JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

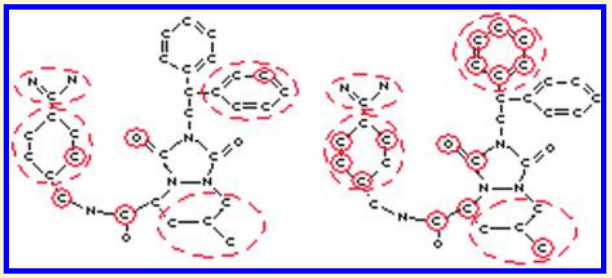
pubs.acs.org/jcim

Modeling Flexible Pharmacophores with Distance Geometry, Scoring, and Bound Stretching

Michael Binns,[†] Sam P. de Visser,[†,‡] and Constantinos Theodoropoulos*[,†]

[†]School of Chemical Engineering and Analytical Science and [‡]Manchester Interdisciplinary Biocentre, University of Manchester, Manchester M13 9PL, United Kingdom

ABSTRACT: The study of pharmacophores, i.e., of common features between different ligands, is important for the quantitative identification of "compatible" enzymes and binding species. A pharmacophore-based technique is developed that combines multiple conformations with a distance geometry method to create flexible pharmacophore representations. It uses a set of low-energy conformations combined with a new process we call bound stretching to create sets of distance bounds, which contain all or most of the low-energy conformations. The bounds can be obtained using the exact distances between pairs of atoms from the different low-energy conformations. To avoid missing conformations, we can take advantage of the triangle distance inequality between sets of three points to logically expand a set of upper and lower distance bounds (bound stretching). The flexible pharmacophore can be found using a 3-D maximal common subgraph method, which uses the overlap of distance bounds to determine the overlapping structure. A scoring routine is implemented to select the substructures with the largest overlap because there will typically be many overlaps with the maximum number of overlapping bounds. A case study is presented in which 3-D flexible pharmacophores are generated and used to eliminate potential binding species identified by a 2-D pharmacophore method. A second case study creates flexible pharmacophores from a set of thrombin ligands. These are used to compare the new method with existing pharmacophore identification software.



1. INTRODUCTION

The *pharmacophore* is defined as the shape and structure required for a molecule to bind to a certain active site.[1,2] There are different methods available for the identification of pharmacophores[3−14] (in addition to related methods for using pharmacophores to search databases[15−17]). The computation or elucidation of pharmacophores allows alternative binding species to be identified. The degree to which a molecule matches the pharmacophore can be used to predict if the molecule will bind and (possibly) to give a predicted binding affinity. In this respect, pharmacophores are important because they offer a computationally convenient way to quantitatively assess which molecules can bind to an enzyme, including potential inhibitors and activators. Hence, the study of pharmacophores can significantly aid in the identification of new drugs that are typically aimed at inhibiting a given enzyme and altering the properties a biological process (e.g., removing a negative effect caused by a disease). Pharmacophores can be constructed directly from experiments that show the molecules docked into the binding site. However, when this is unavailable, a pharmacophore can be computed by comparisons of the ligands that bind to the site. The simplest comparisons involve 1-D and 2-D methods that can identify important features relatively quickly. A 1-D comparison method can only determine common substructures on the basis of the *total number* and the *type* of atoms/functional groups present in each molecule.

The 2-D methods add a topological level of information by considering in addition which atoms, within the molecule, are bonded together and requiring that the common substructure maintain those bonds connecting the common atoms/functional groups. A 3-D method takes into account the size and shape (including the distances between atoms) of the molecule (in addition to the information considered in 1-D and 2-D methods) that allows for a more specific pharmacophore to be identified. By computing more tightly specified pharmacophores, 3-D methods can lead to the identification of more accurate sets of "compatible" binding species compared to 1D and 2-D methods. So for example, compounds that match the 1-D pharmacophore may have atoms connected incorrectly when compared to the 2-D or 3-D pharmacophore. Also, compounds that fit the 2-D pharmacophore may be unable to stretch and twist to fit the 3-D pharmacophore specifications without assuming a very unstable high-energy state (which would make binding in that state infeasible). Hence, the 3-D comparison method will filter out these infeasible compounds.

Many of the existing 3-D methods are intended to create one or more rigid 3-D pharmacophore conformations.[3−14] This assumes that one of the rigid structures found is the correct structure required for the binding (and possible reaction) and that alternative structures will not bind. To account for this possibility, some of the above methods can also be used repeatedly to generate multiple solutions with the anticipation that one of the solutions is the correct one, and some methods are designed to generate multiple solutions anyway.[5,7,12−14]

Received: September 20, 2011
Published: January 11, 2012

An alternative method is to use distance geometry to construct sets of bounds that contain all or most of the important solutions.[18,19] This approach is very promising because the bounds can be used to incorporate molecular flexibility and solutions can be found that are missed by methods using only a finite number of conformations. However, these methods have the drawback that the bounds will usually contain many high-energy or infeasible solutions.

A similar branch of methods have recently been developed using distance bins to group together conformations containing similar distances.[7,9,12] These methods efficiently filter the precomputed conformations to give common pharmacophores, but they do not consider solutions outside the precomputed set.

The method presented here belongs to this class of these distance geometry techniques and is improved with methods we call *bound stretching* and *overlap scoring* to make the most out of the available information. This has the advantage that through bound stretching our algorithm allows the consideration of a large degree of flexibility using only a small number of conformations. The algorithms use clique finding to find all possible pharmacophores that satisfy distance overlap conditions. A scoring routine is also implemented to select the pharmacophores with the greatest overlap. The pharmacophores generated are flexible in that the distances between features identified are specified by calculated upper and lower bounds that allows a range of possible 3-D conformations. These flexible pharmacophores can be used to identify new binding species and are shown to constrain and improve the results of our 2-D pharmacophore methodology,[20] which makes use of the program simcomp[21] to perform the 2-D comparisons.

2. MATERIALS AND METHODS

The new flexible pharmacophore techniques developed here are based on the ideas of distance geometry and involve generating distance bounds between every pair of features (atoms or functional groups) in each of the molecules involved (the ligands and the potential binding species). The flexible representations for the known binding ligands are then compared in order to extract a flexible pharmacophore that contains the atoms or functional groups and the overlapping distance range common to the known ligands.

The computation of distance bounds combining distance geometry with the generation of multiple conformations achieves more accurate bounds than a method using only distance geometry method (starting from standard bond angles and standard bond lengths). Combining the two techniques also means fewer conformations are required than for techniques accounting for flexibility only through the generation a large number of conformations.

To further reduce the number of conformations needed, a new method we call bound stretching based on the triangle inequality is used to logically increase the upper bounds and decrease the lower bounds. Also, in order to ensure the best pharmacophore is found, a new scoring routine based on the size of distance overlaps is added.

2.1. Generating Random Conformations. Generating conformations is the problem of converting 2-D molecular structures into 3-D ones. There are typically many different 3-D conformations for a single 2-D structure. For generating single conformations, there are packages such as Concord[22] and Corina.[23] However, in many cases it is desirable to generate an ensemble of different conformations, which characterize the flexibility of the molecules. It may be possible to modify Concord and Corina to perform this task; however, there are alternative methods and techniques that are designed for generating sets of conformations.[24] In the case of Bostrom and co-workers,[24] they require an initial conformation generated by Corina. However, in the more recent program PHASE developed by Dixon and co-workers[7] they incorporate a program called Ligprep[25] for converting 2-D to 3-D structures and two different methods for generating conformations. They employ a sampling procedure for the torsional bond angles followed by energy minimization and another method involving a Monte Carlo multiple minimum minimization[26] and Low Mode conformational searching[27] using the software MacroModel.[28]

The method used in this work has similarities with other techniques that randomly perturb a configuration then apply geometry/energy minimization.[29] However, it involves energy minimization of randomly generated initial conformations. The method is intended as an alternative strategy to link 2-D structures with 3-D geometry/energy optimization by generating many approximate initial 3-D structures/conformations that can be improved by any energy/geometry optimization algorithm. We have found that it works well for generating conformations when linked with a good geometry/energy optimization program. Starting from the 2-D structure, there are four steps involved in generating 3-D conformations: (1) estimate bond lengths, (2) randomly generate initial bond angles, (3) stretch the initial structure to remove overlapping structures, and (4) energy minimization/geometry optimization.

Bond lengths can be approximated by the sum of the covalent radii of the pair of atoms involved.[30] The orientation of the bonds can be specified with two angles, which allow the relative positions of the connected atoms to be fixed. A third angle representing rotation around the axis of the bond has no effect on the position of the pair of atoms and is not necessary. So to fix the location of n atoms in a molecule, $2(n - 1)$ bond angles need to be specified. This is under the assumption that the structure in question is not split into two or more unconunconnectednected substructures. If there are multiple substructures, then each can be handled one at a time with this procedure.

The bond angles of the initial structure are randomly generated, which can lead to a number of problems. For example, the initial structure might be infeasible because it has an overlapping structure or because of disconnected bonds (Figure 1). Also, because 3-D methods only account for atom types and atom positions and not usually the bonds, this means

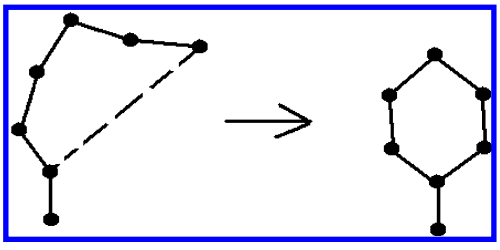


Figure 1. Diagram showing how optimization is used to close loops.

dx.doi.org/10.1021/ci200442h | J. Chem. Inf. Model. 2012, 52, 577−588

passing an initial structure into energy minimization can lead to the 2-D molecular structure being rearranged, which will often fail to reach a stable conformation. To solve this problem, the randomly generated structure is first "stretched out" by minimizing the objective function in eq 1

$$\text{objective} = \sum_{\text{bonded}} (d_{i,j} - d_{i,j}^{\text{ex}})^2 + \sum_{\text{non-bonded}} e^{2\times d_{i,j}^{\text{ex}}/d_{i,j}} \quad (1)$$

This objective function relates the actual distances between atoms i and j, $d_{i,j}$, to the expected distance if they were bonded together, $d_{i,j}^{\text{ex}}$, which is the sum of the covalent radii. In this equation, the distances are measured in angstroms (1 Å = 10^{-10} m). The first part of this objective function ensures that the distance between bonded pairs is fitted to the estimated/expected bond length. The second part maximizes the distance between nonbonded pairs, which is responsible for stretching the molecule to remove overlaps.

The conformations produced by this geometric optimization can then be fed into a more rigorous energy minimization procedure to compute the accurate conformations. In this work, this energy minimization is performed with the quantum chemical package Gaussian 03[31] using the semi empirical method AM1 and the basis set STO-3G. Conformations obtained through this method are local minima of the energy/geometry optimization performed by Gaussian, which are the different possible configurations each molecule can exist in. The AM1 method is used here because it can obtain reasonably accurate results for organic compounds,[32] while still being fast enough to obtain a number of conformations for each molecule involved. More accurate conformations could be obtained using a more accurate method such as B3LYP; however, this would require significantly more computational time. For a single conformation of the molecule sedoheptulose 7, phosphate geometry optimization using AM1/STO-3G required 1 min and 6 s, while using B3LYP/6-31+G(d,p) required 3 days 2 h, starting from the same initial structure and using the same desktop computer with a 3 GHz Intel Core 2 Quad CPU processor.

A recent study[33] shows that N,N-dimethylformamide and phosphoric acid structures generated using AM1 and B3LYP both compare well with experimental data (crystalline and condensed states of phosphoric acid and for the gaseous strructure of N,N-dimethylformamide). Comparing the structures generated by the two methods in this article, we see that the difference in the calculated bond lengths are at most 0.03 Å, and the differences in angles are typically 2−3 degrees. It is also worth noting that in this case one of the bond lengths and many of the angles calculated using AM1 are closer to the reported experimental data than those calculated with B3LYP.[33] Also, although B3LYP is generally regarded as a superior method, there are some exceptions such as the calculation of the structure and vibrational energies of large N_2O clusters for which AM1 gives results closer to the experimental values than any other quantum chemical method.[34] In terms of using different basis sets, deVisser[35] found that when using B3LYP, B3LYP/6-31G compared with B3LYP/6-311+G* for a complete reaction mechanism of substrate hydroxylation by an iron(IV)-oxo oxidant, which gave negligible differences in relative energies for the complete reaction profile and minor changes in geometries.

Methods using molecular mechanics can be used to generate large numbers of conformations in less time than the above methods. However, these less rigorous methods are more commonly used for very large systems that are currently infeasible using DFT (B3LYP) or semi-empirical (AM1) methods. Also, considering that our aim is to demonstrate the utility of bound stretching, it is considered more informative to perform calculations and comparisons with other methods using a small number of conformations.

It is worth noting that the number of conformations required to sample the search space of a molecule are dependent on its size and flexibility. Borodina and co-workers[36] developed a correlation relating the number of conformations to the number of rotatable bonds, the number of nonaromatic atoms, and the required resolution of the coverage. Their fitted correlation shows that smaller molecules with 0.2−2 rotatable bonds might require only 1−6 conformations, but those with 14.2−20 rotatable bonds might require tens or hundreds of thousands of conformations to achieve a higher resolution converage.[33] Making use of distance geometry and using bound stretching, it is possible to consider an extremely large number of conformations, provided the distance bounds contain all possible combinations.

2.2. Distance Geometry Representation. Flexibility in molecules can be accounted for using a distance geometry representation,[37] where distance ranges between every pair of atoms in the structure can be used to represent the allowed range of movement (flexibility). For this purpose, a procedure must be in place for measuring or computing the upper and lower bounds for every distance range. Approximate values for the bounds of these distances can be obtained by starting from a set of standard bond lengths and standard bond angles and using a method such as triangle or tetrangle bound smoothing[38] to tighten the bounds. Starting from initial inaccurate bounds, these methods use distance inequalities based on three points (triangle smoothing) and four points (tetrangle smoothing) in order to reduce upper bounds and increase lower bounds.[38]

However, a more accurate and tighter set of bounds can be found by using a set of energy minimized conformations and extracting the minimum and maximum distances between the corresponding atoms or features. Thus, the upper and lower bounds between atoms i and j are simply found for each of the conformations k

$$\text{Upper Bound}(i, j) = \max[\text{distance}(i, j, k)] \quad (2)$$

$$\text{Lower Bound}(i, j) = \min[\text{distance}(i, j, k)] \quad (3)$$

This will give a set of bounds that contain all the different computed conformations in addition to many other possible conformations that also fit inside these bounds. If enough different conformations are used, the resulting bounds should contain all or most of the possible conformations. This is the main advantage of the distance geometry representation because it removes the necessity to store and work with large numbers of conformations. It however needs to be mentioned that the computed bounds will also usually contain many high energy and unstable conformations. Nevertheless, it is often unclear exactly how many conformations are required to represent the full range of flexibility of a given molecule.

2.3. Bound Stretching. To reduce the number of conformations needed, a new method that we call bound stretching has been introduced in this work. This is based on the process of triangle bound smoothing[38] and stems from a method of computing bounds for the unknown distances in a

dx.doi.org/10.1021/ci200442h | J. Chem. Inf. Model. 2012, 52, 577−588

molecule.[15] The bound stretching here makes use of the triangle inequality equations

$$d_{i,j} \leq d_{i,k} + d_{i,k} \quad (4)$$

$$d_{i,j} \geq d_{i,k} - d_{j,k} \quad (5)$$

that relate the distances between three points labeled i,j and k (d_{ij} = distance between points i and j).

For each of the conformations generated, all the distances between molecules have been computed. In addition, applying eqs 4 and 5 allows bounds to be constructed in a logical way, which effectively stretches the overall bounds for each molecule.

For example, to find the bounds for a certain distance between atoms i and j, the upper and lower bounds are found by searching through every other atom k

$$\text{Upper Bound} = \min(d_{i,k} + d_{j,k}) \quad (i \neq k, j \neq k, i \neq j)$$

$$\text{Lower Bound} = \max(d_{ik} - d_{jk}) \quad (i \neq k, j \neq k, i \neq j)$$

The end effect of this procedure is that the original distances are replaced by a set of bounds. However, if the atom pair is chemically bonded, then bound stretching is not applied because it is not considered necessary to stretch the bonds. In these cases, the bounds obtained from eqs 1 and 2 are considered sufficient.

2.4. Flexible Pharmacophore Elucidation. The pharmacophore elucidation method used here is based on a 2-D method used in our previous work constructing reaction networks[20] with a new 3-D procedure, similar to the method of Raymond and Willett,[16] which uses the distance geometry representation with distance bounds instead of bonds.

The common substructure between a pair of molecules is computed by searching for the maximal common subgraph. This is possible by constructing an association graph, which stores the allowed atom matching.[16] The association graph is stored in a matrix whose rows and lines contain all the possible pairings of the atoms of one molecule with the ones of another molecule. For example, for two triatomic molecules, molecule 1 (M1) with atoms A, B, C and molecule 2 (M2) with atoms B, E, and D, the rows (and columns) of the association graph would be (AD, AE, AF, BD, BE, BF, CD, CE, CF), yielding a square 9 × 9 matrix. If for example, pair AB in M1 "matches" pair DE in M2, then the element corresponding to row AE and to column BD should be equal to 1, else it would be 0 (Figure 2).

Two checks are performed to determine if this matching is possible (and hence if the corresponding element is equal to 1): (1) checking if the atom types are the same (e.g., A = carbon and E = carbon) and (2) checking if the distance bounds A−B and D−E overlap.

The six different possible overlaps of bounds are shown in Figure 3. All need to be accounted for when checking if a set of bounds overlap. This is performed through the following equations

$$U_1 + \text{tolerance} \geq L_2 \quad (6)$$

$$U_2 + \text{tolerance} \geq L_1 \quad (7)$$

where U_1 and L_1 are the upper and lower bounds in molecule 1 and U_2 and L_2 are the upper and lower bounds in molecule 2.

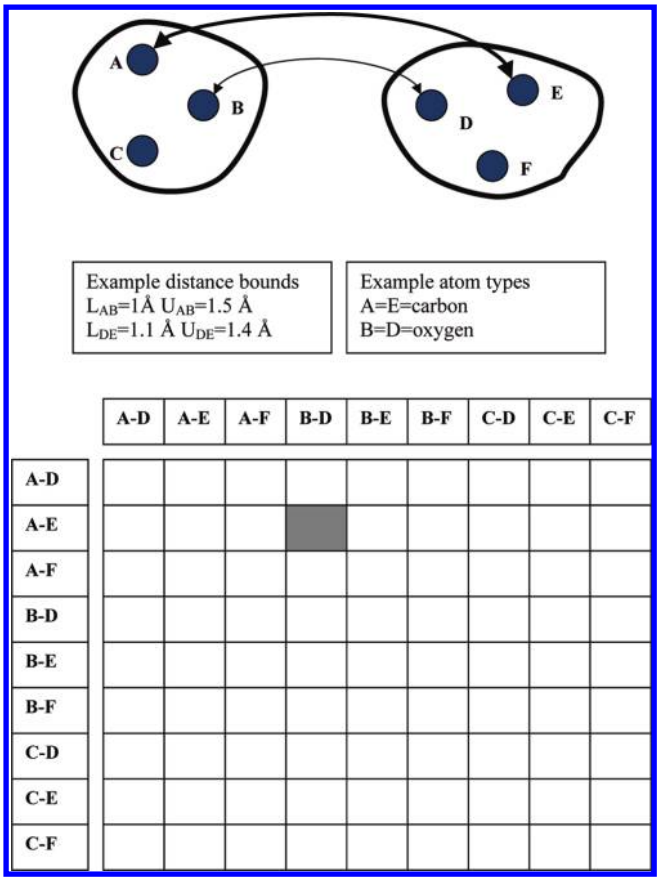


Figure 2. Association graph for two 3 atom molecules. The association graph is a matrix whose elements indicate if a pair of two atoms in one molecule "match" (=1) or not (=0) a pair of atoms in another molecule. The shaded element indicates whether atom "A" in molecule 1 can be mapped onto atom "E" in molecule 2 and at the same time if atom "B" in molecule 1 can be mapped onto atom "D" molecule 2. To determine the feasiblility of this mapping, the types of atoms and the distance bounds between the atoms in the respective molecules are taken into account.

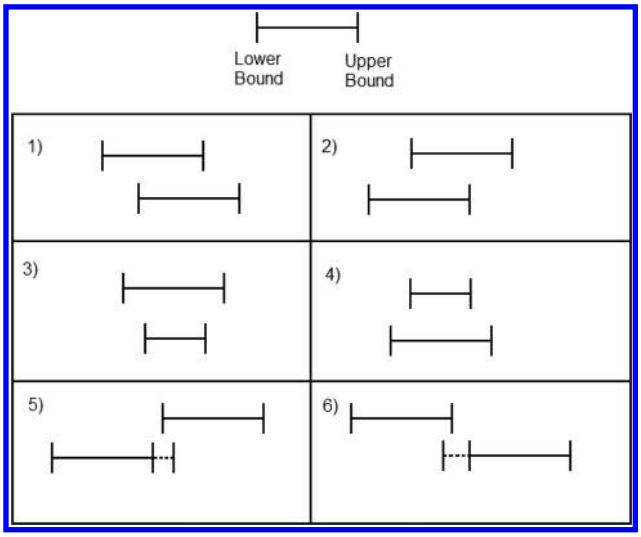


Figure 3. Diagram showing the six different possible overlaps of distance bounds. The horizontal bars representing the bounds and the overlaps in cases (5) and (6) are only possible due to the tolerance added (dashed lines) to the bound checks.

dx.doi.org/10.1021/ci200442h | J. Chem. Inf. Model. 2012, 52, 577−588

Searching for cliques in the association graph is equivalent to searching for common overlapping substructures. So an algorithm for finding the maximum clique[39] is used to find the maximum common substructure. It should be noted that this method generates all possible cliques, which means the process can become computationally expensive.

This method can then be applied to compute pharmacophores by sequential comparison of the binding species. In addition to identifying the nodes involved in the common substructure, this procedure also extracts the overlap of the bounds involved. The pharmacophore computed will hence become a flexible pharmacophore because it is defined by the set of atoms and distance ranges showing the range of possible pharmacophore conformations. This has advantages over the methods that search for rigid active conformations because the flexible pharmacophore representation should contain all the possible active conformations, removing the need to construct multiple rigid pharmacophores.

The extraction of a flexible pharmacophore bound demonstrated in Figure 4 is shown only for a single pair of

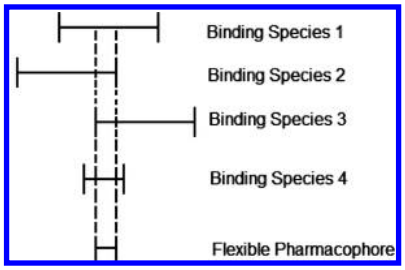


Figure 4. Diagram showing the flexible pharmacophore bound extracted from the bounds of four known binding species.

matching atoms (or features). Flexible pharmacophores identified can depend on the order in which the sequence of binding species are compared and also on the tolerance used.

The tolerance used is of particular importance to pharmacophore elucidation because it is added to each of the comparisons in the sequence. This means if there are seven binding species compared, then the tolerance will be added six times. So a large tolerance could lead to an incorrect pharmacophore with excessively large bounds (in the worst cases). However, the tolerance should still be large enough to account for any variation in bond lengths or any inaccuracy in the structure optimization of the conformations. The tolerance is also useful when the distance bounds become very small so that very close sets of bounds are not missed.

The existence of different orderings of binding species (the sequence in which they are compared) giving different pharmacophores is not necessarily a problem because the pharmacophores are usually quite similar. Nevertheless, the situation can be improved either by modifying the method to compute multiple pharmacophores or by introducing a scoring routine to improve the pharmacophore extraction process (see below).

The computational procedure shown in Figure 5 is a modification of our previous 2-D method[20] and provides improved results in terms of more accurate pharmacophores and better fitting binding species. Here simcomp,[21] which was employed in the 2-D method, has been replaced with separate steps for creating an association graph that accounts for the distance bounds and so that a scoring routine (see below) can be implemented into the clique finding. The step missing from

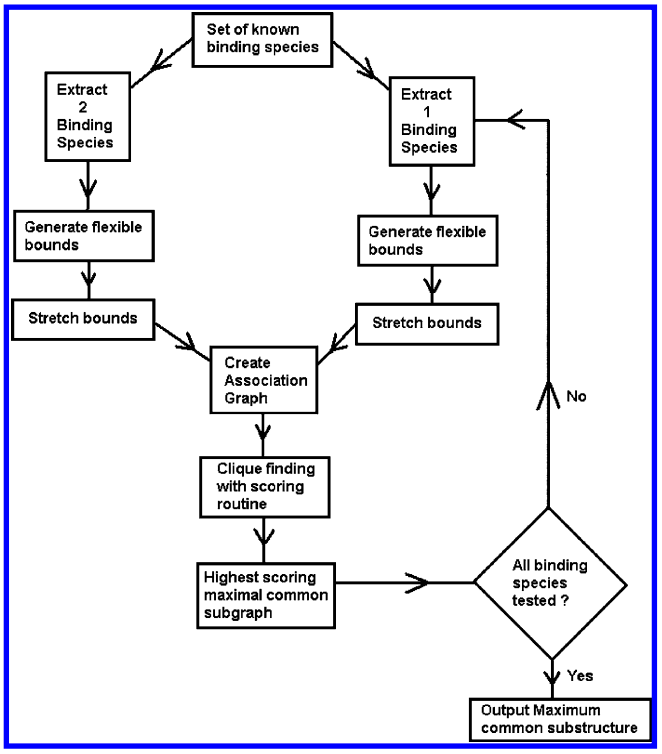


Figure 5. Computational procedure for generating 3-D flexible pharmacophores (with pair wise substructure comparisons).

this procedure is the computation of conformations for each molecule that is carried out before the above procedure. Flexibility is added by extracting distance bounds from the conformations then (optionally) applying bound stretching. As with the 2-D method, the steps are repeated until all the binding species have been included. The final highest scoring maximal common subgraph is the pharmacophore.

2.5. Overlap Scoring. The maximal common subgraph method searches for the common substructure with the most nodes or edges. In most cases there will be more than one structure that includes the maximum number of nodes. To improve the selection, process a scoring routine is applied to rank the different common substructures so the best matching structure can be found. The scoring used here is based on the suggestion of Raymond and Willett[16] that involves weighting the edges according to the size of the overlap and only accepting matching edges with a minimum weight.

A score for each common substructure is computed by adding together the relative overlap of all the matching distances. The score for a common substructure is given by

$$\text{Score} = \sum_{\text{all.atom.pairs}} \frac{\min(U_1, U_2) - \max(L_1, L_2) + \text{tolerance}}{\min(U_1 - L_1, U_2 - L_2)} \quad (8)$$

where L_1 and U_1 are the lower and upper bounds, respectively, for each pair of atoms in molecule 1 and L_2 and U_2 are the (lower and upper) bounds for the corresponding "matched" pairs of atoms in molecule 2. It is also important to note that this equation only applies to cases where the bounds for two matching pairs of atoms (in molecule 1 and 2, respectively) *overlap*, i.e, they satisfy eqs 6 and 7 (which is the case within the common substructure). Equation 8 calculates the length of the overlap in Angstroms (plus a tolerance) divided by the length of

dx.doi.org/10.1021/ci200442h | J. Chem. Inf. Model. 2012, 52, 577−588

the shorter of the two bounds. This yields the *relative overlap* for each pair of matching atoms (which has a value between 0 and 1). Hence, the maximum possible score is equal to the number of atom pairs. The different possible scores for a pair of bounds are shown in figure 6.

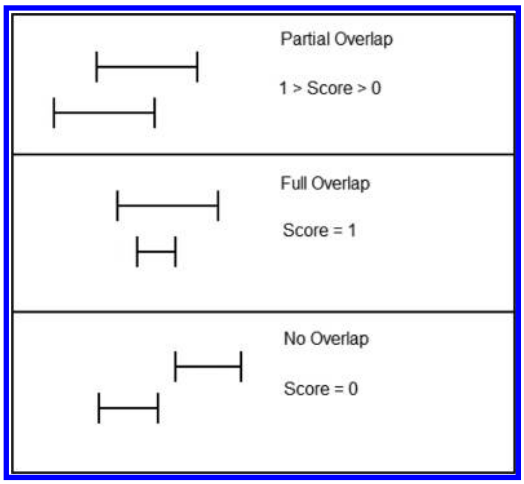


Figure 6. A diagram showing the different possible scores for a pair of bounds.

The maximal common subgraph method with and without any scoring can accept both partial and full overlap (Figure 6) as a match. However, adding the scoring as in eq 8 gives preference to the full overlap cases and to higher scoring partial overlaps.

Scoring is implemented in this work in a modified version of the clique-finding algorithm of and Kerbosch.[39] The original algorithm finds the maximum clique by computing all possible cliques using a recursive algorithm that employs an association graph to give the maximum common substructure. The scoring routine which implements eq 8 is added into this algorithm by computing the score of each common substructure (from the associated clique) and storing the clique with the most nodes and the highest score.

The scoring-based method can be applied to both pharmacophore elucidation and to searching for binding species containing the pharmacophore. The addition of the scoring allows sequential comparisons in pharmacophore elucidation to find better matches and hence give the possibility of finding larger pharmacophores. Applying the scoring to the search for binding species will not increase the size of the common substructure found; however, it will give scores to each of the binding species so that they can be ranked according to how well they fit the flexible pharmacophore.

3. CASE STUDY I: HEXOKINASE BINDING SPECIES

The above methodology has been applied to compute pharmacophores and to search for binding species for a particular binding site of hexokinase (EC 2.7.1.1). This is one of the enzymes considered in our previous work[20] where binding species have been extracted from known reactions found in the KEGG:Ligand database.[40] The binding site in question has nine known binding species (identified by a match-up algorithm).[20] The 2-D methods identified a pharmacophore and found 110 possible binding species in the database that match this pharmacophore. This case study demonstrates how flexible 3-D pharmacophores are used to search through the binding species identified by the 2-D method to find any which can be eliminated. It would also be possible to identify alternative binding species from the KEGG database; however, this would require the computation of conformations for the thousands of compounds in the database. So for simplicity, this option is not considered here.

3.1. Flexible Phamacophore Elucidation. The 2-D pharmacophore (Figure 7) generated using a 2-D method[20]

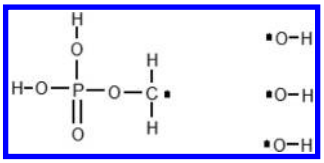


Figure 7. The 2-D pharmacophore for the hexokinase binding site.

contains four separate parts with the large dots representing the free bonds where structure could be attached.

The 3-D flexible pharmacophores are more difficult to visualize because they contain distance ranges instead of bonds. For simplicity, these pharmacophores have been depicted here (Figures 8 and 9) overlaid on the 2-D structure of the last binding species in the comparison sequence.

This case study is also used to test the importance/sensitivity of the bound stretching, scoring, and different tolerances. Six of the pharmacophores generated are shown in Figure 8. All six have been constructed with the scoring routine but with different tolerances and with the even numbered pharmacophores found using bound stretching. The exact specifications used are listed in Table 1, which also shows the sizes of the pharmacophores in terms of the numbers of atoms involved. These results show that increasing the tolerance and adding bound stretching can lead to larger pharmacophores with more matching atoms. This is significant because potential binding species that contain the smaller pharmacophore but not the larger one obtained with bound stretching and higher tolerances would be eliminated, hence, obtaining a tighter specification with respect to the possible binding species. In this way infeasible drug targets could be, for example, eliminated at an early stage. The pharmacophores are also very similar to each other with the same "core" substructure and small variations of the remaining atoms. However, the use of large tolerances could lead to infeasible matches because they allow the molecule to stretch beyond its limits. The maximum tolerance used here is 0.1 Å because it is small compared to the typical length of the bonds. Another disadvantage of using larger tolerances is that it can increase the computational cost of the comparisons because it increases the number of possible common substructures. The addition of bound stretching can also increase the computational cost significantly for the same reasons. This is reflected in the computational times of the six pharmacophores in Figure 8 with those computed without bound smoothing that take seconds to construct and those with the additional bound stretching which take hours. Considering the fact that both methods produce similar pharmacophores, it might seem that the bound smoothing is not useful; however, the small increase in the size of the pharmacophore can have a big effect on the numbers of binding species identified or eliminated.

The above pharmacophores have all been computed using the scoring routine. To show the benefits of the scoring routine here, pharmacophores have also been computed without using

dx.doi.org/10.1021/ci200442h | J. Chem. Inf. Model. 2012, 52, 577−588

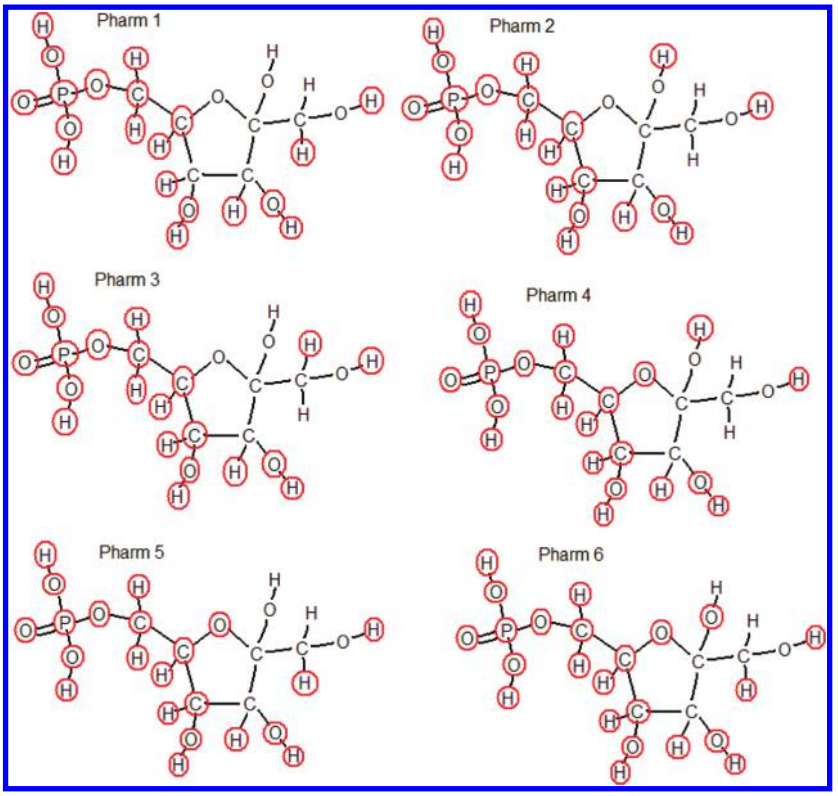


Figure 8. Diagram showing six different flexible pharmacophores with circles highlighting the atoms involved.

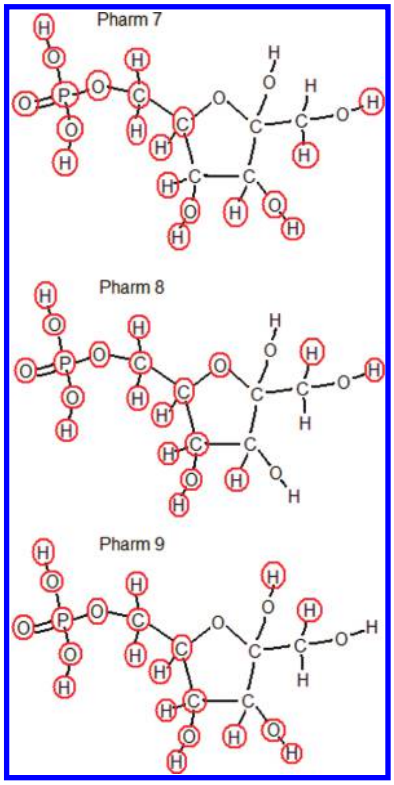


Figure 9. Pharmacophores generated without the scoring routine.

the scoring routine. For simplicity, these are computed without bound stretching using the same tolerances (equivalent to pharm 1, pharm 3, and pharm 5).

The resulting pharmacophores are shown in Figure 9 and are compared against those generated with the scoring routine in Table 2. It is clear that the scoring can find larger pharmacophores (involving more atoms) in cases where the pharmacophore is found through the sequential comparison of known binding species. The only disadvantage of adding the scoring routine is that a small additional amount of computation is required to calculate the scores. However, the benefit of this addition is that an improved pharmacophore can be found without changing any of the bounds.

Table 1. Specifications for Pharmacophore Elucidation

tolerance	without bound stretching	with bound stretching
0.0001 Å	pharm 1 (20 atoms)	pharm 2 (21 atoms)
0.01 Å	pharm 3 (21 atoms)	pharm 4 (22 atoms)
0.1 Å	pharm 5 (22 atoms)	pharm 6 (23 atoms)

Table 2. Summary of Pharmacophores Generated with and without Scoring

tolerance	with scoring routine	without scoring routine
0.0001 Å	pharm 1 (20 atoms)	pharm 7 (20 atoms)
0.01 Å	pharm 3 (21 atoms)	pharm 8 (20 atoms)
0.1 Å	pharm 5 (22 atoms)	pharm 9 (21 atoms)

3.2. Elimination of Infeasible Binding Species. The flexible pharmacophores generated can then be compared to the set of 110 binding species found with the 2-D method. These comparisons can also take advantage of the scoring and bound smoothing techniques applied for pharmacophore elucidation.

For simplicity, the comparisons with binding species have been performed using the same tolerance and method used to construct the pharmacopores. This ensures that the known binding species used to construct the pharmacophore will not be missed. There might be advantages to applying bound stretching

dx.doi.org/10.1021/ci200442h | J. Chem. Inf. Model. 2012, 52, 577−588

Table 3. Size of Comparisons and Number of Binding Species Eliminated by the Six Pharmacophores in Figure 8

pharmacophore	tolerance	size of pharm	range of comparisons	number eliminated (exact)	number eliminated (partial)
pharm 1	0.0001 Å	20	14−20	69	41
pharm 3	0.01 Å	21	16−21	68	39
pharm 5	0.1 Å	22	18−22	67	37
pharm 2	0.0001 Å	21	18−21	47	14
pharm 4	0.01 Å	22	19−22	51	18
pharm 6	0.1 Å	23	20−23	53	19
all				44	7
any				77	51

to a set of potential binding species being compared to a pharmacophore constructed without bound stretching. This can allow, for example, for the identification of binding species that need to assume a high energy conformation to fit the pharmacophore. Also, using a small tolerance to construct a pharmacophore and a larger tolerance to search for binding species will have the same effect. However, these possibilities have not been investigated here because our main purpose is to present the technology developed. A relevant parametric analysis is the scope of a future publication.

The six pharmacophores shown in Figure 8 have been compared to the 110 possible binding species identified by the 2-D method. For each of these potential binding species distance bounds are constructed by computing 20 different conformations for each and applying the above methods. The pharmacophores constructed with bound stretching (pharm 2, 4, and 6) are compared to binding species with bounds constructed by the same method. The results of these comparisons are summarized in Table 3, which shows the range of different comparison sizes (number of atoms matching the pharmacophore atoms) and the numbers of species that can be eliminated by each pharmacophore.

In the 2-D methodology, a binding species would be eliminated if it did not exactly contain the entire pharmacophore. The numbers of species that do not contain the entire flexible 3-D pharmacophore are shown above in the column labeled "exact". To account for the possibility that the pharmacophore might be augmented with additional unnecessary structure, a partial match allowing 1 atom missing from the pharmacophore is also considered (see the column labeled "partial"). These partial matches also account for the possibility that a match is missed because insufficient conformations have been generated. However, this possibility should be accounted for by the addition of higher tolerances and with the bound stretching.

Considering the comparisons in Table 3, the number of binding species that can be eliminated is between 7 and 77 out of the total 110 potential binding species. Eliminating 77 binding species is possible if we require that all the binding species match all the pharmacophores exactly. However, this includes comparisons with pharm 1 that use a very small tolerance and without bound stretching meaning that some of the 77 species might have been eliminated because the comparisons have not fully accounted for the flexibility. Nevertheless, it is a relatively safe assumption that the 33 species that fit all pharmacophores are "good" potential binding species (most likely to bind).

The worst of the potential binding species are those seven that do not fit any of the six pharmacophores. The elimination of only 7 out of 110 binding species might seem insignificant; however, even a small reduction in the number of binding species can dramatically reduce the number of reactions and pathways that can be formed using those binding species.

4. CASE STUDY II: THROMBIN FLEXIBLE PHARMACOPHORES

In this second case study, flexible pharmacophores are computed from a set of thrombin ligands and compared to a known pharmacophore and to those computed by commercially available packages. The pharmacophores for this set of ligands are generated and compared by Patel and co-workers[41] using the pharmacophore elucidation packages DISCO,[4] Catalyst,[6,42] and GASP.[5] In addition, the pharmacophores for this data set are also computed using more recent methods including the package GALAHAD[8] and a multi-objective genetic algorithm (MOGA)[13] method that is related to GASP.[5] Here, the new flexible methodology (see above) is compared to these five methods/packages in order to show the differences and possible advantages of this approach. Pharmacophores are generated using this methodology exploit both the scoring routine and bound stretching. These are compared against the visually determined (known) pharmacophore in the literature[41] and the pharmacophores generated by the five methods mentioned above.

4.1. Comparison of Methodology. The flexible methodology developed in this work is different from the methods of DISCO,[4] GASP,[5] Catalyst,[6,42] GALAHAD,[8] and MOGA[13] in that these methods use flexible methodology to find rigid pharmacophores, and the method developed here uses flexible methods to compute a flexible pharmacophore. A rigid pharmacophore is defined by exact distance between pharmacophore points. However, a flexible pharmacophore is defined here using distance ranges that should contain a number of possible rigid pharmacophores. The idea of constructing flexible pharmacophores is equivalent to earlier methods for generating flexible binding site models[18] and is more commonly applied in database searching with a known query in more recent works.[16] However, the method used for creating flexible pharmacophores is improved here by the addition of the scoring routine and with bound stretching to find more accurate pharmacophores using fewer conformations.

The methods in DISCO, Catalyst, and GALAHAD also rely on a set of precomputed conformations for each of the molecules being compared (although in the case of GALAHAD these conformations are generated as the first step of the program).[8] Hence, they share the same problem where a solution could be missed if insufficient conformations have been generated. Conformations are generated during the search for a pharmacophore in the GASP program and in the MOGA method, which both use a genetic algorithm approach to find the optimum bond angles and the optimum matching features, implemented with limitations to prevent excessively long

searches for the global optimum. The final solutions given will usually be different with each run.

Another important difference is that all five methods DISCO, GASP, Catalyst, GALAHAD, and MOGA identify significant features in the binding species before computing pharmacophores. These significant features could include hydrogen bond donors and acceptors, hydrophobic groups, and positive or negative charge centers.[43] The effect of identifying these important features is that the number of points defining the chemical structure is reduced. For example, a ring structure could be replaced by a single point. This reduction can dramatically reduce the computational cost of comparisons making comparisons of large molecules feasible. However, this simplification could cause problems if the important features are missed or if significant features are lumped together with unimportant ones. If the important features are identified correctly, however, then the correct pharmacophores can be found more easily.

For simplicity, and in order to demonstrate the effectiveness of our methodology (together with bound stretching and scoring), this feature identification has not been applied. Instead, all atoms are considered as features except for the hydrogen atoms in order to reduce computational costs.

4.2. Flexible Pharmacophore Elucidation. The set of thrombin ligands used to extract a pharmacophore is shown in Figure 10, which is the example system used by Patel and co-workers.[41]

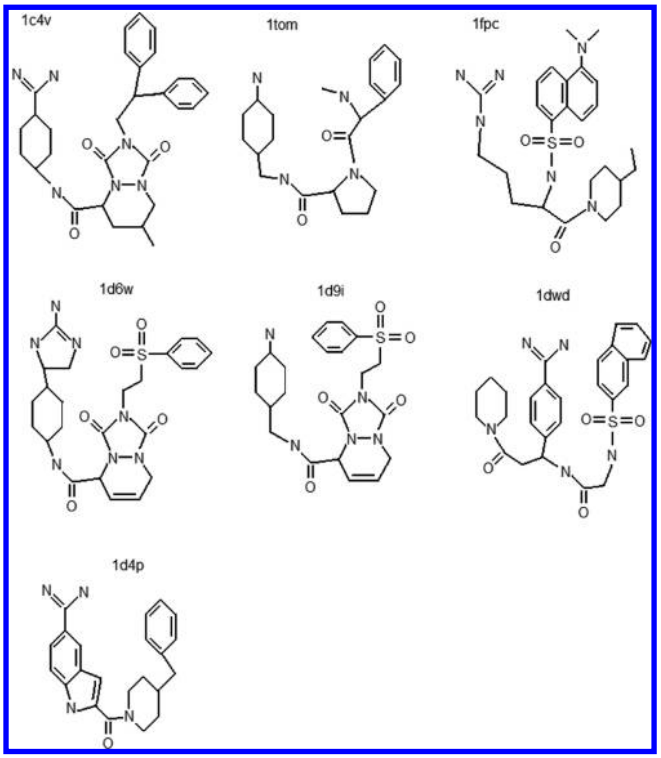


Figure 10. Seven thrombin ligands together with the protein data bank codes of the protein−ligand complexes.

The codes next to each ligand structure are the Protein Data Bank[44] code for the protein−ligand complexes.

The flexible bounds for each ligand are extracted from a set of 20 conformations generated for each to be used with the new methodology. Flexible pharmacophores are computed using different tolerances, with and without the bound stretching. The ligands are compared sequentially starting with the smallest ligands with the fewest atoms. The pharmacophore elucidation here uses fitting with only the non-hydrogen atoms in order to simplify the process.

The results of pharmacophore generation are shown in Figure 11 where the computed pharmacophores are depicted together with the target pharmacophore features identified by Patel and co-workers.[41] It is clear that none of the pharmacophores contain all four pharmacophore features. However, all the pharmacophores can identify at least two of the features, and all of the features have been identified by at least one pharmacophore. The tolerances and method used are specified in Table 4 from which it is found that larger tolerances and the addition of bound stretching lead to larger pharmacophores. The two pharmacophores generated with bound stretching and with the largest tolerances (pharm 6 and pharm 8) are also the only ones that include three out of the four target pharmacophore features. The comparison of DISCO, Catalyst, and GASP[41] shows that DISCO and GASP have similar difficulties finding all four pharmacophore features and require the combination of a number of common substructures. However, Catalyst is shown by Patel to give the best pharmacophore that matches all four features.[41] In addition, the pharmacophores generated by GALAHAD[8] identify only two features that are common to all seven ligands because they have only used a single conformation for each ligand (the crystal structure). The method of GALAHAD does however identify all the important features using partial matches involving fewer ligands, and if alternative conformations were used, it may be able to find all four features. A set of crystal structures are also used by the MOGA method where the authors are able to identify many features.[13] However, in order to compute a pharmacophore, the authors have excluded the worst fitting ligand 1fpc, and they have only found all the features using partial matches that do not include all the six ligands compared. Also, in many cases the crystal structure is unavailable, which is why the MOGA method is also run using a flexible option that gives a fairer comparison with the other methods (including the one presented here). Using this flexible routine (similar to GASP), the method can find three out of the four features, and considering that the worst fitting ligand has been excluded, this shows that even the more recent sophisticated methods find pharmacophore elucidation to be challenging.

There are a number of reasons why the methodology presented above does not find a pharmacophore containing all four features. One reason is because the flexible pharmacophores are constructed using all the non-hydrogen atoms without the identification of significant features beforehand. This is a problem because the method then finds the pharmacophore with the most atoms (and the highest score), meaning that pharmacophores containing many atoms within a small number of features might be favored over a pharmacophore that has one or two atoms found in each of the important features. For example, some of the pharmacophores shown in Figure 11 include six pharmacophore atoms contained in a single pharmacophore feature.

Another reason could be that too few conformations have been used to generate bounds for each of the ligands meaning that some of the flexibility might be missed (the same reason GALAHAD is only able to find two out of four features for all ligands). For this example, only 20 conformations have been used compared to between 100 and 252 conformations used by Catalyst for each ligand in this example.[41]

Considering that pharmacophores matching three out of four target pharmacophore features can be found with these limitations, it seems likely that adding a routine to identify

dx.doi.org/10.1021/ci200442h | J. Chem. Inf. Model. 2012, 52, 577−588

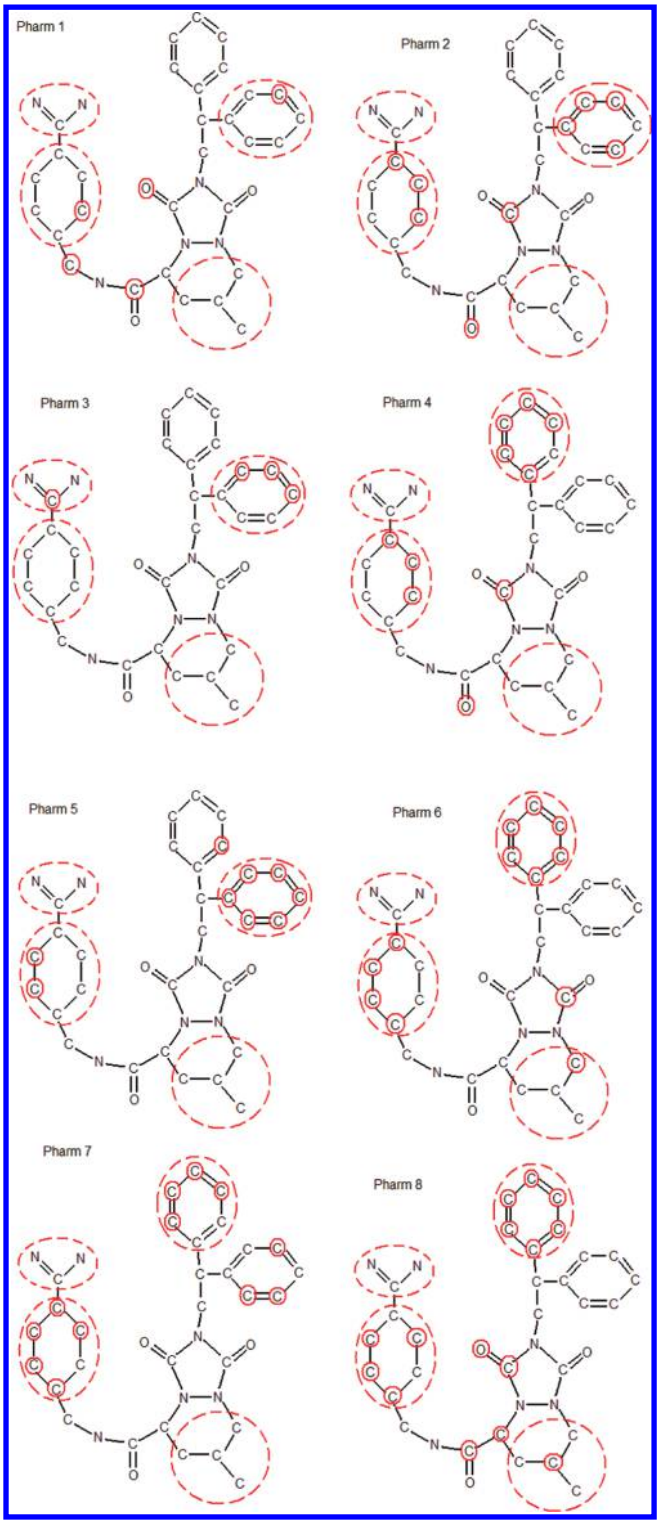


Figure 11. Pharmacophores depicted overlaid on the structure of the ligand 1c4v. The larger dashed circles represent the target pharmacophore identified by Patel and co-workers,[41] and the smaller circles identify the atoms involved in the computed flexible pharmacophores.

important features or computing additional conformations should make it possible to find all four features. The main advantage of the new methodology is that the pharmacophores are created with flexibility built in so that they can contain many possible pharmacophore conformations.

Table 4. Specifications for Pharmacophore Elucidation Using Thrombin Ligands

tolerance	without bound stretching	with bound stretching
0.0001 Å	pharm 1 (5 atoms)	pharm 2 (9 atoms)
0.01 Å	pharm 3 (4 atoms)	pharm 4 (10 atoms)
0.1 Å	pharm 5 (9 atoms)	pharm 6 (12 atoms)
0.5 Å	pharm 7 (12 atoms)	pharm 8 (15 atoms)

The computation times for the pharmacophores in this case study are listed in Table 5 where it is observed that higher

Table 5. Computation Times for Pharmacophore Elucidation Using Thrombin Ligands

tolerance	without bound stretching	with bound stretching
0.0001 Å	pharm 1 (2.9 s)	pharm 2 (8.2 s)
0.01 Å	pharm 3 (2.7 s)	pharm 4 (8.9 s)
0.1 Å	pharm 5 (3.6 s)	pharm 6 (15.3 s)
0.5 Å	pharm 7 (10.3 s)	pharm 8 (7 min 8.8 s)

tolerances will usually lead to longer computation times. The addition of bound smoothing also increases the computation times, although not as dramatically as in the previous case study where they took hours to compute. The increases in computational times in this method are due to the increased numbers of possible matching combinations that are allowed when the bounds are extended.

5. CONCLUSIONS

The methodology presented here combines a number of different methods in order to improve pharmacophore elucidation and database searching using three-dimensional flexible techniques. A flexible distance geometry representation is used where distance bounds between each pair of atoms are generated from a set of low energy conformations. This allows tighter and more accurate bounds than distance geometry methods that rely on an expert system (using standard bond lengths and bond angles) together with bound smoothing (without the generation of conformations). This should also give tighter bounds than the methods on the basis of a systematic search of bond angles, which generate higher energy conformations instead of the energy minimized conformations used in our technique. The addition of bound stretching makes it possible to extend the bounds of nonbonded pairs in a logical way making use of the triangle distance inequality. Its advantage is that it will reduce the number of conformations required to account for the flexibility of the molecules. Applied to pharmacophore elucidation, this bound stretching will lead to larger pharmacophores if insufficient conformations have been generated to fully account for the flexibility.

Pharmacophore elucidation and database searching are possible with this representation using a maximal common subgraph algorithm, which finds the maximum clique in an association graph. Here, the association graph contains the rules determining which pairs of atoms (features) can be matched. These rules include checks to ensure the distance bounds overlap and the atom or feature types match. The maximal common subgraph method includes a new scoring routine based on the relative size of overlap of each of the matching pairs facilitating the computation of the maximal common subgraph with the best matching bounds. The advantage of this approach is that it can select the best match when there are many

dx.doi.org/10.1021/ci200442h | J. Chem. Inf. Model. 2012, 52, 577−588

different common substructures with the maximum number of nodes or edges.

The flexible methodology is applied to two example systems to show the possible advantages of constructing flexible pharmacophores. In the first example, flexible pharmacophores are constructed for a binding site of hexokinase. There are 110 possible binding species identified by our 2-D method reported in a recent publication,[20] and these are compared to the flexible pharmacophores to check if any can be eliminated. Flexible pharmacophores are computed using different tolerances and also testing the effect of bound stretching and scoring. The addition of scoring, bounds stretching and higher tolerances all lead to larger (and more specific) pharmacophores. However, if the tolerances are too high, this could lead to infeasible matching. Here, partial matching allowing binding species containing a substructure that is one atom different from the pharmacophore is also considered.

The second example involves the construction of flexible pharmacophores from a set of thrombin ligands as well as comparison of our method with a number of existing pharmacophore elucidation programs, DISCO, Catalyst, GASP, GALAHAD, and a multi-objective genetic algorithm (MOGA) method.

All the flexible pharmacophores generated contained atoms found in two or three out of the four target pharmacophore features. Although none of the pharmacophores contained all four features, all of the features are found in at least one of the six pharmacophores computed. Also, the only pharmacophores that identified three out of the four features are generated using bound stretching with the higher tolerances (0.1 Å and 0.5 Å). It is shown by Patel and co-workers[41] that GASP and DISCO encounter similar problems requiring the combination of different pharmacophores to find all the target features with Catalyst giving the best pharmacophore. The MOGA method is also shown to have similar difficulties, where a match with three out of four features is possible if the worst one of the problematic ligands is excluded.[13]

There are significant differences between the flexible method used here and these five packages. All five involve an initial step where the significant features of each molecule are identified, which effectively reduces the number of points used in the matching. If the identification is done correctly, the desired pharmacophores can be found more easily. However, this option is not included in the above method in order to avoid complications where important features might be overlooked and to simplify the process so that the benefits of bound stretching and scoring can be seen more easily. A disadvantage of using the atoms for matching is that pharmacophores containing atoms found in all the important features may be overlooked in favor of pharmacophores containing more atoms grouped together in a small number of important features. The other big disadvantage of using the atoms as matching points is that this will usually give more fitting points (atom−atom connections) making the comparisons (involved in pharmacophore elucidation and searching for binding species) more computationally expensive.

Another important consideration is the number of conformations used to represent the flexibility of the ligands. The pharmcophores generated by Catalyst used at least 100 conformations to represent the flexibility of each ligand.[41] For simplicity, the flexible pharmacophores generated here used only 20 conformations for each ligand. It is possible that the addition of more conformations could lead to flexible pharmacophores that match all four target pharmacophore features.

Considering that the flexible methodology could find pharmacophores containing three out of the four pharmacophore features, it seems likely that the addition of more conformations or the addition of a routine to identify significant features would make this method competitive with Catalyst. Considering the difficulties of DISCO, GASP, and the MOGA method, it appears that the current method is already competitive with these packages. It is more difficult to compare our new methodology with GALAHAD because it generated a pharmacophore containing only two out four target features using a single conformation (the crystal structure) for each. We expect that GALAHAD would also achieve much improved results if a sufficiently large number of conformations were used.

In addition to being able to find target pharmacophore features, this flexible methodology has the added advantage that the flexible pharmacophores should directly contain (within the calculated distance ranges) many possible active conformations, overcoming the computational effort of trying to obtain them through multiple computations of rigid pharmacophores.

AUTHOR INFORMATION

Corresponding Author

*E-mail: k.theodoropoulos@manchester.ac.uk. Tel: +441613064386.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Dammkoehler, R. A.; Karasek, S. F.; Shands, E. F. B.; Marshall, G. R. Constrained search of conformational hyperspace. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 3−21.

(2) Cottrell, S. J.; Valerie, J. G.; Taylor, R. Incorporating partial matches within multiobjective pharmacophore identification. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 735−749.

(3) Sheridan, R.; Nilakantan, R.; Dixon, J.; Venkataraghavan, R. The ensemble approach to distance geometry: Application to the nicotinic pharmacophore. *J. Med. Chem.* **1986**, *29*, 899−906.

(4) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; Delazzer, J.; Lico, I.; Pavlik, P. A. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83−102.

(5) Jones, G.; Willett, P.; Glen, R. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532−549.

(6) Clement, O. O.; Mehl, A. T. Hiphop: Pharmacophores Based on Multiple Common-Feature Alignments. In *Pharmacophore Perception, Development, and Use in Drug Design*, 1st ed.; Guner, O. F., Ed.; International University Line: La Jolla, CA, 2000; pp 69−84.

(7) Dixon, S.; Smondyrev, A.; Knoll, E.; Rao, S.; Shaw, D.; Friesner, R. PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 647−671.

(8) Richmond, N.; Abrams, C.; Wolohan, P.; Abrahamian, E.; Willett, W.; Clark, R. GALAHAD: 1. Pharmacophore identification by hypermolecular alignment of ligands in 3D. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 567−587.

(9) Zhu, F.; Agrafiotis, D. Recursive distance partitioning algorithm for common pharmacophore identification. *J. Chem. Inf. Model.* **2007**, *47*, 1619−1625.

(10) Taminau, J.; Thijs, G.; De Winter, H. Pharao: Pharmacophore alignment and optimization. *J. Mol. Graphics Modell.* **2008**, *27*, 161−169.

(11) Schneidman-Duhovny, D.; Dror, O.; Inbar, Y.; Nussinov, R.; Wolfson, H. Deterministic pharmacophore detection via multiple flexible alignment of drug-like molecules. *J. Comput. Biol.* **2008**, *15*, 737−754.

(12) Podolyan, Y.; Karypis, G. Common pharmacophore identification using frequent clique detection algorithm. *J. Chem. Inf. Model.* **2009**, *49*, 13−21.
(13) Gardiner, E. J.; Cosgrove, D. A.; Taylor, R.; Gillet, V. J. Multi-objective optimization of pharmacophore hypotheses: Bias towards low-energy conformations. *J. Chem. Inf. Model.* **2009**, *49*, 2761−2773.
(14) Jones, G. GAPE: An improved genetic algorithm for pharmacophore elucidation. *J. Chem. Inf. Model.* **2010**, *50*, 2001−2018.
(15) Clark, D. E.; Willett, P.; Kenny, P. W. Pharmacophoric pattern matching in files of three-dimensional chemical structures: Use of smoothed bounded distances for incompletely specified query patterns. *J. Mol. Graphics* **1991**, *9*, 157−160.
(16) Raymond, J.; Willett, P. Similarity searching in databases of flexible 3D structures using smoothed bounded distance matrices. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 908−916.
(17) Rhodes, N.; Clark, D.; Willett, P. Similarity searching in databases of flexible 3D structures using autocorrelation vectors derived from smoothed bounded distance matrices. *J. Chem. Inf. Model.* **2006**, *46*, 615−619.
(18) Crippen, G. M. Quantitative structure−activity relationships by distance geometry: Systematic analysis of dihydrofolate reductase inhibitors. *J. Med. Chem.* **1980**, *23*, 599−606.
(19) Guner, O. F.; Henry, D. R.; Pearlman, R. S. Use of flexible queries for searching conformationally flexible molecules in databases of three-dimensional structures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 101−109.
(20) Binns, M.; Theodoropoulos, C. An integrated knowledge-based approach for modelling biochemical reaction networks. *Comput. Chem. Eng.* **2011**, *35*, 3025−3043.
(21) Hattori, M.; Okuno, Y.; Goto, S.; Kanehisa, M. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.* **2003**, *125*, 11853−11865.
(22) Pearlman, R. S. Rapid generation of high quality approximate 3D molecular structures. *Chem. Des. Autom. News* **1987**, *2*, 5−7.
(23) Sadowski, J.; Rudolph, C.; Gasteiger, J. The generation of 3D models of host-guest complexes. *Anal. Chim. Acta* **1992**, *265*, 233−241.
(24) Bostrom, J.; Greenwood, J.; Gottfries, J. Assessing the performance of omega with respect to retrieving bioactive conformations. *J. Mol. Graphics Modell.* **2003**, *21*, 449−462.
(25) *Ligprep*, version 2.0; Schrodinger LLC: New York, 2006.
(26) Chang, G.; Guida, W.; Still, W. An internal coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **1989**, *111*, 4379−4386.
(27) Kolossvary, I.; Guida, W. Low mode search. An efficient, automated computational method for conformational analysis: Application to cyclic and acyclic alkanes and cyclic peptides. *J. Am. Chem. Soc.* **1996**, *118*, 5011−5019.
(28) *MacroModel*, version 9.1; Schrodinger LLC: New York, 2006.
(29) Saunders, M. Stochastic search for the conformations of bicyclic hydrocarbonds. *J. Comput. Chem.* **1989**, *10*, 203−208.
(30) Gillespie, R. J.; Popelier, P. L. A. *Chemical Bonding and Molecular Geometry*; Oxford University Press: New York, 2001; pp 27−33.
(31) Frisch, M. J.; Trucks, G. W.; Schegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.Jr.; Vreven, K. N.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, M.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision B.04; Gaussian, Inc.: Pittsburgh, PA, 2003.
(32) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(33) Fedorova, I. V.; Kiselev, M. G.; Safonova, L. P. Simulating the proton transfer reaction in the phosphoric acid-N,N-dimethylformamide system by means of the AM1 semiempirical method. *Russ. J. Phys. Chem. A* **2011**, *85*, 1917−1922.
(34) Haber, T.; Kevorkiants, R.; Thiel, W.; Suhm, M. A. The performance of the semi-empirical AM1 method on small and nanometre-sized N_2O clusters. *Phys. Chem. Chem. Phys.* **2004**, *6*, 4939−4949.
(35) de Visser, S. P. Trends in substrate hydroxylation reactions by heme and nonheme iron(IV)-oxo oxidants give correlations between intrinsic properties of the oxidant with barrier height. *J. Am. Chem. Soc.* **2010**, *132*, 1087−1097.
(36) Borodina, Y. V.; Bolton, E.; Fontaine, F.; Bryant, S. H. Assessment of conformational ensemble sizes necessary for specific resolutions of coverage of conformational space. *J. Chem. Inf. Model.* **2007**, *47*, 1428−1437.
(37) Crippen, G. M. Distance geometry approach to rationalizing binding data. *J. Med. Chem.* **1979**, *22*, 988−997.
(38) Havel, T. F.; Kuntz, I. D.; Crippen, G. M. The theory and practice of distance geometry. *Bull. Math. Biol.* **1983**, *45*, 665−720.
(39) Bron, C.; Kerbosch, J. Finding all cliques of an undirected graph [H]. *Commun. ACM* **1973**, *16*, 575−577.
(40) Goto, S.; Okuno, Y.; Hattori, M.; Nishioka, T.; Kanehisa, M. LIGAND: Database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.* **2002**, *30*, 402−404.
(41) Patel, Y.; Gillet, V.; Bravi, G.; Leach, A. R. A comparison of the pharmacophore identification programs: Catalyst, DISCO, and GASP. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 653−681.
(42) Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. Identification of common functional configurations among molecules. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563−571.
(43) Greene, J.; Kahn, S.; Savoj, H.; Sprague, P.; Teig, S. Chemical function queries for 3D database search. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1297−1308.
(44) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.