Prediction of New Leads from a Distance Geometry Binding Site Model

Gordon M. Crippen

Department of Chemistry, Texas A & M University, College Station, Texas 77843, U.S.A.

Abstract

In a previous study, we derived a 9 point model of the binding site of *S. faecium* dihydrofolate reductase, given only the binding data for a set of 68 quinazoline inhibitors. We have now used this model to search for novel chemical classes of inhibitors with comparable predicted activity. The computer algorithm consults a library of ring systems and other molecular fragments of known structure and fits the most likely of these into the site, making chemical modifications where appropriate. The search has located diaminopteridines with calculated activity comparable to that of the best quinazolines for which the site was constructed. In addition, the search algorithm's unfettered imagination proposes much more bizarre candidates, such as a benzoxazoline-2-one dimer. Our present purpose is not to claim we have performed an exhaustive search for all possible antifolates, but rather to outline how the calculations are done and to demonstrate that these preliminary results are reasonable.

1 Introduction

Heretofore, distance geometry methods have been used to deduce a structural and energetic model of a binding site given only the binding constants for a series of compounds (see Ref. [1] and references therein). Although the site model is constructed to fit the data for perhaps a series of chemically similar derivatives, the results are coordinates in Angstroms for each of the site points and some interaction parameters representing the contribution to the free energy of binding for bringing some group of atoms of the ligand into contact with a particular site point. Most other QSAR methods can make binding predictions for molecules that are chemically similar to those of the original data set, but the extreme generality of the outcome of a distance geometry study allows testing of a much wider class of compounds. The only requirement is that they consist of the same types of atoms as occurred in the original data. Different ring systems, planar or nonplanar, chiral or achiral, large and small, rigid or conformationally flexible molecules can all be tested for binding to the given site model. Of course, a site designed to fit the binding data from only a restricted class of compounds probably will not make good predictions for radically different molecules.

In an earlier work of ours, we deduced a model for the inhibition site of *S. faecium* dihydrofolate reductase in terms of 9 site points ("study IV" of Ref. [1]), given the free energies of binding for 68 quinazolines. The correlation coefficient was 0.955, the standard deviation 0.69 kcal/mol, and the maximum error 1.6 kcal/mol. Can such a site lead to new classes of inhibitors, and if so, how must one search for them? In what follows, I outline a practical algorithm for carrying out the search and show some expected results and some very novel ones.

2 Methods

The calculations begin with a given site derived from some previous binding study, and with a library of molecular fragments. The site is described in the usual distance geometry terms as a collection of N_s site points s1, s2, ..., sN_s, and a matrix of interaction parameters (additive contributions to the free energy of binding) $e_{tm,si}$, where tm is the type of the molecule point. One method of describing all possible molecules to place in the given site is to specify a list of atom types (tetrahedral carbon, sp^2 carbon, etc.) and lists of corresponding desired bond lengths, vicinal bond angles, and force constants, so that a molecular mechanics calculation could produce the atomic coordinates of any desired molecule given only the structural formula. Although such an approach is very general and compact, the minimization of the forces on the atoms is time consuming and may be trapped in undesirable local minima, even assuming one had a correct set of forces to reproduce all sorts of strained ring systems. Instead, I have simply amassed a small library of atomic coordinates of rigid groups of atoms, including a selection of ring systems, taken from the crystallographic literature. That way, the geometry of the trial molecules is more likely to be correct, albeit at the expense of generality. Rigid groups may be joined by computer simulating "substitution reactions", where a hydrogen atom of each is deleted, and the two half bonds are splinted together at a standard bond length depending on the atom types being joined. The new bond is always assumed to be rotatable, unstrained, and not an attempt to create a new ring. That way, the standard bond lengths are good approximations to reality, as are the bond angles implicit in the deleted hydrogens. That does require the library to contain all ring systems of interest, as well as groups with bonds having restricted rotation, such as ethylene or a trans-peptide.

More specifically the library consists of a list of molecular fragments and a list of standard bond lengths. To each bond length [2] there is associated a pair of atom types. If there is no bond length given for some two atom types, then such a substitution reaction is taken to be chemically impossible. (Since this is the limit of the program's chemical synthetic knowledge, it will propose all correct derivatives and a number of impossible ones in addition.) Each molecular fragment in the library is either a whole molecule taken from the crystallographic literature or a portion of one with severed bonds completed by hydrogens positioned at the experimentally determined bond angles but the standard bond length. A molecular fragment is reduced to a data structure consisting of the number of atoms present, the type of each, an alphabetic label for each, the Cartesian coordinates, and each atom's bonding to its neighbors. The connectivity is in the form of a tree graph with the root at usually a hydrogen atom; each node (atom) has at most three children, since there are no pentavalent atoms present. Since the bonding is represented as a tree, any additio-

0722 – 3676/83/0311 – 0095 $ 02.50/0

nal bonds forming rings must be noted in a separate list, which is consulted as allowed exceptions to van der Waals contacts during conformational alterations. Having described the input data to the problem, I now turn to the algorithm for building prospective drugs from the library of fragments. The following pseudocode presentation starts at the highest level of decision making and works down to computational details, giving enough of the particulars to enable a programmer to reproduce the work without burying the overall logic in minutiae. Two of the low level routines are explained elsewhere: the optimal translation and rigid rotation of the molecule onto the desired site points [3], and the "clique" routine for enumerating all geometrically allowed binding modes for the molecule [4]. Clique is much faster for large molecules than the tree search used formerly but still only checks *necessary* conditions for geometric correctness. Screening out the bogus binding modes requires the atomic coordinate translation and rotation onto the site points.

First consider the overall interaction between the chemist and the computer:

```
Enter site point coordinates and interaction matrix from earlier
study.
While patience lasts,
  Include all molecular fragments of interest in library.
  Use "newleads" procedure to find best binding fragments
  and best binding single substitution derivatives.
  Reject chemical synthetic impossibilities.
  Of those remaining, select best molecules on the basis of:
    the most favorable calculated binding energy,
    ease of synthesis,
    availability of further substitution sites.
  For each selected molecule in turn,
    if further substitution is possible and desirable,
      add the molecule to the library,
      delete the parent molecule from which it was derived,
      and iterate this entire procedure with the revised library.
    otherwise
      propose the molecule as a new lead compound.
  When all selected molecules have been examined, quit.
```

Note that the procedure is recursive, so that the chemist is really traversing a tree of molecular structures. The first layer down is just the contents of the original library, the second layer consists of single substitutions on these, the third layer has another substitution somewhere on the single substitution derivatives, and so on. An exhaustive search of the complete tree of organic compounds is nearly an infinite task, so the method relies on a selective search over a preferred library of molecular fragments. As the tree is descended, most possibilities are rejected according to the criteria above, some of which are admittedly subjective.

The heart of the calculation is the "newleads" routine, which is responsible for synthesizing the derivatives in the computer and eliminating most of the possibilities:

Procedure Newleads

```
Read in the library of molecules.
Read in the site point coordinates and interaction matrix.
Set the energy tolerance to, say, δE = 2.0 kcal.
Sort the molecules according to their estimated binding
energy, best first (in case of ties, the molecule with fewer
atoms is preferred).
  The estimated energy is only a lower (most favorable)
  bound, made without regard to geometric considerations.
    Estimated energy = 0, initially.
    While there are any unused site points and any unused
    atoms,
      find the unused site point and unused atom with best
      energy of interaction.
      If that energy < 0, i.e. not unfavorable, then
        make that contact (atom and site point now
        "used"); add the energy of contact to the estimated
        binding energy.
    End of loop over unused site points and atoms.
The best binding energy so far = 0, i.e., not repulsive.
For each molecule in the sorted list,
  use procedure "dock" to find the energetically optimal
  binding.
  If the calculated binding energy is better than the best so far,
  or at least no more than δE worse, then
    write the molecule out for possible subsequent inclusion in
    the library.
  If binding energy < 0, and the optimal binding mode left
  some site points unused, and there are substitutable hydro-
  gens left, then
    make a new sorting of the library molecules as before,
      except that in the estimation of binding energy,
      the atoms and site points used in the optimal binding
      mode are immediately labeled as "used".
    For each substituent molecule in the sorted list,
      reject it if the optimal binding energy of the parent
      molecule plus the estimated additional binding of the
      substituent is more than δE worse than the best so far.
      Reject it if estimated binding is > 0, i.e. repulsive.
      Otherwise, for each hydrogen on the parent molecule,
        For each hydrogen on the substituent molecule,
          Reject if the proposed bond is not in the standard
          list.
          Otherwise, use procedure "join" to perform the
          substitution "reaction".
          Use procedure "dock" to find the energetically
          optimal binding mode of the derivative without
          regard to the binding mode of the parent molecule.
          If the calculated binding energy is better than the
          best so far, or at least no more than δE worse, then
            write the molecule out for possible subsequent
            inclusion in the library.
          Restore the parent molecule to its original state.
      End of loop over substitution isomers.
    End of loop over substituent molecules.
End of loop over parent molecules.
```

Note that as the search goes on, the best calculated binding energy "seen so far" monotonically improves, so the result of procedure newleads is a list of candidate library molecules and linked pairs of the same that tend to have better binding energies toward the end of the list. As long as $\delta E > 0$, the list is not precisely ordered, and indeed, one would not want $\delta E = 0$ on the grounds that such a choice would reject molecules binding only slightly worse than the best found so far. However, in a subsequent round of substitution modifications, these rejected molecules might yield excellent derivatives. The alternative of a large positive value, of course, would swamp the chemist with many mediocre proposed molecules.

Procedure newleads relegates a lot of bookkeeping details to procedure join, which is where the chosen substitution reaction takes place.

Procedure Join

```
Given are
  the two library molecules to be joined,
  which hydrogen atom of each is to be deleted,
```

and which atom ("kept" atom) of each is bonded to the deletable hydrogens.
The bond on each molecule from the deletable hydrogen is called its "joining" bond.
Make a copy of the two molecules.
Translate and rotate the second molecule to superimpose joining bonds.
Define an axis passing through the second molecule's kept atom oriented normal to the plane defined by the two joining bonds.
Rigidly rotate the second molecule about the axis so that the two joining bonds are antiparallel.
Translate the second molecule so that kept atoms coincide.
Then translate it along the joining bond vector to the desired bond length.
Revise the connectivity lists, etc.
Bonds to the deletable hydrogens must be removed from the connectivity tree.
The new bond between the two molecules must be added to the tree.
If the kept atom of the first molecule is the root of the connectivity tree and is tetravalent, the root must be shifted to another atom to ensure that each atom has at most 3 children.
If the second molecule's kept atom was not the root of its connectivity tree,
the pointers must be reversed between the kept atom and the old root.
Remove the two hydrogens and renumber the atoms in the connectivity lists, etc.
Add the new bond to the list of rotatable bonds.

There remains only to explain the docking algorithm for finding the optimal binding mode. In general terms, the goal is to find the energetically optimal binding mode, checking that it is truly geometrically correct by making a least-squares superposition of the molecule's atoms onto the corresponding site points by overall rigid translation and rotation of the molecule and by adjustment of any internal rotatable bonds. In order to correctly alter torsional angles, you must decide which rotatable bonds move which atoms relative to the rest, and which groups of atoms always have the same relative positions (the "rigid groups" described below). Procedure clique quickly provides a list of all possible binding modes (and sometimes some impossible ones, too) given the (not necessarily best estimated) upper and lower distance bounds between atoms. Finding these bounds is another use for the rigid groups. Those readers interested in greater detail should examine the following outline:

Procedure Dock

Given are the site structure, the interaction energies, and the molecule's atomic coordinates, connectivity, and rotatable bonds. Determine the (overlapping) sets of atoms, called "rigid groups" whose mutual distances are unaffected by rotatable bonds.
For a molecule with N_r rotatable bonds, there are $N_r + 1$ distinct rigid groups. Two atoms are in the same rigid group if either they are linked by a series of non-rotatable bonds, or one of them lies on the common axis of all rotatable bonds linking them.
While there is at least one atom not included in a rigid group,
make a new rigid group and include the atom as its first member.
Put the atom in the stack.
While there is at least one atom in the stack,
Take an atom off the stack.
If its bonded neighbors are not already included in this group,
if the bond joining them is not rotatable,
add the neighbor to a stack for later consideration.
else
put the neighbor in the rigid group,
note the rotatable bond's direction and origin,
and put neighbors beyond the bond into a second stack.
While there are atoms in the second stack,
take an atom off the stack,
Include it in the group if it lies near the rotatable bond axis.
In any case, stack its unincluded neighbors as long as no rotatable bond is crossed which is noncolinear with the original one.
End of the second stack.
End of the first stack.
All atoms included in at least one rigid group.
For all pairs of atoms in the molecule,
if they are in the same rigid group,
upper bound distance = lower bound = actual distance given by the atomic coordinates.
else
upper bound distance = ∞
lower bound distance = sum of van der Waals radii.
Smooth upper bounds by the triangle inequality [5].
Smooth lower bounds by the inverse triangle inequality [5].
Use procedure "clique" [4] to find all geometrically plausible binding modes, given the distances between site points and bounds on the distances between atoms.
Sort the list of binding modes in order of estimated binding energy.
That is, assuming all contacts in a mode can be made, and no other contacts will be made in addition.
Initially the best binding energy = 0.
While the estimated energy of the next mode in the sorted list is still less than the best binding energy actually calculated so far, attempt to form the contacts constituting that mode.
Initially the molecule has translational freedom, two degrees of overall rotational freedom, and whatever rotatable bonds it has.
Initially all rigid groups and contacts are untreated.
While there are degrees of freedom left and there are untreated contacts,
find the untreated rigid group closest to a treated one and otherwise having the greatest number of contacts.
For the first rigid group to be treated, just take the one with the most contacts.
Bind the chosen rigid group to the corresponding site points.
If translation of the whole molecule is available,
Translate all atoms for a least-squares coincidence of those atoms of the rigid group involved in contacts with their corresponding site points.
Translational freedom is no longer available.
If more than one contact is to be made,
optimally rotate all atoms to form contacts [3].
One rotational degree of freedom is used.
If three or more contacts were made to noncolinear site points, then the second degree of rotational freedom is also used.
Else if both rotations are available,
optimally rotate the molecule as above, subtracting one or both rotational degrees of freedom.

The origin for rotation must be the single contact involved in the previously treated rigid group.
Else if only the second rotation is available,
optimally rotate the molecule as above,
deleting the last rotational degree of freedom.
Either the first rigid group had two contacts,
or the first two rigid groups had one each,
so the axis of this last rotation is already determined.
Else only rotatable bonds are left.
Determine which rotatable bonds affect the position of the rigid group in question relative to that of the nearest treated rigid group.
During the determination of the rigid groups, it was noted which rotatable bond joined adjacent rigid groups.
The rigid groups of the molecule are nodes on a graph where the edges link adjacent groups.
Then Dijkstra's algorithm [6] gives the shortest path on the graph from the treated group to the group to be positioned.
The bonds corresponding to the steps on the path are the ones that must be rotated.
A single rotatable bond requires only one iteration, otherwise for 3 iterations,
For each rotatable bond in the list,
rotate all atoms on the side of the connectivity tree including the group to be positioned so that the desired contacts are optimally made.
McLachlan's algorithm [3] calculates an unnormalized vector indicating the axis and magnitude of the rotation to optimally make the desired contacts.
In this case, the axis is just that vector's component along the bond.
Delete all the rotatable bonds in the list from the list of degrees of freedom.
End of forming contacts loop.
The actual calculated energy of the binding mode is the sum of the interactions for each atom closer than δd (= 0.5 Å, say) to the site point, whether or not that contact appeared in the proposed mode.
End of trying different binding modes loop.

Note that the estimated upper and lower bounds on the interatomic distances supplied to procedure clique are not very tight. A more precise method would be to rotate about all combinations of rotatable bonds, noting the greatest and least observed distances for all pairs of atoms. Such a calculation is rather lengthy for more than 3 or 4 torsion angles, so instead we use the much faster but looser triangle inequality bounds. Consequently, procedure clique produces a longer list of binding modes (typically about 100, although sometimes 10000 for a large 50 atom molecule) some of which are in fact not geometrically allowed. However, these are tried in order of estimated binding energy, and typically the first one is successful, and the other modes are disregarded. Although the logic of the docking procedure is rather complex, the actual number of calculations performed is rather small. For molecules with large numbers (say 50) of atoms and small numbers of rotatable bonds (say 3), the procedure offers real speed advantages over the alternative of minimizing with respect to the Cartesian coordinates of the atoms some force field that favors the geometry of the rigid groups and draws contacting atoms to their corresponding site points. The only part of docking with dubious convergence properties is when several rotatable bonds must be adjusted to swing a rigid group into best coincidence with the desired site points. The iterated sequential approach used above has been successfully employed by other investigators [7].

3 Results

In an earlier work [1], we had developed a model of the binding site for dihydrofolate reductase based solely on the binding data for 68 quinazolines. We had found a minimum of 9 site points were required for an accurate fit to the data (see Ref. [1], "study IV"), and these 9 points plus the corresponding interaction matrix are the site description I have taken for the present work. Figure 1 shows the optimal calculated binding

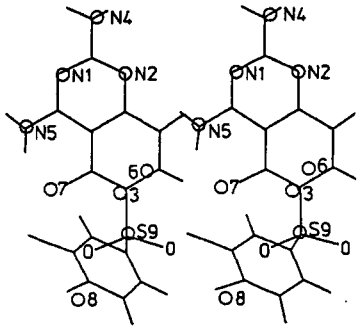

Figure 1. Stereo view of the calculated binding mode [1] of 2,4-diamino-6-(sulfuryl-3',4'-dichlorophenyl)quinazoline, as an illustration of the roles of the 9 site points. Site points are shown as numbered circles. Atoms other than carbon and hydrogen are labeled.

mode for 2,4-diamino-6-(sulfuryl-3',4'-dichlorophenyl)quinazoline, the one of the 68 quinazolines with the best experimentally determined $\Delta G_{bind} = -13.4$ kcal. Imagine that this is all that is known about *S. faecium* dihydrofolate reductase. Would the algorithm presented in the previous section discover new types of inhibitors? I began with a small library of molecules we had employed in other binding studies: CH_4, HCl, HF, HBr, HI, HNO_2, carbazole, β-carboline, H_2O, NH_3, H_2NCOH (peptide group), pyridine, 1,3-dihydro-2H-1,4-benzodiazepin-2-one, pyrimidine, quinazoline, ethylene, H_2S, HCF_3, benzene, naphthalene, H_2CO, H_2SO_2, H_2SO, HCN, pteridine, 2,3,4,8-tetrahydro-β-carboline, benzoxazoline-2-one, 5,6,7,8-tetrahydroquinazoline, and N1-protonated versions of the quinazolines, pyrimidine, and pteridine. The best calculated binding free energies of the unaltered library molecules were tetrahydro-β-carboline (−5.156 kcal), benzoxazoline-2-one (−5.080), protonated pteridine (−4.635), and protonated quinazoline (−4.635). The next best were carbazole and carboline at −4.111, and all the rest were worse than −3.8 kcal. I arbitrarily restricted my attention to the best three molecules. Of those, it is certainly no surprise that quinazoline bound well, since the site had been constructed for that very purpose. Therefore I took the strong binding of the protonated quinazoline to be an expected confirmation of the calculation, but otherwise not worth pursuing further. The tetrahydro-β-carboline is an interesting possibility, but with 12 substitution sites on the molecule, a full investigation of its possible derivatives would require hours more computer time than would be justified for this preliminary study. That left benzoxazoline and pteridine leads to follow up.

N1-protonated pteridine bound in a mode analogous to that of quinazoline in Figure 1. Because amino-substituted quinazolines are such successful inhibitors, I restricted the library at this stage to the pteridine and NH_3. The best single substitution was 4-aminopteridine (−6.842 kcal), and the best choice for substituting that in turn was 2,4-diaminopteridine (−9.126 kcal). Offering 2,4-diaminopteridine a choice of $-NH_2$, −S-, and −SO-substituents resulted in 2,4-diamino-6-SOH-pteridine

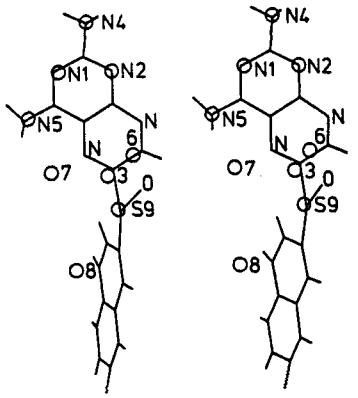

Figure 2. Stereo view of the calculated binding mode of 2,4-diamino-6-(sulfinyl-β-naphthyl)pteridine. The orientation of the site is the same as in Figure 1.

at −10.928 kcal with the close runner up of 2,4-diamino-6-SH-pteridine at −10.811 kcal. Trying further amino, −SH, phenyl, naphthyl, and −SO-derivatives of 2,4-diamino-6-SOH-pteridine resulted in no improvement for subtle geometric reasons. For example, Figure 2 shows 2,4-diamino-6-(SO-β-naphthyl)pteridine (−10.928 kcal) in its optimal binding mode, analogous to that of the quinazolines. The binding calculation has attempted to add an extra contact between site point 8 and one of the naphthyl carbons by optimally rotating the two C–S bonds, but it misses by over 2 Å. The coordinates of site point 8 depend very sensitively on the prescribed distances between it and site points 3 and 9. It would be easy to adjust the position of site point 8 slightly such that quinazolines still bound as in Figure 1 while also allowing the pteridines to contact it, bringing the calculated binding energy in Figure 2 to −12.2 kcal. However, the object of this study is to see what predictions can be obtained from the earlier site determination, not to carry out a combined binding study for quinazolines and pteridines. Even so, if we had known only about quinazoline inhibitors, these calculations could have sent us on the trail toward the pteridine derivative, methotrexate.

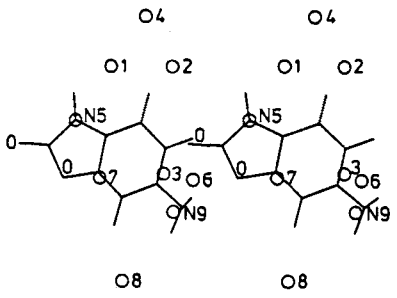

Figure 3. Stereo view of the calculated binding mode of 6-aminobenzoxazoline-2-one in the same orientation as in Figure 1.

The other, very unexpected lead to follow is benzoxazoline-2-one. Offering it a choice of −Cl, $-NH_2$, −SH, phenyl, naphthyl, and −SO− substituents, it chose a variety of different benzoxazoline dimers, all binding in the range of −5.5 to −6.57 kcal, and 6-aminobenzoxazoline-2-one, binding at −6.57 kcal. This last is shown in Figure 3 in its calculated optimal binding mode. Trying to enhance its binding by further substitution by $-CH_3$, −Cl, −F, $-NO_2$, $-NH_2$, pyridyl, −SH, $-CF_3$, phenyl, naphthyl, and sulfinyl groups resulted in

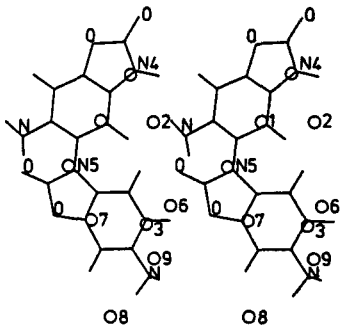

Figure 4. Stereo view of the calculated binding mode of 5-(3′-(6′-aminobenzoxazol-2′-one-yl))-6-aminobenzoxazoline-2-one.

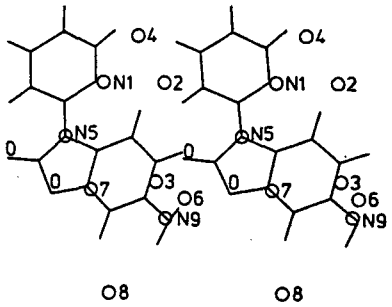

Figure 5. Stereo view of the calculated binding mode of 3-(2′-pyridyl)-5-aminobenzoxazoline-2-one.

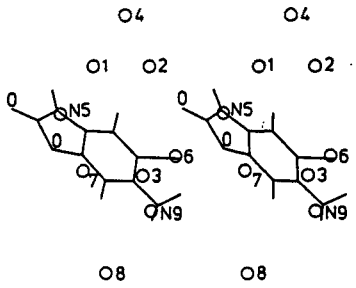

Figure 6. Stereo view of the calculated binding mode of 5-chloro-6-aminobenzoxazoline-2-one.

5-(3′-(6′-aminobenzoxazol-2′-one-yl))-6-aminobenzoxazoline-2-one (−7.36 kcal), 3-(2′-pyridyl)-5-aminobenzoxazoline-2-one (−7.02 kcal), and 5-chloro-6-aminobenzoxazoline-2-one (−6.75 kcal). The 9-point binding site model was constructed without any repulsive site points to delineate the "walls" of the site because none of the 68 quinazolines in the original data set showed the anomalously poor binding characteristic of a steric effect. Consequently we are completely ignorant (as far as Ref. [1] goes) about what space surrounding the quinazoline in Figure 1 may be occupied by the enzyme. The binding mode of the benzoxazoline-2-one in Figure 3 occupies a substantial amount of questionable territory in the vicinity of site points 5 and 7, not to mention the binding mode of the dimer shown in Figure 4. Further pursuit of this line of derivatives is likely to be a fruitless theoretical excercise without some experimental verification of at least benzoxazoline-2-one binding.

4 Discussion

In the past we have shown how distance geometry analysis of binding data can produce models of sites that agree closely in

structure with experiment [8], that reproduce closely the observed binding data [1], that simultaneously account for the binding of ligands from widely different chemical classes [9, 10], and that have good predictive power [10]. Now this study demonstrates that because of the very general description of the site, the model developed in a binding study can be used to systematically search for new lead compounds. The work I have presented here is by no means intended to be an exhaustive search for all possible dihydrofolate reductase inhibitors, but rather an explanation of the method along with an illustrative application in a situation where we have an accurate site developed from a restricted data set for a well studied protein. As expected, a site model developed for quinazolines is discovered to bind those same quinazolines; and 2,4-diaminopteridines are predicted to also bind well, quite in accord with well known experimental fact. The exciting part is that if we had only known about quinazolines as inhibitors, these same data could have lead to a genuine discovery of methotrexate analogues! In order to have "discovered" something along the lines of trimethoprim, the first pass through the library would have to be much more tolerant, since pyrimidine had a binding energy of only −2.9 kcal. My search would not have located the triazines, simply because they were not present in the library. The truly unexpected development of this work was benzoxazoline-2-one. For reasons presented above, it is unlikely this compound really binds tightly, but the prediction that it does is an honest extrapolation from limited data. It is not so important whether or not it is in fact a good inhibitor, but rather that the search algorithm is relatively free from our prejudices and produces some unusual results during its search for the unusual.

Questions of efficiency are important in this sort of algorithm. The longest of the library searches or single-substitution runs I have described in the Results section took 40 minutes on a Data General MV8000 minicomputer (speed equivalent to a VAX 11/780). A much broader library could easily be employed without great computer costs. The sudden escalation in run time comes when there are one or more molecules with say, 10 or more substitutable hydrogens apiece. Then simply examining all dimers of one of these molecules is the equivalent of adding 100 (relatively large) new members to the library. Since there are an astronomical number of possible molecules even in the < 50 atom range of most drugs, the computer time required for these searches can easily get out of hand. I frankly suggest a large role for the chemist in pruning the search tree as one explores different avenues. The human intervention brings a lot of information to bear on the problem which is not readily incorporated into the computer algorithm, while the computer furnishes dogged thoroughness and a useful lack of bias.

Acknowledgement

This work has been supported by grants from the National Institutes of Health (1-R01-AM28140-01) and the Robert A. Welch Foundation. I thank Dr. A. K. Ghose for helpful discussions and for preparing the illustrations.

(Received on June 11th 1983, accepted on September 23rd 1983)

5 References

[1] A. K. Ghose and G. M. Crippen, *J. Med. Chem. 25,* 892 (1982).
[2] L. E. Sutton *et al.* (Eds.), *Table of Interatomic Distances and Configurations in Molecules and Ions,* Special Publication 18, The Chemical Society, London (1965).
[3] A. D. McLachlan, *Acta Cryst. A38,* 871 (1982).
[4] F. S. Kuhl, G. M. Crippen, and D. K. Friesen, *J. Comp. Chem.,* in press (1983).
[5] G. M. Crippen, *Distance Geometry and Conformational Calculations,* Research Studies Press (Wiley): Chichester, England and New York 1981.
[6] E. W. Dijkstra, *Numer. Math. 1,* 269 (1959).
[7] L. Barino, *Computers and Chem. 5,* 85 (1981).
[8] G. M. Crippen, *J. Med. Chem. 24,* 198 (1981).
[9] G. M. Crippen, *Mol. Pharmacol. 22,* 11 (1982).
[10] A. K. Ghose and G. M. Crippen, *J. Med. Chem. 26,* 996 (1983).

Rationalisations among Heterocyclic Partition Coefficients Part 2: The Azines

Susan J. Lewis, Margaret S. Mirrlees, and Peter J. Taylor*)

ICI Pharmaceuticals Division, Alderley Park, Macclesfield, Cheshire SK10 4TG, England

List of Abbreviations and Symbols

P = partition coefficient
π = increment in log$P$
May also refer to π-electrons if this is made clear
σ = Hammett σ-value
May also refer to σ-electrons if this is made clear
f = fragment value
pK_a = negative logarithm of acid dissociation constant pK_a
u.v. = ultra-violet
INH = isonicotinic acid hydrazide

All other abbreviations are conventional or fully referenced.

Abstract

Azine π-values are discussed in terms of $\Delta\pi$, which is the difference in π-value from that expected for benzene. It is shown that $\Delta\pi$ is close to zero for alkyl and most halogen

*) To receive all correspondence.

0722 – 3676/83/0311 – 0100 $ 02.50/0