Journal of Computer-Aided Molecular Design, 7 (1993) 83–102
ESCOM

J-CAMD 189

A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists

Yvonne C. Martin*, Mark G. Bures, Elizabeth A. Danaher, Jerry DeLazzer, Isabella Lico and Patricia A. Pavlik

Pharmaceutical Products Division, Abbott Laboratories, Abbott Park, IL 60064, U.S.A.

Received 2 June 1992
Accepted 16 October 1992

Key words: Molecular design; Bioactive conformation; Molecular superposition; Binding site models; Pattern recognition; Clique detection

SUMMARY

In the absence of a 3D structure of the target biomolecule, to propose the 3D requirements for a small molecule to exhibit a particular bioactivity, one must supply both a bioactive conformation and a superposition rule for every active compound. Our strategy identifies both simultaneously. We first generate and optimize all low-energy conformations by any suitable method. For each conformation we then use ALADDIN to calculate the location of points to be considered as part of the superposition. These points include atoms in the molecule and projections from the molecule to hydrogen-bond donors and acceptors or charged groups in the binding site. These positions and the relative energy of each conformation are the input to our new program DISCO. It uses a clique-detection method to find superpositions that contain a least one conformation of each molecule and user-specified numbers of point types and chirality. DISCO is fast; for example, it takes about 1 min CPU to propose pharmacophores from 21 conformations of seven molecules. We typically run DISCO several times to compare alternative pharmacophore maps. For D_2 dopamine agonists DISCO shows that the newer 2-aminothiazoles fit the traditional pharmacophore. Using site points correctly identifies the bioactive enantiomers of indoles to compare with catechols whereas using only ligand points leads to selecting the inactive enantiomer for the pharmacophore map. In addition, DISCO reproduces pharmacophore maps of benzodiazepines in the literature and proposes subtle improvements. Our experience suggests that clique-detection methods will find many applications in computational chemistry and computer-assisted molecular design.

INTRODUCTION

Frequently small molecules with very different 2D structures displace each other from a binding site on a macromolecule. Even more often, minor modification of the structure of an active molecule renders it inactive. Such structure–bioactivity relationships are an indirect probe of the

*To whom correspondence should be addressed.

3D structure and chemical properties of the macromolecular recognition site for the ligands. The goal of pharmacophore mapping is to transform such 2D structure–activity information into the 3D requirements for binding to the target biomolecule [1,2]. This allows one to search 3D databases for other molecules that match these 3D properties or to design new active molecules [3,4].

A pharmacophore map identifies the bioactive conformation of each active molecule and indicates how to superimpose, compare in 3D, the various active compounds. The map identifies which types of points match in what conformations of the compounds. The decisions as to the required points and the bioactive conformations are interdependent; i.e., the choice of one affects the choices available for the other.

A complication for pharmacophore mapping is that different ligands might approach a polar site point from different directions. For example, a carbonyl group on a protein optionally interacts with hydrogen-bond donors in the directions of its two lone pairs, which are separated by 120° and 2.9–3.0 Å [5]. The result is that, in a pharmacophore map, the positions of the ligand atoms may not overlap, even though their projections to the macromolecule do. The problem this leads to is illustrated in Fig. 1, which shows three dopaminergic agonists and the site hydrogen-bonding points with which they could interact.

In contrast, since hydrophobic interactions involve more area, different ligands would most likely approach the hydrophobic binding site from the same direction. Hence to match these features in the ligands we will superimpose ligand-based points only.

We sought to devise a strategy that would identify the bioactive conformation and the matching points simultaneously. It would tell us if some molecules have one or several conformations that fit a pharmacophore hypothesis. The ideal strategy would tell us if there were several equally attractive superpositions with the same number (but different identity) of matching points. Also it would allow us to compare superpositions with varying numbers of points. For example, how is the superposition affected if one requires that the superposition contains an additional point or one specific type of point?

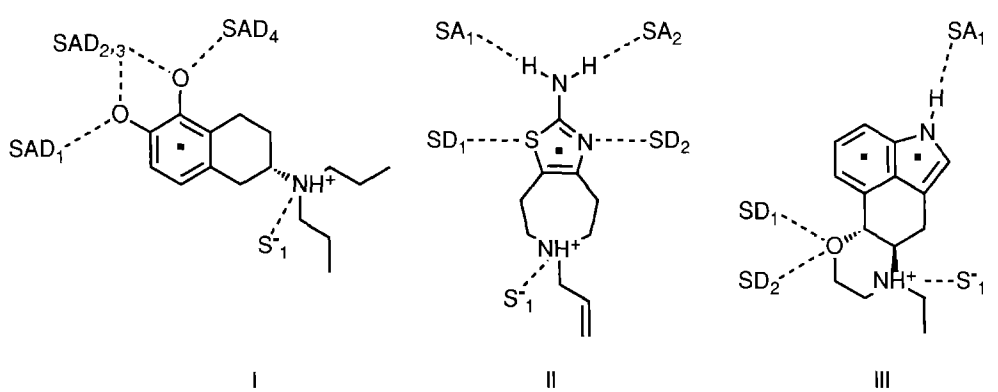


Fig. 1. The possible points to overlap in three typical dopaminergic agonists. SA refers to site hydrogen-bond acceptor points, SD to site hydrogen-bond donor points, and S^- to site anionic hydrogen-bond acceptor sites. In I alternate rotations of the OH bonds lead to site SAD_1 and SAD_3 being hydrogen-bond acceptors and SAD_2 and SAD_4 being hydrogen-bond acceptors or vice versa. The user may also choose to overlap points in the ligands such as the basic nitrogen atoms, the hydrogen-bond donors, or the centers of mass of aromatic rings.

For pharmacophore mapping of potent small molecules, we prefer that the proposed bioactive conformation be a low-energy one. We assume that the interacting atoms of a macromolecule can move slightly and with little energy cost to make optimal interaction with a ligand [6,7]. This assumption is reinforced by observations from small-molecule crystallography that show a variation in the location of the atoms involved in a strong electrostatic interaction or a hydrogen bond [8]. Thus, we accept a larger tolerance (a value at which inter-point distances are considered equivalent) between overlapping points if the resulting pharmacophore map includes significantly lower-energy ligand structures.

Our experience is that it is straightforward but computationally expensive to search conformational space. For example, the generation of 100 structures with distance geometry [9] and energy minimization of these structures may take several hours of computer time using a personal workstation. If the molecule is flexible, 100 structures may not be enough to explore conformational space. Furthermore, if the molecule is not parameterized for molecular mechanics, then quantum-chemical minimization may be necessary. This often takes at least a few hours per starting 3D structure. For these reasons we prefer a strategy in which we search conformational space only once for each molecule.

Conversely, using molecular graphics to compare many conformations of several molecules can take much human time. These comparisons may be biased by the order in which one studies the molecules. We therefore sought an objective and fast method that would compare all low-energy conformations of all molecules simultaneously.

Several groups have explored methods to superimpose two defined 3D structures when one does not know the corresponding atoms [10–19]. The most attractive solution is a method highlighted by Brint and Willett [10] based on an earlier report by Kuhl et al. [20] who used it for docking a small molecule into a protein. Brint and Willett showed that clique-detection methods based on interatomic distances rapidly find the best way to match two 3D structures. For example, the Bron-Kerbosh clique-detection method can find the six matching atoms in two 15 atom structures in 15 s on a PRIME 9950. Brint and Willett extended the method to find the maximum 3D substructure common to all members of a set of molecules.

In graph-theory terms, a clique is a subgraph in which every node is connected to every other node. For 3D structure comparisons, the nodes of this graph are the points for superposition labeled according to type. An edge between two nodes is described by the distance between them. Clique-detection algorithms find cliques in an input graph that match cliques in a reference graph. That is, they find corresponding points in the two 3D structures. These corresponding points are of the same type in the two structures and all corresponding interpoint distances are identical.

Since clique-detection methods are so fast, we decided to explore their utility as a step between conformational searching and molecular graphics comparisons, i.e., to propose pharmacophore maps. This would allow us to vary the numbers and types of points in the possible pharmacophores and to rule out impossible superpositions in little time. We would then study with molecular graphics only the most reasonable potential solutions.

Several improvements of the Brint–Willett implementation are necessary before we could apply clique-detection methods to pharmacophore mapping. First, we must consider multiple alternative conformations of the molecules. Second, pairing should be based not on atomic number but rather on chemical similarity such as hydrogen-bond donating or accepting atoms. Third, the points for matching should include site points as well as ligand points. Lastly, we must address

how to preserve the chirality of molecules yet still match structures with fast distance-based algorithms.

This report describes our program DISCO (DIStance COmparisons), how we overcame the limitations in the method of Brint and Willett, and our experiences with the strategy.

METHODS

The steps in our pharmacophore mapping strategy are generation of the possible conformations of the compounds, calculation of the location of the ligand and site points, execution of DISCO to find potential pharmacophore maps, and molecular graphics analysis of the results. Key features of each of these steps are discussed below.

Conformational analysis

The 3D structures of the compounds can be generated using a simple program such as CONCORD [21], from crystal structures [22], or from conformational searching and energy minimization with any molecular or quantum-mechanical technique [2].

We do not want to include similar conformations in the DISCO comparisons since every conformation considered can lead to a pharmacophore superposition. Thus, our program REJECT compares all the conformations of a molecule to exclude duplicates [23]. If each corresponding interatomic distance between selected atoms in the two conformations is less than a threshold (typically 0.4 Å), then we reject the higher energy conformation. The result is a set of the unique low-energy conformations of a molecule.

Location of potential pharmacophore points with ALADDIN

Since it already had many relevant capabilities, we enhanced our 3D searching program, ALADDIN [24], to prepare the input for DISCO. Specifically, ALADDIN calculates, from the supplied 3D structures, the location of the ligand or other points to be considered for the pharmacophore. For example, one might choose to use centers of rings, the ring atoms themselves, or the location of a halogen as hydrophobic groups for potential overlap. A second variety of points is the location of site hydrogen-bond acceptors or donors. Many of these points are automatically calculated by ALADDIN. In addition, the location of other site points may be calculated by information supplied by the user.

The default locations of site hydrogen-bond donor and acceptor points are based on compilations of observed intermolecular crystallographic contacts in proteins [7] and between small molecules [8,25–28]. Table 1 shows where ALADDIN places the site points for various functional groups. The user may supply a different distance to a site point. This would be helpful if, for example, the macromolecule was thought to contain a metal atom in the binding site.

Hydrogen-bond donors and acceptors such as OH and NH_2 groups can frequently rotate to change the locations of the hydrogen atoms. To consider this flexibility, if at all possible, we do not use the reported location of hydrogen atoms in calculating the location of site points. Rather, we calculate all potential locations of site points. For example, there are three optimum locations for a hydrogen-bond acceptor interacting with an aliphatic hydroxyl group. In ALADDIN they are calculated from the location of the oxygen, the carbon attached to it, and an atom attached to the carbon.

Using ALADDIN solved the problem of generic labeling of points since the program uses a substructure-perception language, GCL (Genie Control Language), to identify atoms based on user-specified descriptions [29]. This allows users, for example, to recognize and give the same label to all hydrogen-bond donors in the ligand. These labels are used to tell DISCO the number and type of points to consider. However, one may also choose to distinguish different types of hydrogen-bond donors and to merge them only if desired.

ALADDIN prepares two new types of files. The first is the input to DISCO. For each 3D structure this file contains the 2D structure as its SMILES [30], the XYZ set name and energy, and the coordinates of each point labeled by type. If any are known, one can use typical 3D constraints to restrict the points or the conformations to be considered by DISCO. ALADDIN also optionally prepares molecular structure files with the whole structure plus the labeled points. This eases the evaluation of DISCO results with our molecular graphics program SWAMI [31].

Options for running DISCO to find pharmacophore maps

A set-up file controls the options of a DISCO calculation. Usually DISCO is instructed to iterate the tolerance at which two inter-point distances are considered the same. The starting and final values and an iteration increment are given. Alternatively, the user may input the tolerance for each type of inter-point distance and the program will not iterate.

The user may direct DISCO to consider all potential points and to stop when a pharmacophore map with a certain total number of points is found. Alternatively, one may specify the types of points and the minimum and/or maximum number of each that every superposition must include. One can also run DISCO so that certain types of points labeled differently are treated as one type. Usually, several runs are made varying these choices to compare the results.

DISCO can be instructed to ignore specific compounds if they do not match a pharmacophore map. This allows one to include inactive compounds in an analysis while not requiring that each appears in all superpositions. One may also direct DISCO to accept superpositions that include fewer than a certain number of unspecified compounds. This option is useful if one is not certain that all low-energy conformations are included.

The user may specify that only the input chirality is to be used for certain molecules and/or that only conformations below a certain relative energy should be considered. These two options prevent DISCO from providing pharmacophore maps that are known to be incorrect.

Lastly, users can request DISCO to prepare molecular graphics macros and/or hard-copy of the 2D structures with the matching atoms labeled.

Algorithm used by DISCO

The molecule with the fewest conformations is used as the reference. DISCO searches each conformation of it separately for pharmacophore maps that contain at least one conformation of every other molecule.

The search begins by reading the ALADDIN output and associating the conformations of each molecule with each other. DISCO then calculates the distances between points in each 3D structure. Next, it prepares correspondence tables that relate inter-point distances in the current reference conformation and distances in every other 3D structure. Distances correspond if the point types are the same and the distances differ by no more than the tolerance. The Bron–Kerbosh clique-detection algorithm [20] then identifies the largest clique of distances common to

TABLE 1
ALADDIN DEFAULT LOCATION OF THE SITE POINTS

Ligand atoms Y and X on which the site point Z is based (in order)	Y...Z length (Å)	X-Y-Z angle (°)	W-X-Y-Z torsion	No. of site points
Site hydrogen-bond donor points				
O=C in carbonyl, carboxylic acid, or carboxylate	2.9	120.0	in plane	2
N-C in 5-membered aromatic ring. Both N's of imidazole are used	3.0	125.0	in plane	1
N-C in 6-membered aromatic ring	3.0	120.0	in plane	1
O(H)-C (aromatic) or O^--C (aromatic)	2.9	119.0	in plane	2
O(H)-C(sp3)-C(sp3)	2.9	117.0	60°, 180°, 300°	3
O(H)-C(sp3)-other	2.9	109.5	along O-H and O-lp[a] directions	3
O(C)C	2.9	109.5	along O-lp[a] directions	2
$N(H_2)$-C-C	2.9	110.0	60°, 180°, 300°	3
$N(H_2)$-C-other	2.9	109.5	along N-H and N-lp[a] directions	3
N(H) (C)-C	2.9	109.5	along N-H N-lp[a] directions	2
N(C) (C)-C	2.9	109.5	along N-lp[a] direction	1
Site hydrogen-bond acceptor points				
N(H)C=O	3.0	120.0	in plane	1 per H
N(H)-C in 5-membered aromatic ring. Both N's of imidazole are used	3.0	125.0	in plane	1
N(H)-C, 6-membered aromatic ring	2.9	120.0	in plane	1
O(H)-C(aromatic)	2.9	119.0	in plane	2
O(H)-C(sp3)-C(sp3)	2.9	117.0	60°, 180°, 300°	3
O(H)-C(sp3)-other	2.9	109.5	along O-H and O-lp[a] directions	3
N=C	3.0	120.0	in plane	1 + no. H's
N^+=C	3.0	120.0	in plane	1 + no. H's
N-C(N)=N (guanidinium)	3.0	120.0	in plane	no. H's
$N(H_2)$-C(sp3)-C (sp3) or $N(H_3^+)$-C(sp3)-C(sp3)	2.9	110.0	60°, 180°, 300°	3
$N(H_2)$-C(sp3)-other or $N(H_3^+)$-C(sp3)-other	2.9	109.5	along N-H (and N-lp[a]) directions	3
N(H) or $N(H_2^+)$ (secondary)	2.9	109.5	along N-H (and N-lp[a]) directions	2
N or N (H^+) (tertiary)	2.9	109.5	along N-lp[a] or NH directions	1
N^+(quat)	3.6	180.0	along the C-N bond farther from the C. Only points that are at least 2.0 Å from every other atom are included	< 4

[a] Lone pair (lp) vectors are placed symmetrically in the structure.

the reference XYZ set and each other 3D structure. It then forms union sets for the cliques of each molecule. If every union set meets the set-up criteria, then DISCO identifies the cliques (individ-

ual conformations) that meet the set-up conditions.

At this point DISCO considers chirality. If there are only three points in the clique, or if all points are in a plane, both enantiomers of the test points will superimpose identically on the points of the reference. However, if the clique contains four points not in a plane, the two enantiomers will not superimpose identically. DISCO calculates the torsion angle of four non-planar points for the reference and test conformations. If a molecule is not labeled as chiral in the set-up file, DISCO accepts the clique and outputs the value of the torsion for consideration in later molecular graphics analysis. However, if the user specified that the molecule is chiral, DISCO keeps a clique only if the torsion angles from the points of the reference and test conformations are within a certain tolerance.

If the cliques meet the set-up criteria a pharmacophore map is found. Any one solution contains all conformations that match the same conformation and points of the reference molecule. A solution may include several conformations of the other molecules. There may be more than one solution, each using different points of the reference molecule, for any reference conformation.

If DISCO finds no solution and the tolerance is below the maximum, it increases the tolerance and repeats the process starting with recalculating the correspondence tables. It continues until it reaches the maximum tolerance or finds a solution.

After DISCO has studied the first conformation of the reference molecule, it repeats the entire process using each additional conformation of the reference molecule in turn. For each solution, DISCO reports the tolerance, conformation, atoms, and torsion of the reference molecule. For each test molecule it reports the conformation, matching atoms, and torsion. Any 3D structure, except for conformations of the reference molecule can appear in several different solutions and any compound can have several conformations in a solution. The results are compared with other DISCO runs with different set-up criteria and are evaluated with molecular graphics.

Implementation

Several important features were added to ALADDIN to facilitate the implementation of DISCO. ALADDIN can now read a file formatted for loading into a Daylight database [29] or it can read a list of SMILES [30] and generate the 3D structures with CONCORD [21]. Thus the structures processed for DISCO input need not be in a 3D database. We extended the Aladdin Control Language (ACL) [24] so that now, as well as lone pairs, the user can describe a point to be the location of a hydrogen-bond acceptor or donor. An advantage of using ALADDIN is that the same ACL file that was used to set up the DISCO input can be used for 3D searching [3,4] of databases with the exception that the constraints implied by the selected pharmacophore model would be added.

DISCO is written in C. We use it on computers running both VMS and UNIX. At this time, ALADDIN runs only on computers using VMS.

APPLICATIONS AND RESULTS

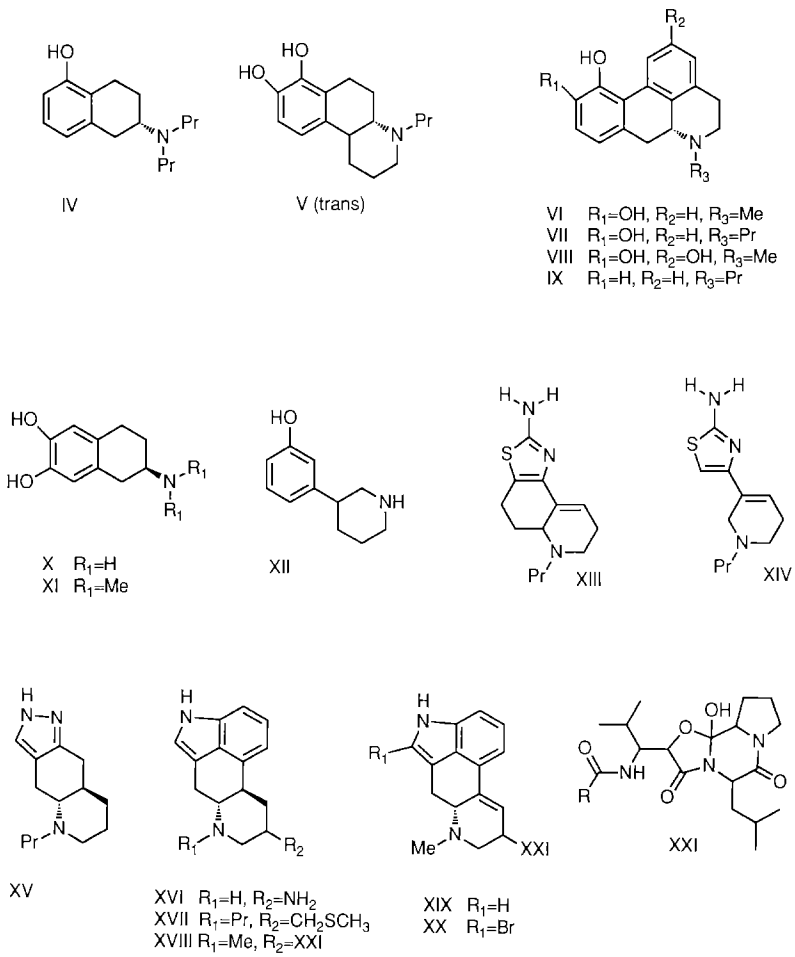
Application to dopamine agonists

To gain experience with DISCO we first tested it on molecules for which we had previous modeling experience [32], dopamine agonists, structures I–XXI (see Fig. 1 and Scheme 1). We

included compounds that Seeman et al. reported as dopamine agonists with an affinity of < 5 nM [33]. We also added three dopaminergics of novel structure (II, XII, and XIV) to see if they also match the others [34].

Since these are conformationally constrained molecules, we generated the 3D structures (one per molecule) with CONCORD [21,v2.9.1]. Because we generated only one conformation per molecule, the choice of the reference molecule does not affect the superposition rule although it may influence the tolerance slightly.

The only functional groups common to all are a hydrogen-bond donor attached to an aromatic ring and a basic nitrogen. For the pharmacophore we considered the location of (1) site hydrogen-bond acceptors of the (protonated) basic nitrogen, (2) site hydrogen-bond acceptors of the donors on the aromatic ring, (3) basic nitrogen atoms, (4) hydrogen-bond donor atoms, and (5) the center of mass of the aromatic ring. Default ALADDIN locations for site points were used.



Scheme 1. Structures IV–XXI.

For these five types of point with IV as the reference compound and a clique of size 4 specified, DISCO found one solution at a tolerance of 0.9 Å that included the site points and the hydrogen-bond donor and basic nitrogen atom points. On the other hand, when only the site points were required in the solution, the tolerance was 0.8 Å. In this case there were two superpositions that each include one or the other of the ligand points. Finally, when the ligand atoms were required to be present the 0.9 Å four-point solution was found. Thus, the ligand atoms superimpose slightly less well than do the site points, and the centers of mass superimpose not at all well.

Figure 2A shows the superpositions of two representative phenols and an indole; Fig. 2B shows the addition of the 2-aminothiazoles. The latter compounds thus fit the same pharmacophore map as do the phenols and indoles. The superpositions shown in Fig. 2A are similar to most models of D2 pharmacophores in the literature [32].

In the four-point solution DISCO found two superpositions for every catechol. This results from the fact that two of the four site hydrogen-bond acceptor points are located only 0.21 Å apart – one originates from the oxygen involved in the superpositions shown, and the other from the other oxygen. Superpositions using the alternative site points are almost indistinguishable in molecular graphics. This may be the way to include 6-hydroxy-2-amino tetralins, which bind weakly, in the model.

We also looked for superpositions that considered points in the molecule only, including the center of mass of aromatic rings. When only the indoles and phenols were considered there was a superposition at a tolerance of 1.00 Å. It is shown in Fig. 3. Notice that the direction of the postulated interaction of the basic nitrogen with the receptor is different in XVI compared to IV and XI. (If we did not know the absolute stereochemistry of the active enantiomer of these compounds, we could just take the enantiomer of XVI for the solution). The tolerance for this superposition rises to 1.8 when the aminothiazoles are also included. This example shows the attractiveness of considering site points in models of pharmacophores, in knowing the absolute stereochemistry of the active enantiomer, and of having diverse structures in the dataset.

The VAX-6310 cpu timings for this investigation were: (i) generating the CONCORD structures, 0.50 min; (ii) preparing the DISCO input with ALADDIN, 0.88 min; (iii) each DISCO iteration, ~ 10 s for 21 molecules with six types of point. Thus the computer requirements for an investigation with this type of molecule are minimal.

Application to benzodiazepine agonists

Our second example includes a more structurally diverse set of compounds, some of which have more than one conformation. We asked if DISCO would reproduce literature superpositions for benzodiazepine receptor ligands [35–40] and if it would suggest alternative ways to superimpose these compounds.

This dataset consists of the compounds shown in Fig. 4. Each is reported to be a full or partial agonist with an IC_{50} for inhibition of [^{3}H]benzodiazepine binding < 150 nM [35]. XXVII and XXVIII are the most structurally different of the series and it has been reported that they bind to a site on the benzodiazepine receptor different from that of the other compounds [40]. In spite of this report, we included them so that we could compare DISCO superpositions to literature superpositions [35].

We used crystal structures from the Cambridge Structural Database [22] as a starting conformation for each compound except XXIV and XXV. For these a crystal structure was not avail-

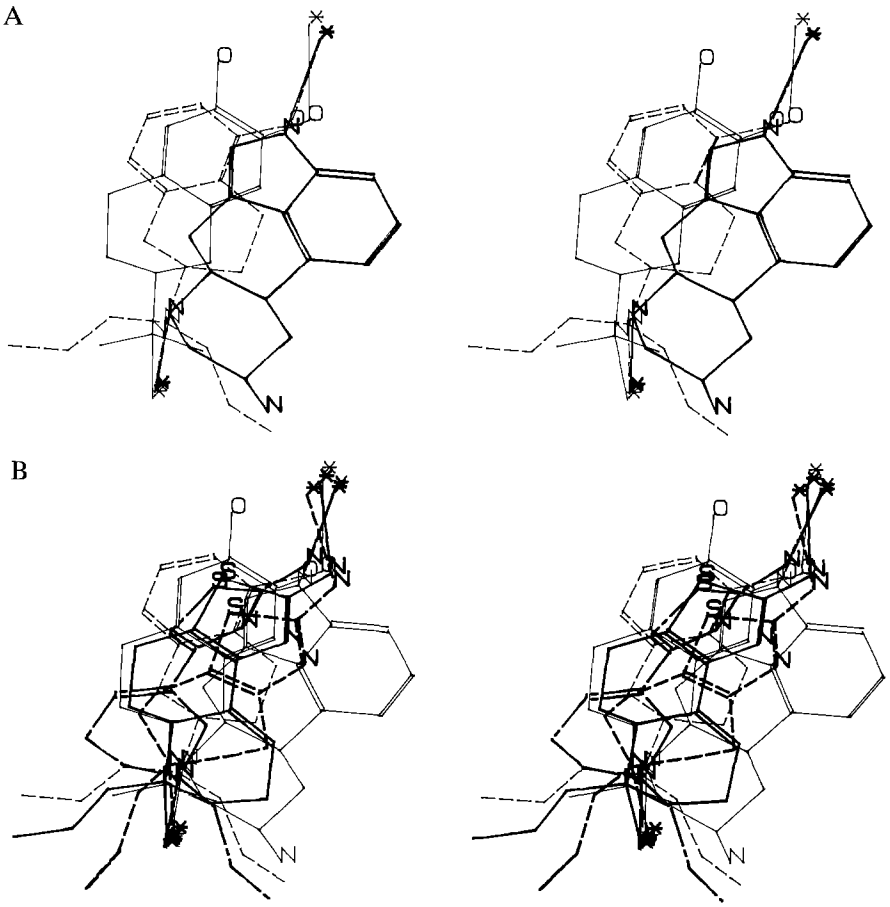


Fig. 2. Superpositions based on two site and two ligand points: (A) The superposition of IV (dashed fine lines), XI (solid fine lines), and XVI (heavy lines). The site points are labeled with an asterisk. (B) The superposition of IV (dashed fine lines), XI (solid fine lines), XVI (solid fine lines), II (dashed heavy lines), XIII (solid heavy lines), and XIV (dotted heavy lines).

able, so the starting conformation was generated with CONCORD [21, v2.9.1]. The conformation with a different ring pucker of the diazepine ring was generated for compounds XXII and XXIII. For XXVI–XXVIII, several low-energy conformations (XXVI: six conformations; XXVII: three; XXVIII: six) were selected by REJECT (using tolerances of 0.6–0.8 Å) from those generated by distance geometry [9].

All starting conformations were minimized in AMPAC, using the AM1 Hamiltonian (with GNORM = 5.0) [41]. We examined in DISCO all conformations within 5 kcal/mol of the lowest energy conformation observed. The conformations used were a good representation of conformational space.

Literature pharmacophore models of benzodiazepine agonists generally superimpose two hydrogen-bond acceptor atoms and a lipophilic region [35,37–39]. For example, in XXII the hydrogen-bond acceptors are the carbonyl oxygen and the doubly-bonded nitrogen, while the pendant phenyl ring represents the lipophilic region [38].

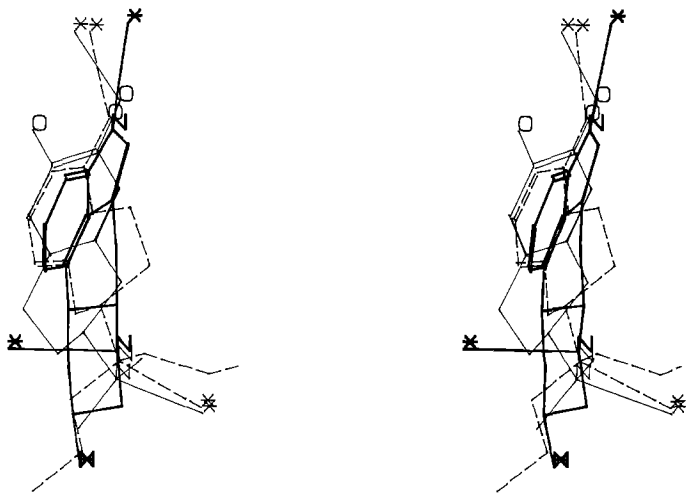


Fig. 3. Superpositions based on ligand points only; molecules distinguished as in Fig 2A.

We selected the locations of three types of points for DISCO examination. The first type of point was for hydrogen-bond acceptors of the ligand. This includes carbonyl oxygen atoms, aromatic nitrogen atoms without an attached hydrogen, and doubly bonded aliphatic nitrogen atoms. The second type of point was for site hydrogen-bond donors, located at the default ALADDIN locations. The third type of point was for lipophilic binding regions. These were located at the centers of mass of six-membered aromatic and heteroaromatic rings. For XXVII and XXVIII, Borea and coworkers suggested that the piperazine ring served this function [35], so a point at the center of mass of this ring was also included as a lipophilic region. The result is that for each compound there are several occurrences of each type of point and five of the seven had more than one low-energy conformation. This was a good case to investigate the utility of DISCO since there are hundreds of possible combinations of bioactive conformations in the pharmacophore map and five of the seven molecules have more than one possible orientation to XXII.

We used XXII as the reference compound for DISCO because it has two representative conformations and we were interested if they lead to different bioactive conformations or superposition rules for the other compounds.

In preliminary DISCO runs we varied the numbers and types of required points. When we used only the ligand points, there were solutions in which the atoms were superimposed well, but the remainder of the structure was not (see Fig. 5). As the subsequent discussion will demonstrate, better results were obtained when one considers the location of the site hydrogen-bonding points, not just the location of the ligand points. The discussion below highlights two representative DISCO runs.

In one run, we required that solutions contain one ligand hydrogen-bond acceptor point, two site hydrogen-bond donor points, and one lipophilic binding point. This model was investigated to verify the importance of including site hydrogen-bond donor points. DISCO found one proposed solution at a tolerance of 1.6 Å for each of the two input conformations of XXII. These two solutions yield the same pharmacophore map, since the only difference in the results is the conformation of XXII.

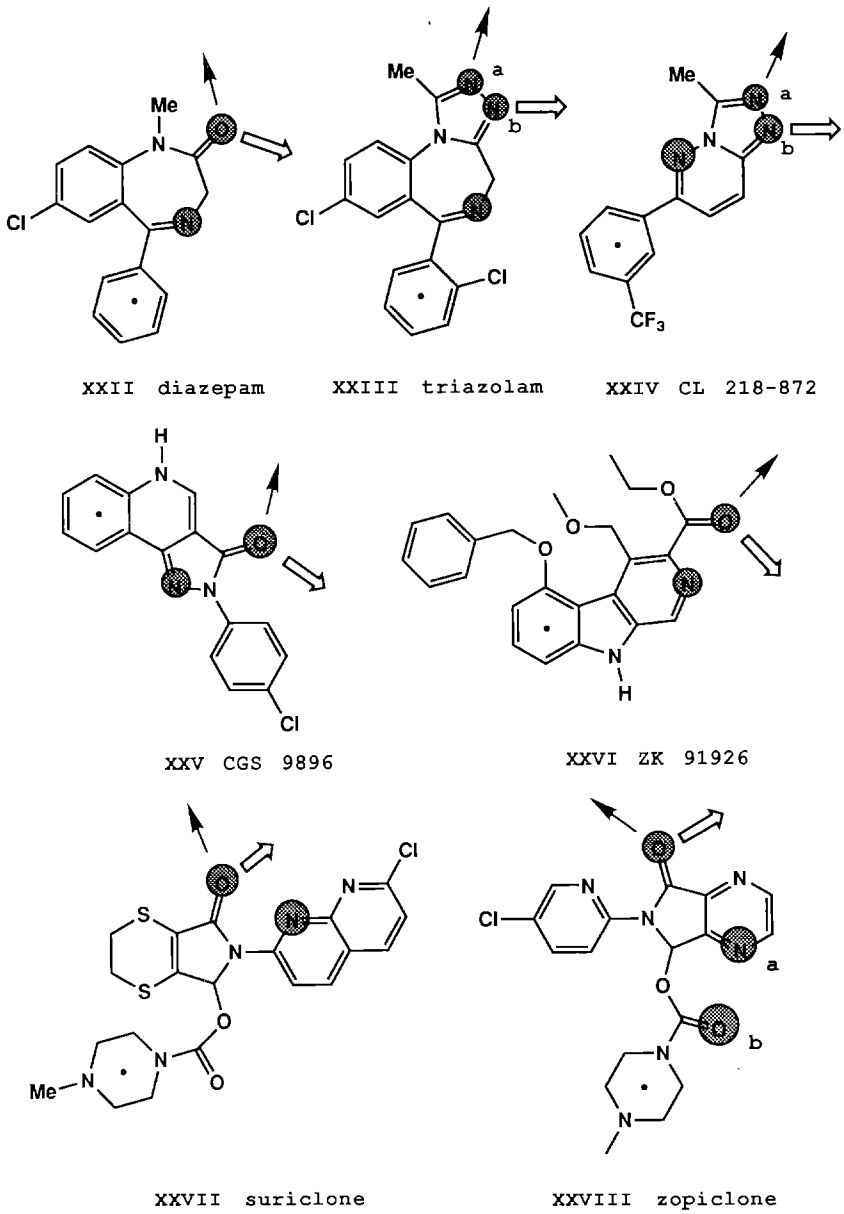


Fig. 4. The seven benzodiazepine receptor agonists studied with DISCO. The identity of the points for the superposition rule proposed by DISCO is shown also. The solution is a five-point model: shaded circles represent ligand hydrogen-bond acceptors; the open and filled arrows represent the site hydrogen-bond donors; and the filled dot is the center-of-mass of a lipophilic binding region. Labels a and b refer to alternate superposition rules.

However, DISCO found that several of the compounds had multiple conformations and orientations that fit this solution. This suggested that a proposed map containing additional points might eliminate some of these superpositions and result in a better model. Because run times for

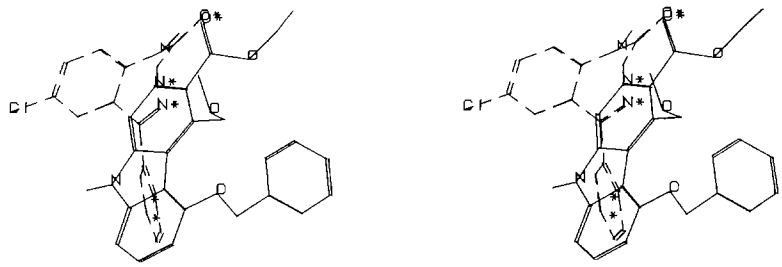


Fig. 5. A stereoview of a DISCO superposition in which the identified atoms superimpose whereas the overall superposition is poor. Compound XXVI (solid lines) is superimposed on the reference compound, XXII (dashed lines). This superposition uses only ligand points, labeled with an asterisk. This type of superposition could be avoided by including additional points, particularly site hydrogen-bond acceptors.

DISCO are minimal (typically ca. 1 cpu min on a VAX-9000), alternate maps can be generated interactively.

Therefore, we evaluated a five-point model, comprised of the four-point model described above, plus an additional ligand hydrogen-bond accepting point. Two ligand hydrogen-bond accepting points were used to match the two corresponding points in the reference compound, XXII. Specifically, we wanted to see if a solution containing both the carbonyl oxygen and the doubly-bonded nitrogen atom in XXII, would be found with DISCO.

Using the five-point model, DISCO found one solution at 1.8 Å for each of the two input conformations of XXII. Again, these two solutions resulted in the same pharmacophore map, the only difference being the conformation of XXII. The remainder of this discussion therefore focuses on pharmacophore maps based on the minimized crystal structure of XXII.

In addition, while several conformations of XXVI–XXVIII were considered by DISCO, the higher energy conformations did not result in additional sensible solutions based on overall superposition. Therefore, each solution discussed below uses a conformation that is within 2 kcal/mol of the lowest energy conformation generated.

The 2D identity of the pharmacophore points in the DISCO superposition is shown in Fig. 4. The shaded circles denote the ligand hydrogen-bond acceptors, the filled and open arrows alternative site hydrogen-bond donors, and the filled dot the center of mass of the lipophilic region. Selected superpositions using these points are presented in Fig. 6. Figures 4 and 6 show that four of the compounds had only one superposition rule, while compounds XXIII, XXIV, and XXVIII had two (labeled a and b in Fig. 4). Thus, the five-point model significantly reduced the number of superpositions for several of the compounds, compared to the four-point model discussed above. It is instructive to discuss selected compound's DISCO superpositions with XXII in turn.

Compounds XXIII and XXIV have similar superpositions. Although they have only one conformation, each has two superpositions. In particular, either of the adjacent shaded nitrogen atoms in Fig. 4 can serve as one of the ligand hydrogen-bond acceptors. In both superpositions, the calculated site hydrogen-bond donor point of both these atoms is present in the superposition. That is, the filled arrow in XXIII corresponds to one calculated site point that interacts with the carbonyl oxygen in XXII, while the open arrow corresponds to the other. The second ligand hydrogen-bond acceptor is the shaded unlabeled nitrogen atom in Fig. 4. Thus, in the superposition of compounds XXIII and XXIV, both the ligand points and two site points are superim-

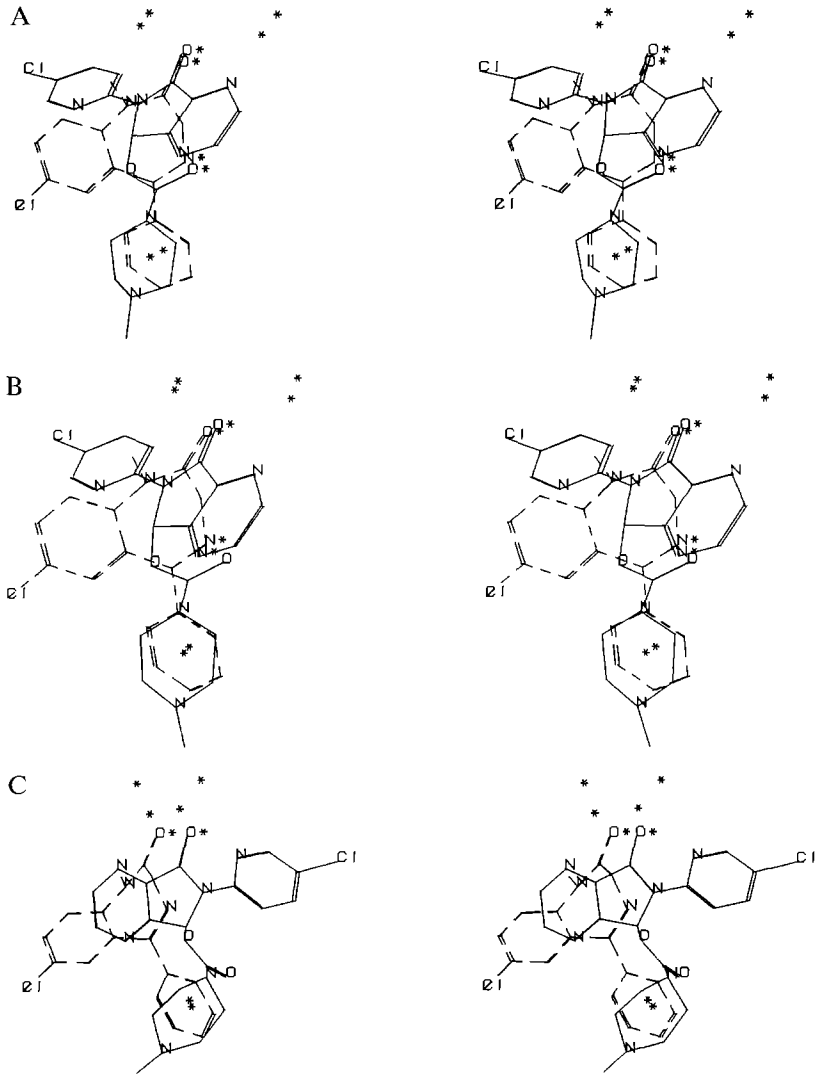


Fig. 6. Stereoviews of selected superpositions for one benzodiazepine agonist using points (labeled with an asterisk) identified by DISCO. (A) shows a superposition for the figures shown in 2D in Fig. 4 for XXII (dashed lines) and XXVIII (solid lines); (B) shows the alternate superposition, b, for XXVIII; (C) shows a superposition for XXVIII, generated using a four-point model.

posed. This is in contrast to the traditional manner of superimposing XXIII on XXII in which only the ligand acceptor points are superimposed directly (see Ref. 36 for a related example). In our opinion, the DISCO solution is attractive because it aligns both site hydrogen-bond donor points. Lastly, in XXIII and XXIV the center of mass of the pendant phenyl ring is proposed as the fourth overlapping point.

The DISCO superposition for XXV, which explicitly includes the site hydrogen-bond donor points, appears to provide a better overall superposition than that presented by Borea et al. [35] (no picture shown).

Compounds XXVI–XXVII each have only one superposition. Interestingly, using the four-point model above, which required only one ligand acceptor point, two superpositions were found for XXVI and XXVII. Adding an additional ligand acceptor point eliminated one of these superpositions in both compounds. In addition, the one superposition found for XXVI and XXVII in the five-point model is better overall compared to the two superpositions resulting from the four-point model.

For XXVIII, DISCO selected the indicated carbonyl oxygen as one of the ligand hydrogen-bond acceptor points. An aromatic nitrogen atom or a second carbonyl oxygen serves as the second ligand point. The superposition rule containing two carbonyl oxygen ligand points (Fig. 6A) results in a slightly better overall superposition, compared to the alternate superposition (Fig. 6B). It had been suggested that the piperazine ring serves as the lipophilic binding region in these two compounds [35], and DISCO also proposes this.

The four-point model produced an additional superposition for XXVIII, in which one site hydrogen-bond donor point comes from the pyridine nitrogen. As can be seen in Fig. 6C, this is not a good overall superposition. Thus, the five-point model eliminated this superposition and generated two better superpositions.

The DISCO runs for this example required approximately 1 min of cpu time on a VAX-9000 (approximately 40 times as fast as a VAX-11/780). The ALADDIN runs to prepare the DISCO input also used about 1 min of cpu time on a VAX-9000. The most cpu intensive part of the procedure was the conformational analysis, requiring about five VAX-9000 cpu min for a distance geometry (100 conformations generated) run on a flexible compound, and about 30 min of VAX-9000 cpu time for an AM1 full-geometry optimization.

This study is representative of the type of results obtained from DISCO and of the most fruitful manner in which to use this tool. Because DISCO generates solutions rapidly, it allows one to explore several proposed pharmacophore maps for each set of structures in a short period of time. Therefore, DISCO can provide the most information when it is run several times in succession, using the current solutions to guide the choice of input criteria for the next run. To do this, the user may use molecular graphics to view the superpositions from each run's solutions, noting both the degree of overlap of the proposed pharmacophore points and the fit of each compound.

Overall, this application of DISCO illustrates the program's ability to propose superposition rules for structurally diverse compounds. When a compound has several occurrences of each type of potential pharmacophore point and several low-energy conformations, DISCO can be useful in finding which set of points and conformations result in a good superposition for the series. In addition, because site binding points were included, DISCO proposed chemically and sterically better ways of superimposing several of the benzodiazepine agonists than some of the literature models.

DISCUSSION

Alternative pharmacophore mapping strategies

Before discussing DISCO in more detail it is useful to highlight selected efforts of other researchers in the field of pharmacophore mapping. Crippen [11] considered how to superimpose molecules of different structure as part of a 3D QSAR method based on distance geometry. His matching algorithm is a simple brute-force investigation of all pairings of distances between

points of similar nature. He has found that some compounds appear to have multiple possible binding modes (different atom pairings and possibly different conformations). In an investigation of thyroxin analogues he found several geometries of the pharmacophore points that are consistent with the structure–activity data. Thus, others have seen multiple pharmacophore maps consistent with a set of structure–activity data.

From our viewpoint, Crippen's strategy has several deficiencies. It stops with the first set of matching points at a given tolerance instead of identifying all superpositions. Only at the 3DQSAR stage are alternative binding modes of individual compounds added back to the consideration. A more serious deficiency is that the strategy does not consider the relative energy of the conformations. The final distance geometry generated binding conformations and orientations might be high in energy.

There are two literature computational strategies that help one propose the bioactive conformations for a pharmacophore map based on a user-supplied superposition rule. The earliest is the active analogue approach [1]. This strategy looks for inter-point distances common to all active molecules by systematic rotation of bonds. Unfortunately, the number of superpositions and the bioactive conformations found can depend on the tolerance at which the user specifies that two distances are the same.

Ensemble distance geometry is a more recent strategy for exploring bioactive conformations given a superposition rule [42]. This program generates ensembles that contain one superimposed conformation of each molecule. A drawback with this methodology is that many ensembles contain high-energy conformations of one or more molecules. When the structures are minimized they may no longer match the original superposition criteria. Additionally, some of the conformations generated are duplicates. Therefore, one must map the relationships of the ensembles to each other.

Neither the active analogue nor ensemble distance geometry approaches to pharmacophore mapping pay explicit attention to the relative energies of the proposed bioactive conformation. In addition, both must be repeated from the beginning to consider an alternative point pairing.

The strategy illustrated in this report has several advantages over the active analogue approach and ensemble distance geometry generation of bioactive conformations. First, the user does not supply the atom or point correspondences between molecules, but is provided this information by the program. Second, one considers only low-energy conformations. An advantage over the active analogue approach is that one can use any method of structure optimization, even quantum mechanics. An advantage over ensemble distance geometry is that DISCO identifies all of the matching conformations of different compounds even if several conformations of one molecule match one (or more) conformations of another. Finally, the potentially time-consuming step of conformational exploration is done only once, rather than being done for each potential pharmacophore map.

The disadvantage of our strategy is that the conformational search is not restricted. Hence if one compound is conformationally constrained but others are quite flexible; many conformations of the flexible compounds might be generated and minimized only to be later discarded.

DISCO

Our results document that the proposed strategy can quickly find superposition rules for sets of molecules. Although it operates on distances, a simple test adds chiral information. The

strategy has the advantage that it is applicable to molecules of any structure and measure of relative energies of the conformations. DISCO alerts the user to alternative superpositions at the same tolerance, and to more complex superpositions at higher tolerances. These alternative superpositions are useful in suggesting different molecules to synthesize.

The proposed superpositions suggested by DISCO are suggestions that one may choose to improve by using more sophisticated criteria such as optimizing the overlap of the total fields surrounding a pair of molecules. One may also choose to add additional points to the superpositions of some of the molecules to take advantage of apparent secondary points such as the hydrogen-bond accepting atom and its site donor point in XXII and XXV. The user may also decide to include other atoms in the final orientation rule to better match features such as overall shape or to overlap features found in some molecules but not in all.

One need not use the default ALADDIN locations of site points. One could instead use the location of minima or maxima in the electrostatic potential surrounding the molecule [43]. This would be especially useful for functional groups for which there is not crystal packing information from a variety of structures. The disadvantage is that electrostatic potential is sensitive to minor conformational changes such as rotation of a hydroxyl group.

We do not explicitly calculate relative energies of the proposed interactions when selecting a favored superposition. For example, an ammonium group forms a stronger hydrogen bond than does a hydroxyl. We would thus prefer a superposition in which the hydrogen bond donor is ammonium over one for the same molecule in which the donor is hydroxyl. We handle this now at the molecular graphics stage, but could handle it in DISCO by a simple weighting scheme.

We have had certain proposed pharmacophore maps for which DISCO found no solution even at a high tolerance. In one case there were 13 molecules, some with 60 conformations to be considered. The lack of a DISCO solution supplied objective information that there is no superposition that meets the criteria for the proposed model. This would have been difficult to prove with molecular graphics. We then used the same set of conformations to examine other pharmacophore hypotheses.

This method is not appropriate for sets of molecules each of which has thousands of low-energy conformations of different distances between pharmacophore points. If the set does contain molecules with different distance constraints between the proposed pharmacophore points, an alternative strategy, similar to that of Crippen [11] can be envisioned. In this case the input would be the distance range between possible pharmacophore points of different types. The clique-detection method would find the best match between these points and also find the distances common to all molecules. The output would be a distance matrix for each molecule and a correspondence list. These form the input to ensemble distance geometry. After energy minimization the conformations would be compared with DISCO. If there are still thousands of structures, then this set of data does not supply enough information for a pharmacophore mapping study.

Improvements and future plans for DISCO

If it were possible to include information on inactive compounds in the search the power of the method would be improved. Unfortunately, molecules are inactive either because they cannot assume the pharmacophore geometry or because in the potential complex they occupy space that is also occupied by the target biomolecule. One strategy to include information on inactive compounds is to contrast the results found with DISCO calculations including and excluding

inactive molecules. This might identify a pharmacophore map that the inactive molecules cannot attain. On the other hand, if the inactive compounds show the same distances as the active compounds, then one must rely on volume considerations. The multitude of potential superpositions supplied by DISCO will help one find the one or several consistent hypotheses.

If one were to calculate the reference points for DISCO from a 3D structure of a macromolecule and use only one test ligand, the results would be the orientation and conformation of the docking of the ligand with the macromolecule. For such a use, the site points would be coordinates of atoms in the macromolecule and the ligand points would be calculated by projection from them. This is similar to the strategy of Kuhl et al. [20]. One difference is that DISCO has routines that discard cliques based on criteria such as requirements for inclusion of certain points and chirality detection. This suggested that use of DISCO is also similar to the strategy of Kuntz et al. [44], who use the centers of spanning spheres to describe the shape of the binding site. In their case a different point-matching algorithm is used. Conformational flexibility can be added by considering the lower and upper bounds of allowed interatomic distances [45].

If one uses clique-based docking to examine molecules in 3D databases, the result is a 3D substructure searching program that can accept partial matches between the search query and the database molecule [45,46]. Such partial matches could be spliced together at overlapping points to invent molecules different from anything in the database. We expect that this strategy will be especially applicable to the computer design of molecules to fit a binding site of known but complex 3D structure.

We are also exploring a simple modification of this method to identify subsets of closely related analogues in the large set of hits identified or designed with 3D-substructure searching. Here the clique search is based on more points in the molecule and uses a smaller tolerance. The result is families of molecules that are quite similar in 3D. One could then screen or synthesize only one member of each family to explore a variety of ways to position the pharmacophore atoms. Once a hit was found, the other members of the family would be attractive candidates for further investigation.

CONCLUSION

DISCO provides the user with suggested bioactive conformations and superposition rules for a set of compounds. DISCO also allows for rapid evaluation of alternative hypotheses. These two factors make the program a powerful aid to pharmacophore mapping. However, the final choice of a pharmacophore model is still the judgment of the user who must balance considerations such as relative conformational energy, closeness of fit of the pharmacophore points, closeness of fit of the complete structures, and the angles made to site hydrogen bonds. Knowing which choices, as provided by DISCO, to examine with molecular graphics saves false starts. Further development of tools that assist pharmacophore mapping will continue to be important to advance the design of bioactive molecules.

ACKNOWLEDGEMENTS

We appreciate the helpful discussions with Peter Willett on the characteristics of clique-detection methods and the implementation of the Bron-Kerbosh algorithm.

REFERENCES

1 Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.A. and Dunn, D.A., In Olson, E.C. and Christoffersen, R.E. (Eds.) Computer-Assisted Drug Design, American Chemical Society Symposium, No. 112, ACS, Washington, 1979, pp. 205–226.
2 Martin, Y.C., Methods Enzymol., 203 (1991) 587.
3 Martin, Y.C., J. Med. Chem., 35 (1992) 2145.
4 Martin, Y.C., Bures, M.G. and Willett, P., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry, VCH Publishers, New York, 1990, pp. 213–263.
5 Murray-Rust, P. and Glusker, J.P., J. Am. Chem. Soc., 106 (1984) 1018.
6 Kelly, J.A. and Knox, J.R., In Jensen, B., Jorgensen, F.S. and Kofod, H. (Eds.) Frontiers in Drug Research – Crystallographic and Computational Methods, Alfred Benzon Foundation, Copenhagen, 1990, p. 252.
7 Ippolito, J.A., Alexander, R.S. and Christianson, D.W., J. Mol. Biol., 215 (1990) 457.
8 Taylor, R. and Kennard, O., Acc. Chem. Res., 17 (1984) 320.
9 Crippen, G.M. and Havel, T.F., In Bawden, D. (Ed.) Chemometrics Research Studies Series, Research Studies Press, Wiley, New York, 1988. We use DGEOM by Blaney, J., Crippen, G.M., Dearing, A. and Dixon, J.S. from QCPE, program number 590.
10 Brint, A.T. and Willett, P., J. Chem. Inf. Comput. Sci., 27 (1987) 152.
11 Crippen, G., J. Med. Chem., 22 (1979) 988; 23 (1980) 599; 24 (1981) 198.
12 Crandell, C.W. and Smith, D.H., J. Chem. Inf. Comput. Sci., 23 (1983) 186.
13 Danzinger, D.J. and Dean, P.M., J. Theor. Biol., 116 (1985) 215.
14 Namasivayam, S. and Dean, P.M., J. Mol. Graphics, 4 (1986) 46.
15 Kato, Y., Itai, A. and Iitaka, Y., Tetrahedron, 43 (1987) 5229.
16 Chau, P.-L. and Dean, P.M., J. Mol. Graphics, 5 (1987) 88, 97.
17 Dean, P.M. and Chau, P.-L., J. Mol. Graphics, 5 (1987) 152.
18 Dean, P.M., Callow, P. and Chau, P.-L., J. Mol. Graphics, 6 (1988) 28, 38.
19 Hermann, R.B. and Herron, D.K., J. Comput-Aided Mol. Design 5 (1991) 511.
20 Kuhl, F.S., Crippen, G.M. and Friesen, D.K., J. Comput. Chem., 5 (1984) 24.
21 Pearlman, R.S., Rusinko III, A., Skell, J.M., Balducci, R. and McGarity, C.M., CONCORD, Distributed by Tripos Associates, Inc., 1969 S. Hanley Road, Suite 303, St Louis, MO 63944, U.S.A.
22 Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrea, C.F., Mitchell, E.M., Mitchell, G.F., Smith, J.M. and Watson, D.G., J. Chem. Inf. Comput. Sci., 31 (1991) 187.
23 Martin, Y.C. and Rys, J., unpublished program.
24 Van Drie, J.H., Weininger, D. and Martin, Y.C., J. Comput.-Aided Mol. Design, 3 (1989) 225.
25 Thanki, N., Thornton, J.M. and Goodfellow, J.M., J. Mol. Biol., 202 (1988) 637.
26 Boobbyer, D.N.A., Goodford, P.J., McWhinnie, P.M. and Wade, R.C., J. Med. Chem., 32 (1989) 1083.
27 Vedani, A. and Dunitz, J.D., J. Am. Chem. Soc., 107 (1985) 7653.
28 Jeffrey, G.A. and Saenger, W., Hydrogen Bonding in Biological Structures, Springer-Verlag, Berlin, 1991.
29 Daylight Chemical Information Systems, Inc., 1991, 3951 Claremont St. Irvine, CA 92714.
30 Weininger, D. and Weininger, A., J. Chem. Inf. Comput. Sci., 28 (1988) 31.
31 Martin, Y.C., Kim, K.-H. and Bures, M.G., In Wermuth, C.G. (Ed.) Medicinal Chemistry in the 21st Century, Blackwell Scientific Publ., Oxford, 1992, pp. 295–317.
32 Martin, Y.C. and Danaher, E.B., In Williams, M., Glennon, R. and Timmermans, P. (Eds.) Receptor Pharmacology and Function, Marcel Dekker, New York, 1988, pp. 137–171.
33 Seeman, P., Watanabe, M., Grigoriadis, D., Tedesco, J.L., George, S.R., Svensson, U., Lars, J., Nilsson, G. and Neumeyer, J.L., Mol. Pharmacol., 28 (1985) 391.
34 Caprathe, B.W., Jaen, J.C., Wise, L.D., Heffner, T.G., Pugsley, T.A., Meltzer, L.T. and Parvez, M., J. Med. Chem., 34 (1991) 2736.
35 Borea, P.A., Gilli, G., Bertolasi, V. and Ferretti, V., Mol. Pharmacol., 31 (1987) 334.
36 Codding, P.W. and Muir, A.K.S., Mol. Pharmacol., 28 (1985) 178.
37 Crippen, G.M., Mol. Pharmacol., 22 (1982) 11.
38 Loew, G.H., Villar, H.O., Jung, W. and Daview, M.F., In Rapaka, R.S., Makriyannis, A., Kuhar, M.J. (Eds.)

National Institute on Drug Abuse Research Monograph Series, 112, U.S. Department of Health and Human Services, 1991, pp. 43–61.
39 Tebib, S., Bourguignon, J.-J. and Wermuth, C.-G., J. Comput.-Aided Mol. Design, 1 (1987) 153.
40 Trifiletti, R.R. and Snyder, S.H., Mol. Pharmacol., 26 (1984) 458.
41 AMPAC, version 2.1 (QCPE No. 506), available from Quantum Chemical Program Exchange, Indiana University, Bloomington, IN, U.S.A.
42 Sheridan, R.P., Nilakantan, R., Dixon, J.S. and Venkataraghavan, R., J. Med. Chem., 29 (1986) 899.
43 Weiner, P.K., Langridge, R., Blaney, J.M., Schaefer, R. and Kollman, P.A., Proc. Natl. Acad. Sci. USA, 79 (1982) 3754.
44 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T., J. Mol. Biol., 161 (1982) 269.
45 Smellie, A.S., Crippen, G.M. and Richards, W.G., J. Chem. Inf. Comput. Sci., 31 (1991) 386.
46 Moon, J.B. and Howe, W.J., Tetrahedron Comput. Methodol., 3 (1990) 697.