ligand molecule into the chymotrypsin X-ray coordinates in a reasonable fashion, as shown in Figure 3. The view of the active site is essentially the same as in Figure 1 of ref 11, and the positioning of the inhibitor is similar to the experimentally observed location of formyl-L-tryptophan according to ref 11. Table V lists the parts of chymotrypsin corresponding to our proposed site points when the site is placed in the active site in this way. The point of Figure 3 is not so much to conclusively verify the correctness of the proposed site points but merely to show that they are not inconsistent with the best experimental evidence available.

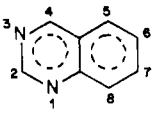
In spite of the complexity of the final result, it was relatively easy to achieve. Of course, there is no guarantee that the proposed site is the simplest one to account for the given data, but one could now attempt to simplify it. There is, furthermore, no guarantee that the proposed site will account for the binding energies of other ligands outside of the data set, although at least any molecule made up solely of the groups we have chosen as ligand points can have its binding energy predicted. Indeed, we have already seen how the addition of ligand 8 to the data set required extensive revision of the site points. At least

Table V. Description of Chymotrypsin Site Points, Including for Each Its Probable Chemical Nature, What It May Coincide with in the X-Ray Crystal Structure of Chymotrypsin,[a] and the Distance from that Protein Atom to the Corresponding Site Point

site point no.	chem type	located near chymotrypsin atom	distance, Å
1	H-bond donor	Ser-195 O^{γ}	2.68[b]
2	polar		
3	nonpolar pocket	several active site residues	
4	weakly polar	open to solvent	
5	steric blocking	Ser-214 O	1.24
6	steric blocking	Gly-216 N	1.36
7	steric blocking	Val-213 C^{γ}	1.60
8	steric blocking	interior of protein	

[a] See Figure 3. [b] To ligand carbonyl O.

Table VI. Binding of Quinazoline Derivatives to *S. faecium* Dihydrofolate Reductase



no.	R_2	R_4	R_5	R_6	ΔG_{obsd},[a] kcal	ΔG_{calcd}, kcal
1	H	NH_2	H	SO_2-2-$C_{10}H_7$	−5.8	−7.7
2	SH	SH	H	S-2-$C_{10}H_7$	−6.0	−6.2
3	SH	OH	H	S-2-$C_{10}H_7$	−6.2	−6.5
4	NH_2	NH_2	SO_2-2-$C_{10}H_7$	H	−6.5	−7.0
5	H	NH_2	H	S-2-$C_{10}H_7$	−6.5	−7.4
6	OH	SH	H	S-2-$C_{10}H_7$	−6.8	−6.5
7	OH	OH	H	S-2-$C_{10}H_7$	−6.9	−6.8
8	OH	NH_2	H	S-2-$C_{10}H_7$	−6.9	−8.0
9	NH_2	NH_2	SO-2-$C_{10}H_7$	H	−6.9	−6.7
10	H	NH_2	H	SO-2-$C_{10}H_7$	−7.2	−7.3
11	NH_2	OH	S-2-$C_{10}H_7$	H	−7.4	−8.0
12	SH	NH_2	H	S-2-$C_{10}H_7$	−7.4	−7.7
13	NH_2	SH	H	SO_2-2-$C_{10}H_7$	−8.0	−8.0
14	NH_2	OH	H	SO-2-$C_{10}H_7$	−8.2	−7.9
15	NH_2	OH	SO_2-2-$C_{10}H_7$	H	−9.0	−8.3
16	NH_2	SH	H	S-2-$C_{10}H_7$	−9.3	−7.7
17	NH_2	OH	H	SO_2-2-$C_{10}H_7$	−9.6	−8.3
18	NH_2	OH	H	S-2-$C_{10}H_7$	−10.2	−8.0
19	NH_2	NH_2	S-2-$C_{10}H_7$	H	−10.9	−11.6
20	NH_2	NH_2	H	S-2-$C_{10}H_7$	−12.1	−11.6
21	NH_2	NH_2	H	SO_2-2-$C_{10}H_7$	−12.4	−12.0
22	NH_2	NH_2	H	SO-2-$C_{10}H_7$	−12.8	−11.6

[a] See ref 12.

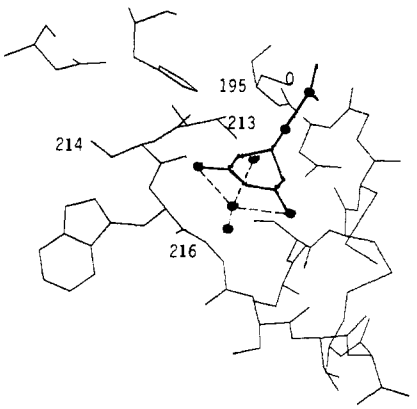


Figure 3. A possible positioning of inhibitor no. 8 in the active site of α-chymotrypsin. The residues of the active site are indicated by light lines, and certain important residue sequence numbers are given. See Table V and text for explanation. The ligand molecule is drawn in heavy lines, while the site points are shown as heavy dots. The dashed lines connecting some of the site points are meant only to convey a sense of depth and have no physical significance. The ligand lies thrust into the active-site cleft with its carbonyl group in the foreground near the side-chain O (marked) of Ser-195. Site point 5 is in the plane of the ligand benzene ring to the left, and site point 4 is to the right. Site points 6 and 7 are found above and below the plane of the ring, and point 8 is deep in the background in the plane.

the result appears to be in accordance with the X-ray diffraction evidence, and our hypothesized site could be directly verified or disproven by an X-ray crystal study of chymotrypsin with one of the eight ligands of Table I bound to it.

Dihydrofolate Reductase. As a test of the method on a larger, more complex data set, we next tried a series of quinazoline inhibitors of *S. faecium* dihydrofolate reductase. Hansch et al.[2b] have formulated a quantitative structure–activity relationship for 68 such derivatives, achieving a root mean square fit of the observed free energies of binding of only 1.05 kcal. We chose to work with a limited subset of these compounds on the grounds that some of the 68 derivatives are so complex that more computer time would be required than is warranted for this preliminary study. The 22 compounds we chose, given in Table VI, are all 2-, 4-, 5-, and 6-substituted quinazolines, where either the 5 or 6 position involves some sort of sulfur linkage to the 2 position of naphthalene. The experimental data[12] were once again given in terms of I_{50}, but these were converted to the ΔG_{obsd} values given in Table VI according to the usual equation, using the K_m value given in ref 13.

Since these 22 compounds are relatively complex, the exhaustive approach is out of the question, and it is instead necessary to propose an active site using the interactive method. To aid the inspection of the experimental data, we systematically noted the difference in binding energies for every pair of compounds which differed by only a single feature. It soon became apparent that generally 2,4-diamino derivatives bind much better than those where one or more of these two positions are occupied by any other type of group. As has been shown experimentally for folate and methotrexate,[15-17] we assumed the effect is due to a shift of the pK_a of the ring nitrogens for the 2,4-diamino derivatives. In accordance with the quantum mechanical calculations of Perault and Pullman,[18] we modeled this by declaring the nitrogen in the 1 position to be of a different type when the 2 and 4 positions are amino substituted. However, this does not explain away all the differences in binding energies. Although ligand **19**, being diamino substituted, binds very well, changing the 5-thioether linkage to sulfoxide (**9**) or sulfone (**4**) results in anomalously poor binding. The diamino effect is, nevertheless, in force for these 5-position linkages, as can be seen by comparing ligands **11** and **19**. On the other hand, when the sulfur linkages are on the 6 position, as in ligands **20–22**, there is not a large difference in binding among the three types of linkages. Not surprisingly, Hansch et al.[2b] also had trouble with this point, mispredicting the binding of ligand **15** by 4.82 kcal. Their explanation was that apparently a different mode of binding was involved, but taking such

Table VII. Final Proposed Dihydrofolate Reductase Binding Site Showing Site Point Numbering, Types, and Distance Matrix[a]

point no.	point type, point no.										
	t_{s1}, 1	t_{s2}, 2	t_{s3}, 3	t_{s4}, 4	t_{s4}, 5	t_{s4}, 6	t_{s5}, 7	t_{s6}, 8	t_{s3}, 9	t_{s3}, 10	t_{s5}, 11
1	0.0	1.5	2.9	2.6	4.3	3.2	5.1	4.5	7.5	9.0	5.9
2	1.5	0.0	2.9	2.6	3.2	4.3	4.5	5.1	7.5	9.0	5.9
3	2.9	2.9	0.0	5.2	3.8	3.8	3.2	3.2	5.5	7.6	3.2
4	2.6	2.6	5.2	0.0	5.1	5.1	6.8	6.8	9.9	10.0	8.3
5	4.3	3.2	3.8	5.1	0.0	4.0	2.4	4.2	5.0	6.0	5.2
6	3.2	4.3	3.8	5.1	4.0	0.0	4.2	2.4	5.0	6.0	5.2
7	5.1	4.5	3.2	6.8	2.4	4.2	0.0	4.0	3.4	5.6	2.9
8	4.5	5.1	3.2	6.8	4.2	2.4	4.0	0.0	3.4	5.6	2.9
9	7.5	7.5	5.5	9.9	5.0	5.0	3.4	3.4	0.0	2.5	3.5
10	9.0	9.0	7.6	10.0	6.0	6.0	5.6	5.6	2.5	0.0	5.6
11	5.9	5.9	3.2	8.3	5.2	5.2	2.9	2.9	3.5	5.6	0.0

[a] In Angstroms.

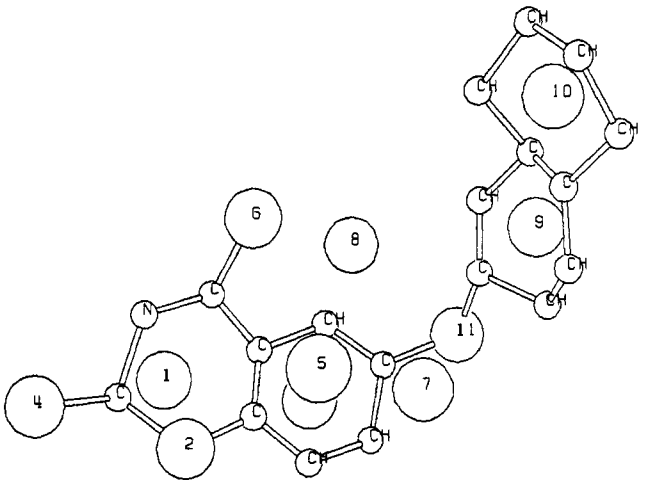


Figure 4. Proposed dihydrofolate reductase binding-site geometry with inhibitor no. 22 bound. Large spheres are the locations of site points, numbered as in Table VII, and the small spheres are the nonhydrogen atoms of the ligand connected by bonds. Site point no. 3 is in the plane of the quinazoline ring behind site point no. 5. Number 1 lies behind the plane, while no. 7 is in front. Number 2 is coincident with the N-1, no. 4 with the 2-amino group, no. 6 with the 4-amino group, and no. 11 with the SO linkage.

things into account is much more difficult in the Hansch approach than with ours. At least one solution is to propose a "rocking" site, whereby the quinazoline ring pivots about an axis running through the 2 and the 6 positions. At one place in the range of motion allowed, the nitrogen-1 in the ring would bind at a site point favorable to an unprotonated nitrogen, while at the other extreme of the swing, there would be a site to accept a protonated nitrogen-1. Since the quinazoline ring system is rigid, there are two corresponding sets of sites for 4 and 5 substituents. The 2 and 6 substituents, however, always bind in the same place, because they lie along the axis of rotation. Our final proposed site then, as shown in Figure 4 and Table VII, consists of 11 points: no. 1 binds the N-1 for all but 2,4-diamino derivatives, no. 2 binds the N-1 in the case of 2,4-diamino substitution, no. 3 provides a hydrophobic site for the center of the second (4–7 positions) quinazoline ring, no. 4 binds the 2 substituents, no. 5 is for the 4 substituents when site point no. 1 is being used, no. 6 is the alternative 4-group site in the case of 2,4-diamino derivatives, no. 7 is the 5-position binding site corresponding to the use of sites no.'s 1 and 5, no. 8 binds a 5-S– (but not the bulkier 5-SO– or 5-SO_2–) in the 2,4-diamino case, no. 9 is for the first (1–4 positions) naphthalene ring, no. 10 binds the other naphthalene ring, and no. 11 is the

Table VIII. Final Proposed Dihydrofolate Reductase Interaction Energy Table[a] for the Eight Ligand Point Types and the Six Site Point Types

ligand point types	site point types					
	t_{s1}	t_{s2}	t_{s3}	t_{s4}	t_{s5}	t_{s6}
t_{l1}	−1.0	10.0	0.1	0.1	0.1	0.1
t_{l2}	0.1	0.1	−1.5	0.1	0.1	0.1
t_{l3}	0.1	0.1	0.1	−1.8	0.1	0.1
t_{l4}	0.1	0.1	0.1	−0.3	−0.1	−0.0
t_{l5}	0.1	0.1	0.1	0.1	−0.0	10.0
t_{l6}	0.1	0.1	0.1	0.1	−0.4	10.0
t_{l7}	0.1	0.1	0.1	−0.6	0.1	0.1
t_{l8}	1.5	−3.4	0.1	0.1	0.1	0.1

[a] In kcal.

site for all 6 substituents. The 22 ligands were described in terms of eight types of points: no. 1 is the ordinary type of N-1, no. 2 is for benzene rings (the nonheterocycle in quinazoline, the two naphthalene rings, etc.), no. 3 is $-NH_2$ in any position, no. 4 represents either –S– or –SH, no. 5 is an –SO– linkage, no. 6 an $-SO_2-$ linkage, no. 7 is an –OH, and no. 8 represents the N-1 when 2,4-diamino substituted. Proceeding as before, the geometries of the 22 ligand molecules were represented by upper and lower bound distance matrices for the interpoint distances, as measured from molecular models. Having already in mind the binding scheme outlined above, the intersite point distances given in Table VII were chosen to match corresponding interligand point distances.

Determination of appropriate interaction energies was somewhat more difficult. According to Table VII, six site point types were chosen: 1 and 2 for the two sorts of N-1, 3 for binding benzene rings in various places, type 4 for binding the small substituents in the 2 and 4 positions, 5 for the sterically permissive 5 and 6 positions, and 6 for the sterically restrictive 5 position when in the binding mode for 2,4-diamino derivatives. Because of the small distance differences between the intended alignment of the ligands in the site and alternative alignments, it was necessary to reduce the allowable distance error to 0.5 Å. Then interaction energies were roughly chosen in hopes of achieving the intended binding modes, simply by entering small negative values for desired interactions, +10 kcal for sterically disallowed interactions and a default +0.1 for all other entries. The default value is mildly unfavorable, yet it is small enough in magnitude that such contacts will be made if geometrically required by other attractive contacts. Least-squares refinement of the interaction energies was more difficult than for chymotrypsin, because slight alterations in energy values brought about undesired modes of binding. Eventually, we arrived

at the values given in Table VIII, which along with the geometry of Table VII produce the ΔG_{calcd} entries in Table VI. As was originally intended, the 2,4-diamino substituted ligands have their slightly different mode of binding, the steric constraints on the 5 position permitting. The root mean square deviation between the calculated and observed binding energies for the 22 inhibitors is 0.99 kcal, which is probably comparable to the experimental error.

It is interesting to compare the proposed binding site with the only available X-ray crystal study, that of Matthews et al.,[14] on the methotrexate complex of an *Escherichia coli* dihydrofolate reductase. One should keep in mind that these enzymes tend to be species specific, but at least it is known that methotrexate is also a good inhibitor of the *S. faecium* reductase considered in our study. Furthermore, methotrexate contains a pteridine ring system instead of the quinazolines we have investigated, but the first ring in either case involves an N-1 and N-3 heterocycle, which is 2,4-diamino substituted in methotrexate while having 2-amino, 4-hydroxyl substitution in dihydrofolate, the natural substrate. Just as in the quinazolines, methotrexate binds much more strongly than dihydrofolate, apparently due to the 2,4-diamino effect. Matthews et al. suggest that the N-1 is protonated in methotrexate and show that there is a strong interaction between that atom and the side chain of Asp-27. All this is in good agreement with our hypothesized special type for the N-1 when 2,4-diamino substituted, and Asp-27 may be thought of as being responsible for our site point no. 2. They further find that there is a hydrogen bond between the 2-amino as donor and the side-chain hydroxyl of Thr-113 as acceptor. Thus, the empty space near this hydroxyl corresponds to our site point no. 4. Similarly, the hydrogen bond observed in the crystal structure between the 4-amino and the carbonyl oxygen of Ile-5 corresponds to our site point no. 6. Our hydrophobic pocket for the rest of the quinazoline ring, site point no. 3, may be thought of as being indicated by the close proximity in the X-ray results of the pteridine ring and side chains of Ile-5, Ala-7, Leu-28, Phe-31, and Ile-94. The crystal structure even indicates a second hydrophobic pocket for the aromatic ring of the *p*-aminobenzoyl portion of methotrexate in a position corresponding to our naphthalene binding sites, points no.'s 9 and 10. Unfortunately, the hypothesized alternate tilted binding mode for analogues without 2,4-diamino substitution cannot be verified without a crystal study on such an inhibitor–enzyme complex. Thus site points no.'s 1, 5, and 7 remain hypothetical at this time. Indeed, it is quite possible that there is some other alternate binding geometry which accounts for the inhibition data, but there is nothing in the set of 22 compounds that compels us to seek it.

Discussion

We have outlined two approaches to rationalizing ligand binding data: the first is a thorough search of all possible site descriptions that results in the simplest picture of the site consistent with the data; the second is an interactive method which results in some sort of adequate site description, although it may be more complicated than necessary. Since we have found that the exhaustive algorithm is practical only for particularly simple data, we will consider only the interactive approach in this section.

As we have seen in the previous section, it is possible to account for the experimentally determined binding energies of a few inhibitors of chymotrypsin and several inhibitors of dihydrofolate reductase up to a reasonable estimate of the experimental error. It was not necessary to assume that the ligands within each set were chemically similar, although in these test cases they were. The computations work equally well, if not better, if the ligands were conformationally more restricted, but any flexibility is easily handled. The resultant description of the site consists of some points representing energetically attractive pockets of either a hydrophobic or polar nature and some points corresponding to the location of steric blocking groups, repulsive to any part of a ligand molecule. Thus, the site involves both energetic and steric features, as well as their relative positions. The result is not just an empirical restatement of the input data but rather can be used to predict the binding energy of any molecule built up out of the same types of atomic groups. Hence, from a set of binding data one could deduce the site and then predict what sort of drug molecules would be most likely to bind even better. Of course, any later discrepancies between predictions and observations would require an alteration of the site. There is the added feature that not only is the binding energy predicted but also the *mode* of binding. Thus, one could suggest conformationally restricted analogues having presumably high binding specificity for the protein site in question, just from noting the predicted mode of binding of more flexible molecules.

The interactive algorithm is quite feasible. Calculation of the binding of the eight ligands to the final 8-point site of chymotrypsin takes about 4 min on a PDP 11/70 computer, with the program written in fast Commercial Union Leasing Corporation's fortran-4-plus. Computer time goes up in a complicated fashion, although not necessarily exponentially, when more ligand point and more site points are involved. Computing the optimal binding mode for all 68 dihydrofolate reductase inhibitors considered by Hansch et al.[2b] required only 24.7 s on the Lawrence Berkeley Laboratory's CDC 7600 at a cost of $3.86, even though there were 11 site points and a number of the ligands contained 16 ligand points. The outlook is poor for handling large problems by the exhaustive enumeration algorithm, however. Deducing a two-point site for the first seven chymotrypsin inhibitors cost only $1, whereas considering simple four-point sites for all eight inhibitors cost in excess of $30, and expenses were rising exponentially with the number of site points.

Certainly, devising a suitable site requires some imagination on the part of the user, but, as we have shown in the previous section, one can follow an easy build-up principle to arrive at a rather complicated site. We intend to develop more automatic procedures for proposing new sites based on the method mentioned above for the quinazolines, involving comparisons of the binding energies and chemical structures for pairs of very similar inhibitors. Although the resulting site may not be the simplest possible, at least the binding calculation prevents the user from building in any preconceptions as to *how* the ligands bind to the site. Indeed, the ligands give the impression of being remarkably slippery.

It is desirable to compare the present method with the Hansch approach. In general, both are to some degree an empirical correlation of the chemical structure of inhibitors to their binding energy. Whereas the Hansch method could just as well correlate structure to very complicated experimental observations, such as in vivo assays of drug effectiveness, we have constrained ourselves to the much simpler physical chemical problem of accounting for observed free energies of binding to a single site on a single receptor. As we have shown, our method does not simply attribute binding to a sum of factors but gives a direct spatial interpretation of steric factors and alternate binding modes. Our proposed binding sites, although not nec-

essarily unique, at least contain much more geometric detail than is, in principle, directly testable. As a more precise comparison, Hansch et al.[2b] examined the 22 quinazoline inhibitors of dihydrofolate reductase that we did but included in their data set an additional 46 related compounds. We did not specifically fit the whole data set for two reasons: the remaining compounds involve as many as 16 ligand points apiece, compared to the maximum of 8 per inhibitor in our restricted data set, and incorporating these larger ligands would have required hypothesizing a considerably larger site, which would have required much more computer time. We estimate that the calculation would still have been feasible but probably not of sufficient clinical interest, since the data are not for the human enzyme. Secondly, it became clear that further development of systematic ways for proposing sites is required for handling large data sets, so we consider it worthwhile to concentrate our efforts on methods rather than extensive "cut and try" applications. Unfortunately, the result is that a fair comparison between Hansch's work and our own is difficult. They fit three times as many compounds to a root mean square deviation of observed and calculated binding energies of 1.05 kcal, their worst error being 4.82 kcal for compound 15 (25 in the numbering of ref 2b), using only six parameters. In comparison, we fit the smaller data set to a root mean square error of 0.99 kcal, with a worst error of 2.2 kcal for compound 18, using considerably more than six parameters. It is rather difficult to accurately estimate the number of mutually independent parameters our method does employ, but a crude estimate can be obtained as follows. In order to specify the relative positions of 11 site points, one must give values for $3 \times 11 - 6 = 27$ coordinates. This is probably an overestimate of the number of geometric parameters, since, for instance, site point no. 3 is likely to be redundant altogether, and quite a range of positions for points 9 and 10 would still bind the naphthalene ring adequately. Of the 48 entries in the energy matrix of Table VIII, only 14 are set to any value other than the default. Of these, three are equal to +10, and any other large positive value would do as well at merely indicating steric repulsion. Solving the least-squares equations for the remaining 11 entries resulted in only nine linearly independent variables, so we may say there are nine energetic parameters. A total of 36 geometric and energetic parameters used to fit only 22 binding energies indicates that the proposed site is by no means the simplest possible solution and that the Hansch approach is much more economical in this respect. On the other hand, our proposed tilting in the binding site better allows us to account for the data on the most difficult inhibitors in a manner completely outside the scope of their method. Hansch et al. indicate their deduced picture of the dihydrofolate reductase binding site in Figure 2 of ref 2b. It is of course much simpler than our Figure 4, but it also qualitatively differs in that they suggest "the region adjacent to position 6 must be open to solvent".[2b] We, however, have site points 9 and 10 located in that area. Furthermore, if we attempt to calculate the binding energies of those analogues outside of our set of 22 having much longer chains attached to position 6, we find that the binding energy is consistently underestimated by as much as 6 kcal. In other words, for our method to account for these larger inhibitors, we would have to propose some extra site points further out beyond the 6 position, rather than letting that part of the ligand interact with no site points (i.e., solvent).

We conclude that we have a workable, novel method for rationalizing binding data. It is capable of dealing with large data sets without requiring unreasonable human or computer effort. Rather subtle steric considerations can be included to give geometrically realistic results better than any other method to date. The resultant site geometries and interaction energies involve more adjustable parameters and detail than is absolutely necessary, but at least they appear to be consistent with the available X-ray crystallographic data. It is hoped that this very geometric and energetic detail, along with the calculated modes of binding for even flexible ligands, will stimulate a more effective exploration of binding sites in the process of drug design.

Acknowledgment. Drs. I. D. Kuntz, M. E. Wolff, R. B. Meyer, and E. C. Jorgensen have contributed many valuable suggestions for the development and presentation of this work. Dr. Paul Weiner and Martin Pensak were of great assistance in preparing the illustrations by computer graphics techniques.

Appendix

We have explored two basic approaches to the problem as set up in the Introduction. The first consists of automatically finding the simplest binding site consistent with the binding data by an exhaustive search of all combinations of number of site points, their types, their distance matrix, and the interaction energy matrix. The second approach has been to propose a site and then compute whether it fits the data, in an interactive "cut and try" fashion. We will first describe the exhaustive approach.

Clearly, the task of finding a site which accounts for the observed binding energies is an open-ended process, in that if a site with n points is adequate then so is one with $n + 1$ points, where the extra one is located so as not to interfere with the preferred binding arrangements. Therefore, the only fair question is what is the *simplest* site that accounts for the data? The algorithm simply consists of trying all possible site geometries and interaction energies. In order to make the search finite, although still very lengthy, we restrict the values of the intersite point distances to a certain list of choices, specified in advance. Similarly, the values of interaction energies are taken from another list, chosen in advance. The search is then organized in the following seven levels: (i) select the number of points to be in the proposed site, usually starting with 2, since a single site point is rather trivial; (ii) for a given number of site points, choose the number of *types* of site points, which may run from one to the number of site points; (iii) for a given number of site point types, choose a particular assignment of types to the points (the number of distinguishable assignments is given by the binomial coefficient of the number of points minus 1 over the number of types minus 1); (iv) choose a site-point distance matrix, where each entry is taken from the allowed distance value list; (v) choose an interaction energy matrix, where each entry is taken from the allowed energy value list; (vi) given the above choice of site, try each ligand molecule in turn to see if the calculated optimal binding energy agrees with the observed value; (vii) for a given ligand, try all possible combinations of ligand-points and site-point contacts, rejecting the geometrically impossible ones and choosing the energetically optimal one. It is clear from this very brief outline that the total number of combinations can be astronomical, even when the number of site points is kept small and the distance value and energy value lists are very restricted. The search can be speeded up many orders of magnitude by cleverly eliminating possibilities in the earlier levels and restricting the choices in the later levels, sometimes according to the

outcome of the energy calculation in the seventh level. However, we find that the cost of computing becomes prohibitive when the number of site points reaches 4. This method is, nevertheless, useful if the experimental data can be fit by a very simple site description, since the combinatorial search will terminate at the very simplest suitable site. These minimal solutions are often surprisingly difficult to guess by inspection. Because of the time limitation, we will not describe the algorithm in greater detail, except for levels six and seven which are the same in the interactive approach, the next topic.

In order to avoid the extremely lengthy combinatorial search, we alternatively propose a binding site by inspection, specifying the number of site points, the type of each, the distance matrix giving their relative locations, and the interaction energy table showing the contribution to the total binding energy for a contact between a ligand point of a given type and a site point of a given type. Specifying the geometry of the proposed binding site in terms of a distance matrix is ordinarily the most convenient way, particularly since the geometries of the ligands it must bind are expressed as upper and lower bound distance matrices. In order to obtain coordinates at the end, one must use the algorithm given in ref 8. Alternatively, one could propose site-point coordinates and then calculate the corresponding distance matrix. In any event, it is not so difficult for a person to propose a site by inspection, and the examples in the preceding sections make it clear how one might go about it. What is persistently difficult, however, is to avoid preconceptions concerning how the ligands must bind to the site. This is overcome by using a computer algorithm to automatically locate the optimal binding arrangement for each ligand to the proposed site and inform the user of any significant discrepancies between calculated and observed binding free energies. In the case of disagreement, the user then alters his proposed site in a likely fashion and tests it again until it satisfactorily accounts for the data.

The heart of the interactive approach (and the exhaustive one also) is the algorithm for finding the energetically optimal mode of binding of a given ligand to a given proposed site, according to given interaction energies. The flow chart in Figure 5 is helpful in following the logic outlined below. (1) Generate successively all possible combinations of contacts of ligand points with each site point, including the possibility that some of the site points may have no ligand points in contact with them. (2) Reject any contact combination which has one ligand point in contact with two site points (or of course one site point in contact with two ligand points), although several unused site or ligand points are allowed. (3) Reject any contact combination which includes contacts with an unfavorable (i.e., positive) interaction energy. (4) Check that for each pair of used site points, i and j, the distance between them, d_{ij}, is in the range of distances allowed to the corresponding ligand points, I and J, with which they are respectively in contact. That is, $u_{IJ} + \Delta d \geq d_{ij}$ and $l_{IJ} - \Delta d \leq d_{ij}$ both hold, where u_{IJ} and l_{IJ} are the upper and lower bounds, respectively, on the distance between ligand points I and J, and I is in contact with site point i and J with site point j. The parameter Δd represents the allowed flexibility in the site. (5) Having now found a contact combination which is geometrically allowed and contains no unfavorable energy interactions, it is now necessary to determine if some (possibly energetically unfavorable) contacts follow as the necessary consequences of those already chosen. These are referred to as "forced contacts" and are classified as being the consequence of certain combinations of three

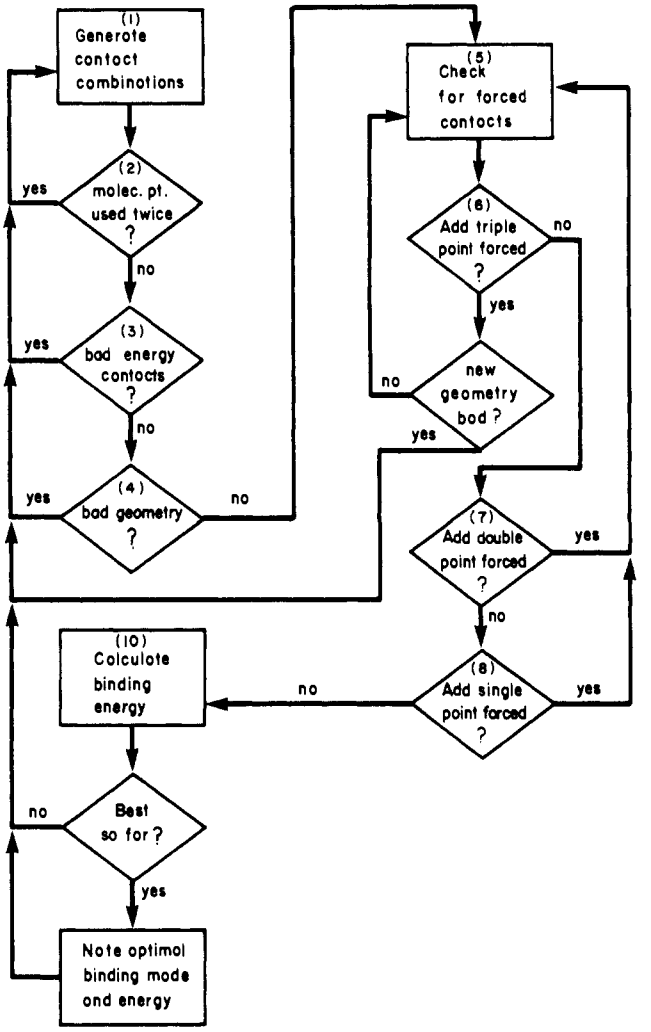


Figure 5. Flow chart of the algorithm for calculating the energetically optimal mode of binding of a given ligand to a given site. Numbers in parentheses refer to the explanation in the text.

contacts, two contacts, or perhaps even of one contact. Each case is considered, in turn, in this order. (6) A triple-point forced contact occurs whenever there is an unused ligand point whose position is fixed in space by having rather invariant distances to three other ligand points which are in contact with some three site points, *and* there is a fourth site point having distances to the other three site points which match the three invariant distances of the ligand points. For example, suppose we were fitting neopentane into a tetrahedral 4-point site, one point for each methyl group. If a proposed contact combination paired methyl groups 1, 2, and 3 with their corresponding sites 1, 2, and 3, then necessarily methyl no. 4 would have to lie on site no. 4. The invariant distance proviso requires that the upper and lower bounds on the interligand points differ by no more than an arbitrary upper limit, taken to be 1 A. That way the fourth ligand point is accurately triangulated in space and has no way of avoiding a correspondingly placed site point. Of course there are some fine points of handedness and mirror inversion to be considered in the case of an asymmetric center in the ligand, but these are neglected in the present version of the program. Any new forced contact is checked for geometric compatibility with the existing contacts. If the test of step 4 is failed, then the whole contact combination is rejected. (7) After all triple-point forced contacts have been deduced, a weaker sort of double-point forced contact is considered. When an unused ligand point

has two invariant distances to some two used ligand points, the position of the unused one is restricted to lie on a circle in space. If there are four or more site points which also lie on that circle (as indicated by their having the same corresponding distances to the two site points involved in contacts with the two used ligand reference points), then the unused point in question must be in contact with one of these four site points. Take the new contact to be with the energetically most favorable unused site point of the four. Of course, the choice of 4 as the number of site points to completely occupy a circle in space is rather arbitrary, and increasing the number would amount to making the search more detailed. The main use of this sort of forced contact is to require a part of the ligand that can rotate to either find an energetically favorable orientation in the site or be excluded altogether in the case that all four site points are repulsive. (8) The last sort of forced contact to be considered is the single-point variety, where making one contact constrains an unused ligand point with invariant distance to the used one, to lie on a spherical shell in space. If there are six site points that also lie on this shell, then the unused ligand point is taken to be in contact with the most energetically favorable unused one. This is certainly the situation in a concave site "pocket", and once again the choice of six is arbitrary (except that it should clearly be greater than the number of site points necessary to force a double-point contact). (9) As indicated above, triple-point forced contacts are the most specific and are tried first until no more can be made. Only then are double-point contacts attempted. If one is formed, then perhaps triple-point contacts can again be deduced, so that must be tried again exhaustively. Only when no more triple- or double-point contacts can be formed are the single-point forced contacts tried. Once again, success results in trying triple points again. When at last no contacts of any variety can be forced, the proposed contact combination is considered to be complete. (10) The energy of the (possibly revised) contact combination is evaluated simply by summing the energetic contributions of each contact. The contribution is taken to be the given interaction energy table entry for the corresponding ligand point type and site point type. Unused ligand or site points contribute zero to the sum. The calculated mode of binding is the contact combination that gives the minimal calculated binding energy.

Fortran programs exist for the above algorithms. Since this work is still in its early stages, these programs are probably difficult for the uninitiated to use. We anticipate improving the methodology, so that our approach may be easily employed by others, but in the meantime the author may be contacted about possible applications.

References and Notes

(1) This work was supported by grants from the Academic Senate of the University of California, by the National Resource for Computation in Chemistry under a grant from the National Science Foundation and the U.S. Department of Energy (Contract W-7405-ENG-48), and by the National Science Foundation directly under Grant PCM78-05468. We are also grateful for the use of the UCSF Computer Graphics Laboratory (NIH RR 1081).
(2) (a) C. Hansch, C. Grieco, C. Silipo and A. Vittoria, *J. Med. Chem.*, **20**, 1420 (1977); (b) C. Hansch, J. Y. Fukunaga, P. Y. C. Jow, and J. B. Hynes, *ibid.*, **20**, 96 (1977).
(3) Z. Simon, I. Badilescu, and T. Racovitan, *J. Theor. Biol.*, **66**, 485 (1977).
(4) P. Gund, W. T. Wipke, and R. Langridge, *Comput. Chem. Res. Educ., Proc. Int. Conf.*, **3**, 5/33 (1973).
(5) K. E. Platzer, F. A. Momany, and H. A. Scheraga, *Int. J. Pept. Protein Res.*, **4**, 187 (1972).
(6) G. M. Crippen, *J. Comp. Phys.*, **24**, 96 (1977).
(7) G. M. Crippen, *J. Comp. Phys.*, **26**, 449 (1978).
(8) G. M. Crippen and T. F. Havel, *Acta Crystallogr., Sect. A*, **34**, 282 (1978).
(9) B. R. Baker and J. A. Hurlbut, *J. Med. Chem.*, **10**, 1129 (1967).
(10) A. Tulinsky, I. Mavridis, and R. F. Mann, *J. Biol. Chem.*, **253**, 1074 (1978).
(11) T. A. Steitz, R. Henderson, and D. Blow, *J. Mol. Biol.*, **46**, 337 (1969).
(12) J. B. Hynes, W. T. Ashton, D. Bryansmith, and J. H. Freisheim, *J. Med. Chem.*, **17**, 1023 (1974).
(13) J. H. Freisheim, C. C. Smith, and P. M. Guzy, *Arch. Biochem. Biophys.*, **148**, 1 (1972).
(14) D. A. Matthews, R. A. Alden, J. T. Bolin, S. T. Freer, R. Hamlin, N. Xuong, J. Kraut, M. Poe, M. Williams, and K. Hoogsteen, *Science*, **197**, 452 (1977).
(15) J. S. Erickson and C. K. Mathews, *J. Biol. Chem.*, **247**, 5561 (1972).
(16) M. Poe, N. J. Greenfield, J. M. Hirshfield, and K. Hoogsteen, *Cancer Biochem. Biophys.*, **1**, 7 (1974).
(17) K. Hood and G. C. Roberts, *Biochem. J.*, **171**, 357 (1978).
(18) A. M. Perault and B. Pullman, *Biochim. Biophys. Acta*, **52**, 266 (1961).

Notes

Synthesis of Benzo-15-crown-5 Polyethers, Anticoccidial Ionophore Analogues

George R. Brown* and Alan J. Foubister

Imperical Chemical Industries Limited, Pharmaceuticals Division, Alderley Park, Macclesfield, Cheshire, England.

Received October 23, 1978

Synthesis of eight benzo-15-crown-5 derivatives I (R = H, CO_2Me, CO_2H, Me; R_1 = H, CO_2H, CO_2Me, CHO, $CH{=}CHCO_2H$, $CH_2CH_2CO_2H$) designed as rigid cyclic analogues of the anticoccidial ionophores is described. No anticoccidial activity was observed in chickens, but moderate activity in tissue culture was found for I (R = Me, R_1 = H; R = R_1 = H) and dibenzo-18-crown-6.

The synthesis of acidic derivatives of benzo-15-crown-5 I designed as rigid cyclic analogues of the anticoccidial ionophores is described. No anticoccidial activity was observed in chickens but moderate activity in tissue culture

0022-2623/79/1822-0997$01.00/0