Technique for Generating Three-Dimensional Alignments of Multiple Ligands from One-Dimensional Alignments

Andrei V. Anghelescu,[†] Robert K. DeLisle,[‡] Jeffrey F. Lowrie,[§] Anthony E. Klon,[†] Xiaoming Xie,[†] and David J. Diller*[,†]

Department of Molecular Modeling, Pharmacopeia, Inc., CN5350, Princeton, New Jersey 08543, Array BioPharma, Inc., 3200 Walnut Street, Boulder, Colorado 80501, and Schrödinger, Inc., 120 W45th Street, New York, New York 10036

Received October 31, 2007

We describe and demonstrate a method for the simultaneous, fully flexible alignment of multiple molecules with a common biological activity. The key aspect of the algorithm is that the alignment problem is first solved in a lower dimensional space, in this case using the one-dimensional representations of the molecules. The three-dimensional alignment is then guided by constraints derived from the one-dimensional alignment. We demonstrate using 10 hERG channel blockers, with a total of 72 rotatable bonds, that the one-dimensional alignment is able to effectively isolate key conserved pharmacophoric features and that these conserved features can effectively guide the three-dimensional alignment. Further using 10 estrogen receptor agonists and 5 estrogen receptor antagonists with publicly available cocrystal structures we show that the method is able to produce superpositions comparable to those derived from crystal structures. Finally, we demonstrate, using examples from peptidic CXCR3 agonists, that the method is able to generate reasonable binding hypotheses.

INTRODUCTION

Despite the rapid increase in the number of protein–ligand cocrystal structures, large biologically important protein families, such as G-protein coupled receptors and ion channels, continue to be extremely difficult to crystallize. While homology modeling can provide insight,[1–3] rational drug design is, for these protein families, still largely dependent on ligand-based methods, such as quantitative structure–activity relationships, pharmacophore modeling, and lead hopping. Three-dimensional ligand alignment in one form or another is the basis for much of ligand-based drug design.

Because of the importance of ligand alignment there have been a large number of methods proposed for generating three-dimensional ligand alignments.[4–26] Most of these structural alignment methods can be broken down into two components: a scoring function to quantify the quality of an alignment and an algorithm to search for the optimal alignment. The scoring functions typically consist of a force field based internal energy to ensure reasonable geometry for each molecule and an alignment scoring function. The alignment scoring functions are usually empirical functions of factors that are critical components of protein–ligand binding. These alignment scoring functions include shape,[4,25,27] molecular surface similarity,[6,8] steric/electrostatic field similarity,[10,21] similarity based on pairwise atomic similarity and overlap,[15] and combinations thereof.[22]

The diversity in alignment scoring functions is more than matched by the number of approaches developed to find alignments that optimize these scoring functions. The reason that many alignment algorithms have been developed is that solving the full three-dimensional alignment problem for several flexible molecules involves a global optimization problem whose solution is likely beyond standard global optimization algorithms. The full three-dimensional problem has $6(N-1) + \sum_{i=1}^{N} R_i$ degrees of freedom where N is the number of molecules and R_i is the number of rotatable bonds of the ith molecule. In addition, this search problem can have discrete degrees of freedom in the form of unspecified stereocenters, ring conformations, and cis/trans orientations around double bonds. Since the complexity of the search grows exponentially in the number of degrees of freedom, the problem becomes intractable with even modestly sized data sets unless some simplifying assumptions are made. The approach to multiple alignment most methods take is to align a single flexible ligand to a rigid template molecule. A multiple alignment is then built ligand by ligand thereby limiting the number of degrees of freedom at any one step.

The alignment technique described here differs from most previous methods in that it attempts to solve the fully flexible multiple alignment problem of all ligands of interest simultaneously. The assumption that the optimal multiple alignment necessarily occurs from optimal pairwise alignments is likely not to be true particularly with molecules lacking sufficient similarity to discern a single clearly optimal alignment. The chief advantage of aligning multiple molecules simultaneously is that partial structural overlap over many molecules may be easier to identify and therefore exploit.

* Corresponding author phone: (609)452-3783; fax: (609)655-4187; e-mail: ddiller@pcop.com.
[†] Pharmacopeia, Inc.
[‡] Array BioPharma, Inc.
[§] Schrödinger, Inc.

10.1021/ci700395f CCC: $40.75
Published on Web 04/16/2008

The chief simplification of the three-dimensional alignment method described here is that it begins by solving the alignment problem in a lower dimensional search space. In particular, the one-dimensional representation[28] is used for each molecule to derive a multiple one-dimensional alignment.[29] Constraints derived from the multiple one-dimensional alignment are then used to drive the three-dimensional search. In the lower dimensional space the search problem becomes tractable. For N molecules the solution of the one-dimensional multiple alignment problem involves solving a global optimization problem in $N-1$ continuous variables (the translation of each molecule relative to the alignment) and $N-1$ discrete variables (the orientation of each molecule relative to the alignment). For example, if the mean number of rotatable bonds within a data set is 4 then working in one dimension decreases the number of degrees of freedom by a factor of 10. Once good solutions to the multiple alignment problem are found in one dimension the conserved features of the multiple one-dimensional alignment are used as constraints during the three-dimensional alignment. By doing so, this approach is able to perform a fully flexible alignment of several molecules simultaneously rather than via an incremental pairwise alignment.

A fair analogy can be made between this three-dimensional ligand alignment method and a standard approach to multiple protein structure alignment in the following manner. In order to align several structures of homologous proteins a common approach is to first perform a multiple sequence alignment from which conserved amino acids are identified. Then the protein structures are aligned by selecting the rigid body transformations that minimize the spatial rms deviations between the corresponding C_α atoms of the conserved amino acids. For this ligand alignment method, the amino acids are replaced by pharmacophoric features, the protein sequences are replaced by the one-dimensional representations, and the multiple sequence alignment is replaced by the multiple one-dimensional ligand alignment. The chief failing of this analogy is that the search for the optimal three-dimensional ligand alignment is an all atom search through conformational space, whereas the search for the optimal protein structural alignment is just one of finding optimal rigid body transformations.

We demonstrate our approach to the three-dimensional alignment problem on several test cases representing a range of relevant problems. First, to demonstrate the advantage of the approach, we apply it to the alignment of 10 hERG channel blockers with a combined 72 rotatable bonds. Next we show that the method is able to reproduce superpositions derived from available cocrystal structures of 5 estrogen receptor antagonists and 10 estrogen receptor agonists. Finally, we apply the alignment technique to the identification of short sequences of amino acids from IP-10, a natural occurring agonist of CXCR3, that are similar to known small molecule CXCR3 agonists.

METHODS

The general approach (see Figure 1) is first to solve, to the extent possible, the multiple alignment problem in one dimension[29] rather than in three dimensions and then to use the one-dimensional alignment to guide the construction of the three-dimensional alignment. Conserved pharmacophore features in the multiple one-dimensional alignment, the rectangular boxes in the upper right corner of Figure 1, are identified using a nonparametric clustering algorithm called combinatorial principle component analysis (cPCA).[30] The conserved pharmacophore features in the one-dimensional multiple alignment are converted into constraints for the three-dimensional alignment. A three-dimensional alignment is then sought that simultaneously aligns the atoms in each conserved feature from the one-dimensional alignment and satisfies the typical geometric constraints of low-energy conformations, such as the preservation of ideal bond lengths and ideal bond angles and the avoidance of steric clashes. A solution that satisfies all these constraints is sought initially via stochastic proximity embedding, a self-organizing procedure developed by Agrafiotis and co-workers[31–33] for, among other applications, the related problem of conformational searching. The alignment is then subjected to gradient-based all atom minimization using an objective function consisting of the sum of a force field-derived internal energy of each molecule and the alignment score developed by Labute and co-workers.[15] Each of these steps is described in more detail below.

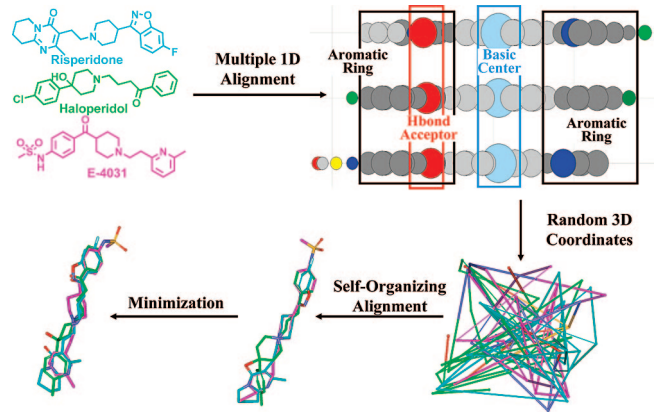

Figure 1. An overview of the three-dimensional alignment technique. The molecules in the data set are first aligned using their one-dimensional representation[28]. The one-dimensional alignment of risperidone, haloperidol and E-4031 is shown in the upper right corner. Each horizontal row of colored circles is the one-dimensional representation of one of the three molecules on the left. The circles are colored according to atom type: gray – carbon, red-oxygen etc. The circles are sized according to their contribution to the alignment score. Conserved features are then identified in the multiple one-dimensional alignment using the cPCA algorithm.[30] For this example the boxes superimposed on the multiple alignment indicate 4 conserved features. The two black boxes highlight conserved aromatic rings, the red box a conserved hydrogen bond acceptor, and the blue box the conserved basic center. Using the self-organizing algorithm, SPE,[31–33] these conserved features are then used to guide the three-dimensional alignment. This is heuristically described in the lower half of this figure. We emphasize here that the steps in the top half of the figure, one-dimensional alignment and identification of the conserved features, do not depend on any three-dimensional coordinates. The three-dimensional coordinates enter only in the steps in the bottom half of the figure.

Scoring Multiple Ligand Alignments. The similarity between two atoms from different molecules in a multiple ligand alignment, either one- or three-dimensional, is given by

$$S_{ij} = \begin{cases} 0 & \text{if atoms i and j are in the same molecule} \\ \sigma_{ij} f(d_{ij}) & \text{if atoms i and j are in different molecules} \end{cases} \quad (1)$$

where σ_{ij} is the similarity of the two atoms based solely on their atoms types, d_{ij} is the distance between atoms i and j,

and f is a distance dependent function that decreases as the distance between the two atoms increases, i.e., atoms that are closer to one another are deemed more similar. The atom type similarity matrix is that proposed by Labute and coworkers.[15] Here each atom is assigned a short bit string according to its pharmacophoric character: bits indicating whether the atom is aromatic, hydrophobic, a hydrogen bond acceptor, a hydrogen bond donor, positively charge, or negatively charge. The similarity between two atoms is simple: the weighted inner product of the two bit strings. For the distance dependent function in three dimensions we use the Gaussian overlap function used by Labute and coworkers.[15] In one dimension we use the overlap function used by Dixon and Merz[28]:

$$f(d) = \begin{cases} 1 - d & \text{for } 0 < d < 1 \\ 0 & \text{for } d > 1 \end{cases} \quad (2)$$

The One-Dimensional Alignment. A one-dimensional representation for a small molecule is simply the assignment of a single coordinate to each atom. Any molecule could be assigned many one-dimensional representations. For the work described here each molecule is assigned its canonical one-dimensional representation according to the originally published scheme.[28] In essence, the canonical one-dimensional representation for a molecule is derived via multidimensional scaling from the two-dimensional topological atomic distances so that the resulting one-dimensional atomic distances are as close as possible to their corresponding two-dimensional topological atomic distances. The canonical one-dimensional representation for a molecule is unique up to the translation of its one-dimensional coordinates, i.e. adding a constant to each coordinate, and to reorientation of its one-dimensional coordinates, i.e., multiplying the coordinates by –1. A key aspect of the one-dimensional representation is that it depends only on the topological pairwise atomic distances and does not depend on either a two-dimensional depiction or three-dimensional coordinates of the molecule.

The first step in the alignment procedure is to find the multiple one-dimensional alignment that maximizes the alignment score. This involves a search through the possible relative translations and orientations of the molecules being aligned. In our initial work we used a genetic algorithm to search for the optimal one-dimensional alignment.[29] While this approach proved to give near optimal alignments, it often takes in excess of an hour to run and limits the user's ability to generate different alignments. For this work, we adopted a different heuristic approach. In this approach an initial one-dimensional alignment of the N molecules is created at random. Using a small step size every potential translation of both orientations of molecule 1 is scored relative to the one-dimensional alignment of molecules $2-N$. The highest scoring alignment of molecule 1 is kept, effectively creating a new multiple one-dimensional alignment. The procedure is repeated successively with each molecule in turn. Once molecule N has been realigned, the individual one-dimensional realignment procedure is repeated starting with molecule 1 and iteratively realigning all N molecules. This iterative realignment continues until it converges, i.e., until one pass through all N molecules results in no change in the one-dimensional alignment. We generally run this one-dimensional alignment procedure 50 times with different starting positions, and in most cases the alignment converges on nearly the identical solution every time. In cases where multiple solutions are found from the 50 runs we only pass the best scoring one-dimensional alignment to the three-dimensional alignment procedure though in principle each unique multiple one-dimensional alignment could be passed into the three-dimensional alignment phase to generate alternate three-dimensional alignments.

Identification of the Conserved Features from the One-Dimensional Alignment. The next step in this procedure is to identify the conserved features, that is, those pharmacophoric features that are consistently well aligned in the multiple one-dimensional alignment. For example, the conserved features in the one-dimensional alignment shown in Figure 1 are highlighted by the vertical rectangles superimposed on the alignment in the upper right portion of the figure. It is important to note that conserved features are not necessarily conserved across all molecules of the data set. To identify the conserved features, first we partition each molecule into six types of features: hydrogen bond donors, hydrogen bond acceptors, positively charged groups, negatively charged groups, aromatic rings, and hydrophobic groups. Each pharmacophoric feature may consist of one atom, such as a hydrogen bond donor or acceptor, or multiple atoms, such as an aromatic ring. The similarity between two features from different molecules is simply the sum of the atomic similarities, using eq 1, over the atoms of the two features. To have nonzero similarity two features must be similar at the atomic level and must also overlap in the one-dimensional alignment.

To identify and rank the conserved features in the one-dimensional alignment we use the cPCA[30] algorithm. Except for the choice of the similarity measure the method provides a nonparametric way of identifying and ranking clusters. At the center of the cPCA method is a pair of linkage functions, π and F, that measure similarity between sets of features. Heuristically, the clusters contain features from multiple molecules that are well aligned in the one-dimensional alignment. The highest ranked clusters, via the function F, are taken as the conserved features.

The cPCA algorithm works as follows. Suppose the set of ligands being aligned has N total pharmacophoric features which we number arbitrarily from 1 to N. Set $W = \{1, 2, 3, \ldots, N\}$. For any subset H of W we define the two linkage functions, π and F, via

$$\pi(i;H) = \sum_{\substack{j \in H \\ j \neq i}} s_{ij} \quad (3)$$

where i is any member of H and s_{ij} is the similarity between features i and j. The function π is a measure of the consistency of a given member of H to the remaining members of H. Next we define F by

$$F(H) = \min_{i \in H} \{\pi(i;H)\} \quad (4)$$

The linkage function F quantifies the self-consistency of a set of features according to the pharmacophore feature that least belongs to the set. The function F is ultimately used to rank clusters and select the conserved features. The first conserved feature is the largest subset H^* of W that maximizes $F(H)$. Heuristically, the search for the optimal subset of features begins with the set of all features. The features are then removed one at a time according to the feature that

is least consistent, as defined by the function π, with the remaining features. This creates a sequence of sets of the features each set with one fewer feature than the previous. The set features in this sequence that maximizes the function F will be the conserved feature. Mathematically, the subset of W that maximizes F is found by the following algorithm:

Construct a sequence, $H_0, H_1, \ldots, H_{N-1}$, of subsets of W via:

a. $H_0 = W$.

b. Given H_k define H_{k+1} by selecting i in H_k so that $F(H_k) = \pi(i;H_k)$ and setting $H_{k+1} = H_k \backslash \{i\}$.

c. From $H_0, H_1, \ldots, H_{N-1}$, choose k^* such that $F(H_{k^*})$ is the largest of all the $F(H_k)$.

d. H_{k^*} as defined by step b is the maximal subset of W.

To prove that this algorithm does indeed give us the maximal subset of W we first note that if H^* is the maximal subset of W and $H^* \subset H \subseteq W$ with k in H such that $\pi(k;H) = F(H)$ then k cannot be in H^*. To prove this we note if to the contrary k were in H^* then

$$F(H^*) = \min_{i \in H^*} \pi(i;H^*) \leq \pi(k;H^*) \leq \pi(k;H) \leq F(H) \quad (5)$$

which contradicts the assumption that H^* is the largest maximizer of W.

Once the first conserved feature is identified, the second conserved feature is found by removing from consideration the features involved in the first conserved feature and repeating the procedure on the remaining subset of features. The third, fourth, etc. conserved features are found analogously. We emphasize that no three-dimensional coordinates are used to identify the conserved features. Rather the conserved features are identified solely from the multiple one-dimensional alignment of the molecules.

Generating the Three-Dimensional Alignment. Generating the Three-Dimensional Alignment. To convert a one-dimensional alignment to a three-dimensional alignment we use three types of constraints: force field constraints that encourage reasonable geometry for each ligand, alignment constraints that enforce the three-dimensional alignment of conserved features from the one-dimensional alignment, and knowledge-based constraints that allow the user to incorporate domain specific information such as a known active conformation. The three types of constraints are described below. Following the selection of constraints, the SPE algorithm[31–33] is used to generate an initial three-dimensional alignment that aligns the conserved features from the one-dimensional alignment and has reasonable low-energy conformations for each molecule. Briefly, random three-dimensional coordinates within a prescribed box, $15 \times 15 \times 15$ Å^3, are initially assigned to each atom. For example, this step is shown in the lower right corner of Figure 1. For a number of steps a constraint (force field, alignment, or knowledge-based) is selected at random, and the atoms involved in the constraint are moved to better satisfy the constraint. For example, if a constraint between two bonded atoms is selected, then the two bonded atoms are moved closer to or further from one another to more closely match their ideal bond length. The random selection of constraints and movement of the atoms involved is performed a fixed number of steps after which α is decreased. In practice, we perform 1,000,000 steps of randomly selecting a constraint and moving the atoms involved for each of $\alpha = 0.75$, 0.5, 0.25, and 0.1. The values of these parameters were not systematically tested but are likely excessive for the task, particularly the number of steps taken. Even with an excessive number of steps, the minimization step was the time limiting step in the process. Thus there was no need to optimize these parameters.

Force Field Constraints. To ensure that the final conformation of each molecule results in reasonable geometry, standard force field constraints for each molecule are included. These constraints are taken and used as described by Xu and co-workers[31] for conformational searching and are only briefly described here. The conformational constraints include terms for ideal bond lengths and ideal bond angles, terms to avoid internal steric clashes, and terms for defined stereocenters. Any two bonded atoms are constrained to a distance equal to their ideal bond length. Any two atoms separated by two bonds are constrained to the distance determined by their ideal bond angle and the ideal bond lengths of the two intervening bonds. Thus both bonded atoms and atoms separated by two bonds result in distance constraints. Unlike the bond length and bond angle constraints the steric constraints offer only a lower bound on the distance between the two atoms involved: any non-hydrogen atoms separated by four or more bonds are assumed to be separated by at least 4.0 Å. To these standard force field constraints, volume constraints are added to any defined stereocenter to ensure that the molecule adopts the correct stereochemistry. The volume constraints are implemented exactly as described by Xu et al.[31]

Alignment Constraints. The alignment constraints are based on the conserved features identified from the one-dimensional alignment. When an alignment constraint is selected, the atoms are moved closer to the center of mass of all the atoms involved in the constraint with the caveat that all atoms involved in the constraint from a single molecule are moved together as a rigid body so that their center of mass approaches the center of mass of all the atoms involved in the constraint. Mathematically, when an alignment constraint is selected, the coordinates for an atom from molecule n involved in the alignment constraint are changed via

$$X_{\text{new}} = X_{\text{old}} + (C - C_n)\alpha \quad (6)$$

where C is the center of mass of all the atoms involved in the constraint, C_n is the center of mass of the atoms from molecule n involved in the constraint, $0 < \alpha < 1$, and the old and new subscripts indicate the coordinates of the atom in question before and after the move. It is straightforward to show that after the move the center of mass of the atoms from molecule n involved in the constraint is $C_n(1-\alpha) + C\alpha$.

Knowledge-Based Constraints. In cases in which the user knows or has strong evidence for the active conformation of a molecule, additional constraints can be added to ensure that the active conformation will be the one that the corresponding ligand ultimately adopts in the three-dimensional alignment. There are two cases in which a user might wish to include additional knowledge-based constraints. The first case occurs when the user has a cocrystal structure. In this case, the user wants the template molecule to adopt its absolute coordinates so that the other molecules in the alignment will be aligned in the coordinate frame of the

protein structure from which the template molecule was taken. In this case each atom of a molecule with known geometry is constrained to adopt its input coordinates. These constraints force the molecule to adopt its exact input coordinates in the final alignment. The second case occurs when the user has strong evidence for the active conformation of any members of the data set. This often occurs when sufficient structure–activity data are available to identify a likely active conformation. In this case every pair of atoms is constrained so that their distance remains the same as in the input conformation. This allows knowledge-based constrained molecules to reorient to one another while still maintaining the user defined conformation. In the examples described below knowledge-based constraints are applied only in the CXCR3 example.

The force field constraints always significantly outnumber the alignment constraints. To counteract this imbalance, these two types of constraints are kept as separate lists. An alignment constraint is selected 25% of the time, and a force field constraint is selected the remaining 75% of the time. Knowledge-based constraints are included as part of the force field constraints because they are used to define the geometry of a molecule.

The Final Optimization of the Three-Dimensional Alignment. The final step in the generation of the three-dimensional alignment is to minimize the entire alignment using standard force field terms for each molecule individually and the intermolecular alignment score collectively to maintain the overall character of the alignment from the self-organizing procedure. Briefly, we use the Dreiding force field[34] to evaluate the internal energy of the individual molecules. The score for the intermolecular overlap, described above, is essentially that described by Labute and co-workers.[15] Since both the force field and alignment score are continuously differentiable, any gradient-based minimization algorithm can be used to optimize their combination. This step essentially allows the conformation of each molecule to be minimized while maintaining or improving the character of the three-dimensional alignment.

Timing. The time required for the three-dimensional alignment grows quadratically in the total number of atoms involved. The quadratic increase in time is due to the final gradient-based minimization of the three-dimensional alignment. Each pair of atoms involves one interaction, either an overlap interaction for two atoms from different molecules or a force field interaction for two atoms from the same molecule. Thus the number of interactions increases quadratically in the total number of atoms. For the 10 hERG channel blockers, the multiple one-dimensional alignment takes approximately 38 s, while the three-dimensional alignment procedure takes approximately 67 s per run. For the estrogen receptor agonists, the multiple one-dimensional alignment takes approximately 6 s, while the three-dimensional alignment procedure takes approximately 29 s per run. For the estrogen receptor antagonists, the multiple one-dimensional alignment takes approximately 16 s, while the three-dimensional alignment procedure takes approximately 53 s per run. All timing information is reported from calculations performed on a single processor of a 64 bit dual core 2.4 Ghz AMD Opteron processor.

Comparisons to Other Available Methodology. Here we compare our method to two commonly used state of the art techniques for generating ligand alignments: PHASE[35–37] and the MOE flexible alignment methodology.[15,38] PHASE generates pharmacophore models from a set of ligands which results in an implicit alignment whereas the MOE flexible alignment protocol explicitly searches through the translational, rotational, and conformational space of each ligand while trying to find the global optimal alignment score. Both programs were largely run from default parameters. In all cases, in order to find the optimal PHASE pharmacophore model the number of features and the minimal number of matched ligands was systematically varied.

Examples. *Example 1 – Identifying Common Pharmacophoric Elements in 10 hERG Channel Blockers.* To demonstrate our approach to three-dimensional alignment we align the 10 most potent hERG channel blockers from the study of Cavelli and co-workers.[39] The 10 ligands, shown in Figure 2, are astemizole, sertindole, E-4031, cisapride, droperidol, pimozide, dofetilide, haloperidol, terfenadine, and thioridazine. The molecules are oriented in this figure to approximate their relative orientation in the final three-dimensional alignment. These ligands have a total of 72 rotatable bonds which together with the six orientational and rotational degrees of freedom for nine of the 10 molecules makes a total of 126 degrees of freedom in the full three-dimensional alignment problem.

Figure 2 further highlights the top five conserved features from the one-dimensional multiple alignment. Conserved feature 1 consists of at least one phenyl ring from each of the 10 molecules. In addition five of the molecules contribute a second phenyl ring to the first conserved feature. The second conserved feature consists of aromatic rings from eight of the 10 molecules. Conserved feature 3 consists of the central basic center, a tertiary amine in all cases. It is noteworthy, however, that the basic centers of dofetilide and astemizole are not part of this conserved feature. The reason for this is that in optimizing the placement of other functional groups within the one-dimensional alignment, the basic centers for these two molecules are not able to align to those of the remaining eight. Both conserved features 4 and 5 consist of hydrogen bond acceptors. Conserved feature 4 contains acceptors from six of the 10 molecules, while conserved feature 5 contains acceptors from four of the 10. A conserved feature not shown in Figure 2 is the cluster of hydrophobic carbons surrounding the basic centers.

Figure 3 a shows the well aligned portions of the top 5 conserved features as they are aligned in three dimensions. Figure 3 b−f shows the alignments of all 10 hERG channel blockers, two at a time, relative to the alignment of the conserved features. All 10 molecules align a phenyl ring to conserved feature 1: the cluster of aromatic rings on the left of Figure 3a. The alignment of this phenyl ring is very tight. The rms deviations between these 10 phenyl rings ranges from essentially 0 to 1 Å. As noted above, five of the 10 molecules contribute a second phenyl ring to the first conserved feature. As is apparent in Figure 3e, the second phenyl ring of terfenadine and pimozide is nearly perfectly superimposed, whereas the second phenyl ring of sertindole and astemizole, Figure 3d, does not superimpose due to very different geometric constraints.

The second conserved feature, the cluster of aromatic rings on the right in Figure 3a, includes aromatic rings from eight of the 10 molecules. Only thioridazine and sertindole do not

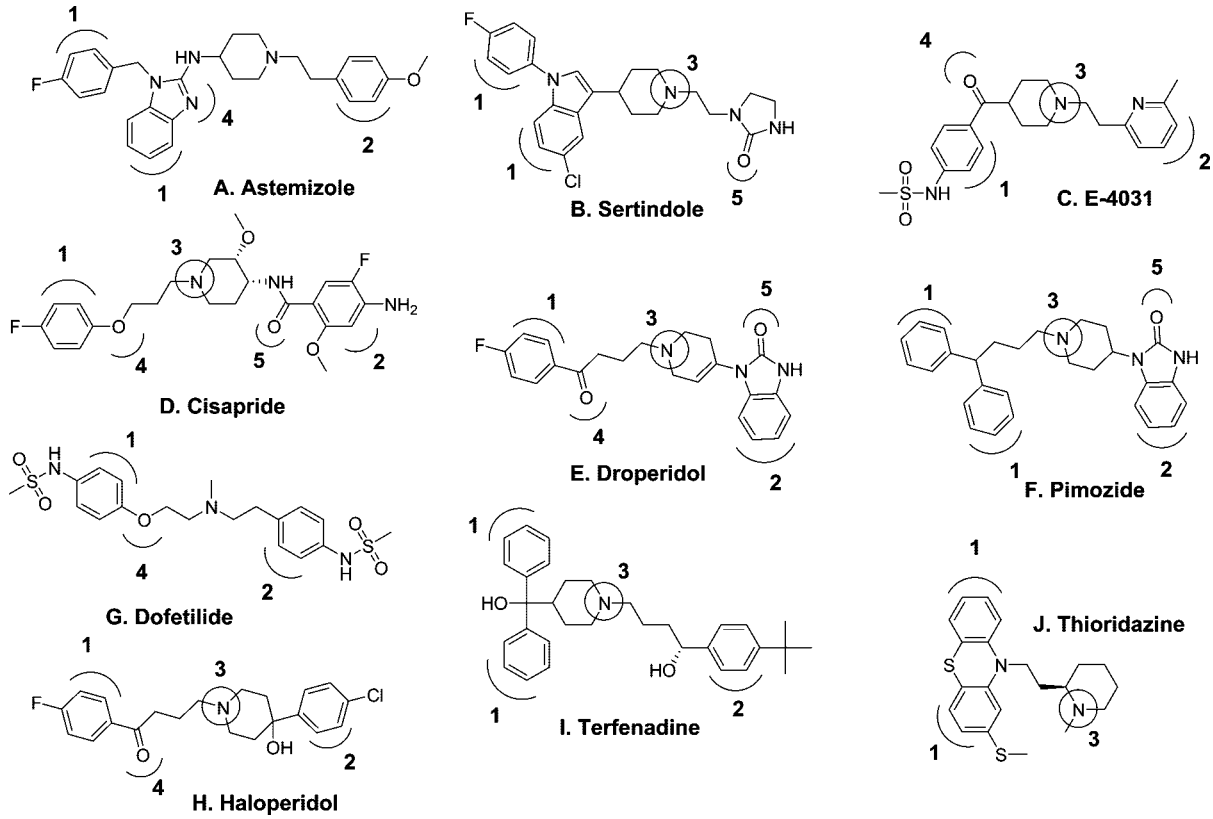

Figure 2. The 10 hERG channel blockers and the conserved features from the one-dimensional alignment: **A.** astemizole, **B.** sertindole, **C.** E-4031, **D.** cisapride, **E.** droperidol, **F.** pimozide, **G.** dofetilide, **H.** haloperidol, **I.** terfenadine, **J.** thioridazine. The molecules are drawn and oriented approximately as they oriented in the 3D alignment. The conserved features from the one-dimensional alignment of the five highest ranked clusters are labeled 1−5 throughout the figure.

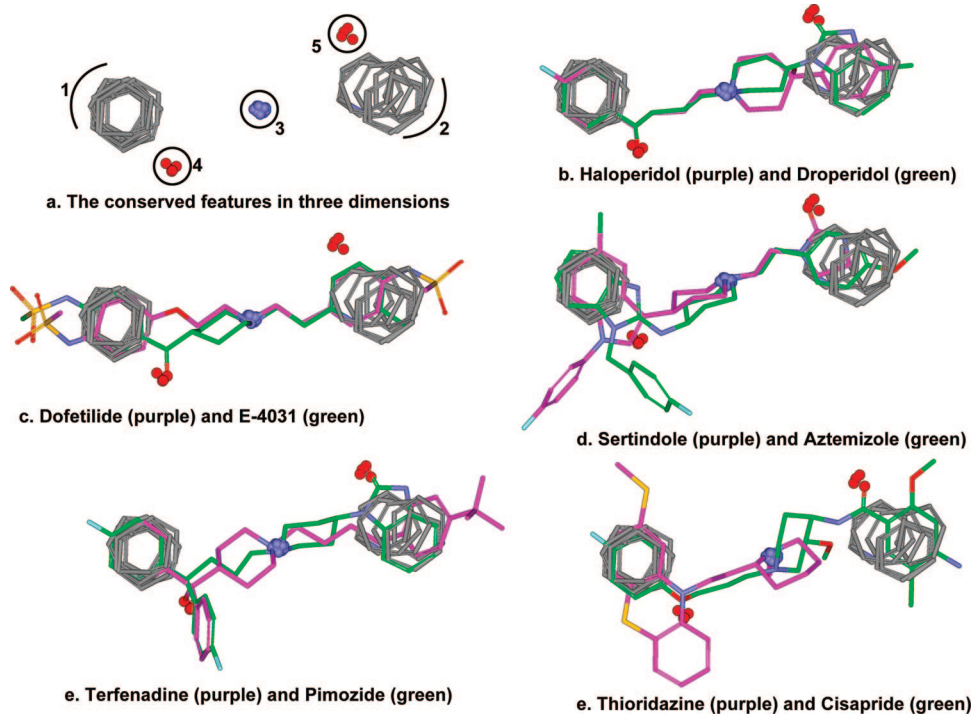

Figure 3. The three-dimensional alignment of the 10 hERG channel blockers. **A.** The alignment of the conserved features. The numbering of the features and the orientation of the ligands is consistent with that in Figure 2. **B.** The alignment of haloperidol and droperidol relative to the conserved features. **C.** The alignment of dofetilide and E-4031 relative to the conserved features. **D.** The alignment of sertindole and astemizole to the conserved features. **E.** The alignment of terfenadine and pimozide to the conserved features. **F.** The alignment of thioridazine and cisapride to the conserved features.

have an aromatic ring to match this conserved feature. As is apparent from Figure 3a the alignment of the aromatic rings from the second conserved feature is not as tight as that from the first conserved feature. The rms deviations for the aromatic rings in the second conserved feature are primarily above 1 Å and go as high as 2.2 Å.

The third conserved feature consists of the tertiary amine of eight of the 10 ligands. Due to the constraints from the one-dimensional alignments of other functional groups, the tertiary amines of dofetilide and astemizole were unable to align in one dimension to the tertiary amines of the remaining eight ligands. In the three-dimensional alignment, however, the tertiary amines of all 10 ligands align very well: the positions of the tertiary amine deviate by no more than 0.5 Å.

Conserved features 4 and 5 both represent hydrogen bond acceptors. The fourth conserved feature from the one-dimensional alignment contained acceptors from six of the molecules. The alignment of these acceptors changed dramatically in going to three dimensions. Due to the geometric constraints imposed by the alignment of other conserved features the acceptors of dofetilide, E-4031, and astemizole were unable to maintain sufficient alignment in three dimensions with the remainder of the acceptors in this feature. Terfenadine, which did not contribute a hydrogen bond acceptor to conserved feature 4, did ultimately align its OH to the acceptors of this feature. In contrast to conserved feature 4, the fifth conserved feature maintained the same four hydrogen bond acceptors in one and three dimensions.

This example demonstrates several important aspects of this approach to the multiple three-dimensional alignment problem. First the algorithm is able to generate reasonable alignments for flexible ligands from different chemotypes. Second, the one-dimesional alignments clearly lose some information when compared to three dimensions. For example, conserved feature 3 shows where functional groups that could not be aligned in one dimension were ultimately aligned in three simply due to the constraints of nearby functional groups. Conversely, conserved feature 4 gives an example where functional groups that were well aligned in one dimension could not be aligned in three due to the geometric constraints imposed by the alignment of other functionality. The key point of this example is that while the one-dimensional alignment loses information with respect to three dimensions, it is able to maintain sufficient information so that when the one-dimensional alignment constraints are combined with the typical constraints of low-energy conformations the extrapolation into three dimensions is able to recapture the lost information.

Of the examples presented here the three-dimensional alignment of the 10 hERG channel blockers is the most complex: there is no common scaffold upon which to build the alignment, and the ligands have a combined 72 rotatable bonds. Not surprisingly, the MOE alignment protocol which tries to search through the degrees of freedom to find the optimal alignment was unable to find a reasonable solution. The top scoring PHASE pharmacophore model, shown in Figure 4a, has two aromatic ring features, a hydrogen bond acceptor and a positive charge. It matches eight of the 10 ligands, missing only thioridazine and sertindole. The resulting alignment is shown in Figure 4b with a comparison to the alignment of all 10 hERG blockers using the method describe here (Figure 4c).

Without knowledge of the binding modes for these molecules in the hERG channel it is impossible to quantitatively compare the alignments of the hERG blockers. This example highlights the differences in the approaches to

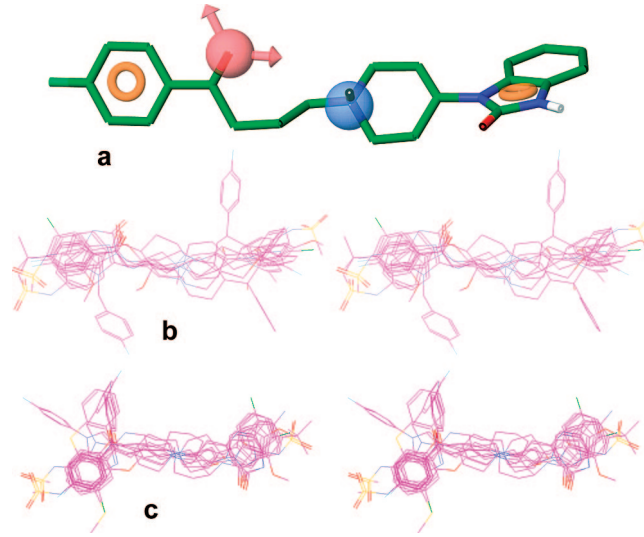

Figure 4. A. The top scoring PHASE pharmacophore model. **B.** A stereoview of the alignment of the eight hERG blockers that match the pharmacophore in **A**. **C.** For comparison purposes, a stereoview of the alignment of the 10 hERG blockers using the methodology described here.

deriving three-dimensional alignments and how this affects the final alignments. The key difference is that the optimal alignment from our method will sacrifice alignment of a single feature for partial alignment of several features: two features were matched by all 10 molecules, one feature by eight the 10, and two features by four of the 10 (see Figures 2 and 3). In contrast, pharmacophore approaches emphasize finding a few features that are conserved throughout a ligand data set, all four features in eight of the 10 molecules (Figure 4a). Fundamentally different approaches will ultimately lead to different results which is particularly important in cases with nonobvious answers such as this one.

Finally, to verify that the one-dimensional multiple alignment provides constraints that significantly guide the three-dimensional alignment we tested three alternate approaches to generate constraints. These three methods are intended to act as controls to better quantify the importance of the one-dimensional alignment. The first control was to include no alignment constraints. For this approach there is no attempt to align any features during the stochastic embedding step. Alignment occurs only during the minimization step. The second control was to generate a fixed number of alignment constraints purely at random. To generate a random alignment constraint, we select at random a single feature of each molecule allowing no feature to occur in more than one constraint. For this example we chose to select 6 random constraints. The third method is to identify the conserved features exactly as in our alignment protocol but then create constraints that randomly switch each feature in the conserved feature with one of the same pharmacophoric types from the same molecule. Our full three-dimensional alignment technique generated an alignment of the 10 hERG inhibitors with an alignment overlap score of 390. The first control approach, no constraints, generated an alignment with an overlap score of 170. The last two control approaches generated alignments with overlap scores below 70. While the score should not be interpreted too literally the differences between our method and the control methods are highly significant. Visual examination of the three control align-

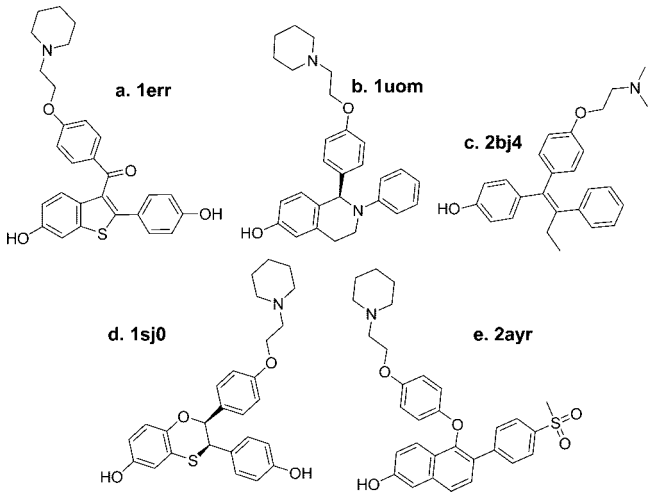

Figure 5. The five estrogen receptor antagonists. **A.** From 1err. **B.** From 1uom. **C.** From 2bj4. **D.** From 1sj0. **E.** From 2ayr. The molecules are drawn so as to approximate their relative orientations in the X-ray alignment.

ments indicates that the significantly lower alignment overlap scores are consistent with poor three-dimensional alignments; the data are not shown.

Example 2: Reproducing the Crystallographic Alignment of Five Estrogen Receptor Antagonists. For the second example we use five estrogen receptor antagonists extracted from PDB structures 1err,[40] 1uom,[41] 2bj4,[42] 1sj0,[43] and 2ayr[44] with resolutions of 2.6, 2.3, 2.0, 1.9, and 1.9 Å. The structures of these five ligands are shown in Figure 5, oriented approximately as they bind to the estrogen receptor. This case represents a difficult example for de novo alignment because the ligands have a combined total of 34 rotatable bonds which together with the six orientational and translational degrees of freedom for four of the five molecules makes a total of 58 degrees of freedom.

First, the protein structures in the five PDB files were aligned using the sequence and structure alignment capability within MOE.[38] The structures of the estrogen receptor from these five cocrystal structures superimpose with pairwise rms deviations over the C_α atoms ranging from 0.4 to 1.1 Å. These rms deviations give some insight into the reasonable achievable limits for alignment procedures with these five ligands. The resulting coordinates for the five ligands were extracted from the protein alignment which we subsequently refer to as the X-ray alignment. The X-ray alignment of the five estrogen receptor antagonists is shown in Figure 6a. Next, using no information from the crystal structures, the three-dimensional alignment procedure was used to align these five ligands starting from SMILES strings. This we refer to as the 3D alignment. The 3D alignment is shown in Figure 6b in approximately the same orientation as the X-ray alignment. The most noticeable difference between the two alignments is that the 3D alignment is more tightly aligned. In particular, the phenyl moiety depicted in the lower right corner of the alignment is nearly superimposed on an atom by atom basis in the 3D alignment but shows considerable variation in the X-ray alignment. In the 3D alignment, these phenyl rings deviate from 0.3 to 0.7 Å, whereas in the X-ray alignment they deviate up to 2.4 Å.

To quantify the quality of the alignment, a rigid body transformation was calculated that best superimposed the coordinates of the 2bj4 ligand in the 3D alignment to those of the X-ray alignment. This single rigid body transformation was then applied to the entire 3D alignment to bring it into the frame of reference of the X-ray alignment. This allows for a quantitative comparison of the 3D and X-ray alignments. A pairwise comparison of the two alignments after the rigid body transformation is shown in Figure 7. For four of the five ligands the rms deviations are approximately 1 Å with the majority of the deviation coming from the group containing the amine depicted at the top in Figures 5−7. The largest deviation between the two alignments is with the 1err ligand where the rms deviation is 1.4 Å. The additional deviation arises from the pendant phenol on the benzthiophene depicted in the lower right-hand portion of Figure 7a. The deviation here occurs because the 3D alignment protocol strives for a tighter alignment in this portion of the molecules than that observed in the X-ray alignment. As an example consider 1err compared to 1sj0 both of which have a phenol in this portion of the molecule. The interaction of this portion of these two molecules with the estrogen receptor is noticeably different. The right-hand side OH of the 1err ligand interacts with Nδ of His524, whereas the right-hand side OH of the 1sj0 ligand interacts with both the Nδ of His524 and the backbone carbonyl oxygen of Gly521. It is not at all clear why the 1err ligand is unable to or prefers not to make this second hydrogen bond.

The top ranked PHASE pharmacophore model for the five estrogen receptor antagonists is shown in Figure 8a. This model finds all three conserved aromatic rings, and it treats the ether oxygen common to all five ligands as a hydrogen bond acceptor, the tertiary amine as a positive charge, and the hydroxyl depicted in the left-hand side of all molecules in Figure 5 as a hydrogen bond donor. This pharmacophore model is consistent with the conserved features of the X-ray alignment, Figure 6a. The alignment from this pharmacophore model is shown in Figure 8b. As above the 1err ligand was used to bring the PHASE alignment into the same frame of reference as the X-ray alignment. The resulting rms deviations are 1err - 2.6 Å, 1uom - 2.2 Å, 2bj4 - 1.6 Å, 1sj0 - 3.0 Å, and 2ayr - 2.3 Å. The rms deviations are, however, misleading. The majority of the deviation arises from placement of the amine. The exception is with the 1sj0 ligand (purple). The bicyclic portion of the 1sj0 ligand is twisted relative to the X-ray orientation. This likely arises from an incomplete conformational coverage.

The top scoring alignment from the MOE flexible alignment of the five estrogen receptor antagonists is shown in Figure 8c. When using the 1err ligand as the reference ligand to bring the alignment into the same frame of reference as the X-ray alignment, 3 of the 5 have rms deviations at or below 2 Å: 1err - 1.6 Å, 1uom - 1.8 Å, and 2ayr - 2.1 Å. The 2bj4 ligand, cyan, is oriented with its amine properly aligned, but the remaining portion of the molecule is horizontally flipped which results in an rms deviation of 4.9 Å. The 1sj0 (purple) ligand is incorrectly oriented and has an rms deviation of 8.0 Å. Since this procedure is stochastic, we ran it several times on this data set and obtained qualitatively similar results each time. This result suggests that with this data set the full translational, rotational, and conformational space is too vast and complex to allow for a sufficiently thorough search.

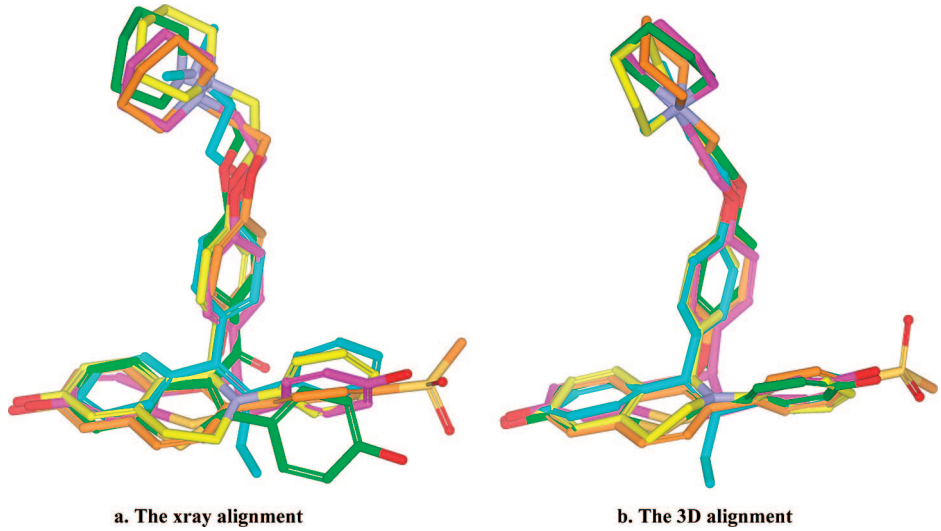

Figure 6. The estrogen receptor antagonist three-dimensional alignment. **A.** The X-ray alignment. The X-ray alignment was generated by aligning the estrogen receptor structures of the 5 structures shown in Figure 5and then extracting the resulting coordinates of the ligands. **B.** The 3D alignment. The 3D alignment was generated using the procedure described here. The two alignments are shown in approximately the same orientation. In both cases the ligands are colored as 1err – green, 1sj0 – magenta, 1uom – yellow, 2ayr – orange, and 2bj4 – cyan.

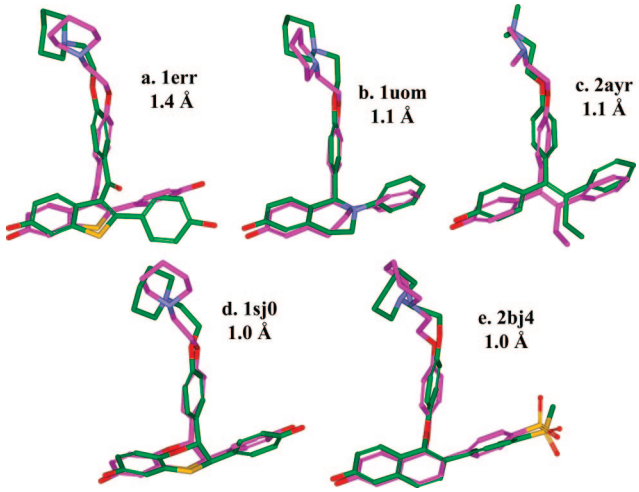

Figure 7. A comparison of the 3D alignment to the X-ray alignment of the five estrogen receptor antagonists. This figure was created by choosing the rigid body transformation that best superimposed the coordinates of the 2bj4 ligand from the 3D alignment to those of the X-ray alignment. This rigid body transformation was then applied to the entire 3D alignment (magenta) to bring it into the same coordinate frame as the X-ray alignment (green). **A.** 1err. **B.** 1uom. **C.** 2ayr. **D.** 1sj0. **E.** 2bj4. All numbers shown are rms deviations between the coordinates from the X-ray alignment and those from the transformed 3D alignment.

Example 3 – Reproducing the Crystallographic Alignment of 10 Estrogen Receptor Agonists. As a third example we extracted 10 estrogen receptor agonists from publicly available cocrystal structures (see Figure 9): 1a52,[45] 1gwq,[46] 1yy4,[47] 1x7r,[48,49] 2b1z,[47] 3erd,[50] 1x7e,[48,49] 2i0j[51], 1l2i,[52] and 1yye[47] with resolutions of 2.8, 2.5, 2.7, 2.0, 1.8, 2.0, 2.8, 2.9, 2.0, and 2.0 Å. As with the estrogen receptor antagonist case we created an X-ray alignment of the 10 ligands. Similar to the ER agonists case above the rms deviations for the proteins in this alignment range from 0.25 to 1.1 Å. Though both cases use the estrogen receptor, this case presents a different challenge for the alignment technique. In this case the ligands are relatively rigid having a total of only 13 rotatable bonds (67 total degrees of freedom). The problem

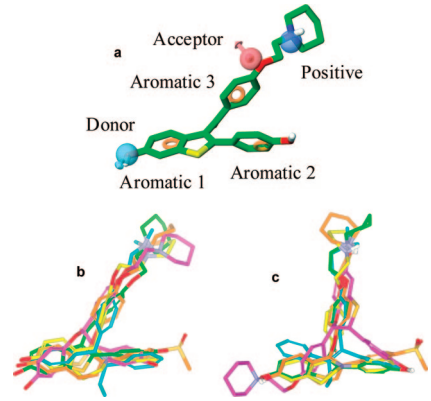

Figure 8. Alignments of the five estrogen antagonists with the other methodologies. **A.** The top scoring PHASE pharmacophore model. **B.** The resulting alignment from the pharmacophore model in **A**. **C.** The alignment from the MOE flexible alignment protocol. In both cases the ligands are colored as in Figure 6: 1err – green, 1sj0 – magenta, 1uom – yellow, 2ayr – orange, and 2bj4 – cyan.

is, however, made difficult because the molecules are 4-fold pseudosymmetric, roughly along both the horizontal and vertical axes of Figure 9.

The overall 3D and X-ray alignments are shown in Figure 10. As with the estrogen receptor antagonist case above, the area of greatest deviation occurs where the 3D alignment has portions of the molecules superimposed more tightly than that observed in the X-ray alignment. In particular, the largest difference occurs in the right-hand portion of the alignment: in the 3D alignment the rms deviations between these phenols range from 0.1 to 0.4 Å, whereas in the X-ray alignment the rms deviations between these phenols often exceed 1.0 Å with a maximum of 1.2 Å.

As with the estrogen receptor antagonists, the 3D alignment was brought into the same frame of reference as the X-ray alignment via choosing the rigid body transformation that best superimposed the coordinates of one of the ligands, in this case 1a52, of the 3D alignment to those of the X-ray alignment. The resulting comparison between the two alignments is shown in Figure 11. The ligands from the 1a52,

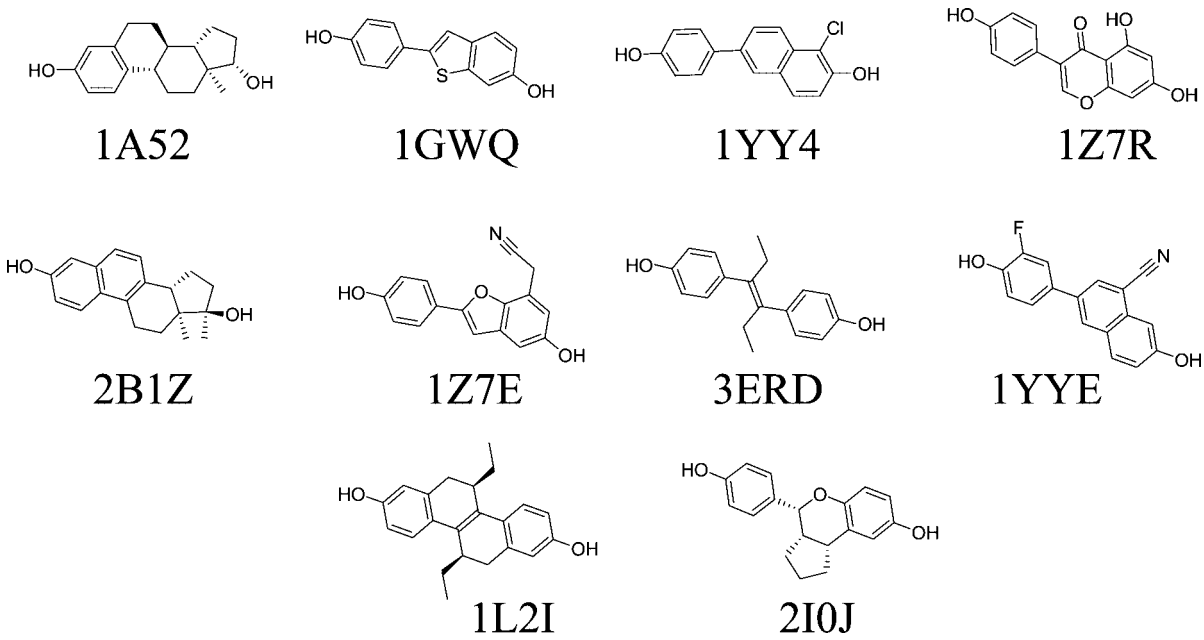

Figure 9. The 10 estrogen receptor agonists. **A.** From 1a52. **B.** From 1gwq. **C.** From 1yy4. **D.** From 1x7r. **E.** From 2b1z. **F.** From 1z7e. **G.** From 3erd. **H.** From 1yye. **I.** From 1l2i. **J.** From 2i0j. The molecules are drawn so as to approximate their relative orientations in the X-ray alignment.

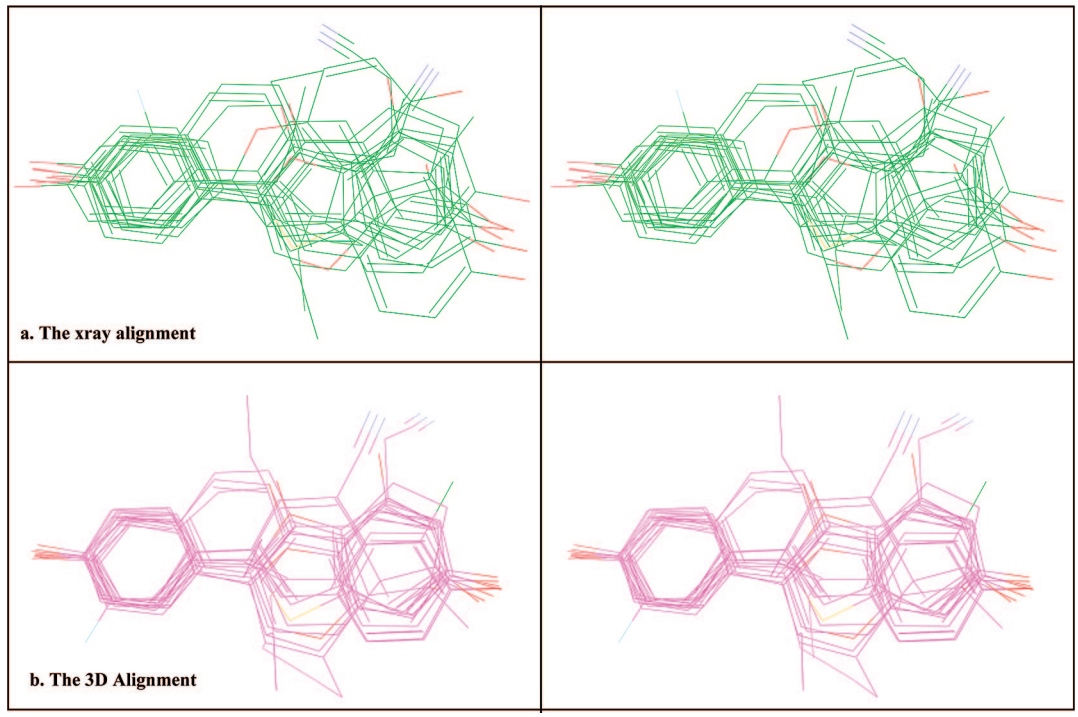

Figure 10. A stereoview of the X-ray and 3D alignments of the 10 estrogen receptor agonists. **A.** The X-ray alignment was generated by aligning the estrogen receptor structures mentioned in Figure 9 and extracting the resulting coordinates. **B.** The 3D alignment was generated using the procedure described here. The chief difference between the two alignments occurs in the portion of the molecules depicted in the right-hand side of the figure. The difference is that the 3D alignment has superimposed the phenol in this portion of the alignment much more tightly than that observed in the X-ray alignment.

1gwq, 1yy4, 1x7r, and 2b1z structures are all aligned with rms deviations well below 1Å. The remaining five ligands demonstrate the challenges inherent in building and assessing 3D alignments and are discussed below.

For the 1x7e ligand, the majority of the deviation occurs with the pendant cyano functionality. The rms deviation for the full molecule is 1.5 Å. If the cyano is excluded from the rms deviation calculation, the rms deviation drops to 0.7 Å. The positioning of the cyano changes freely from run to run indicating that its position is not determined by the constraints in the alignment.

The ligands from crystal structures 3erd and 1l2i both exhibit moderate rms deviations (1.2 and 0.9 Å) for essentially the same reasons. In these two cases the pendent ethyl groups depicted in the center of parts g and i of Figure 11 lead to the increased rms deviations. From run to run the orientation of these ethyl groups varies significantly. The only constraint is that the ethyl groups of 3erd align to those of

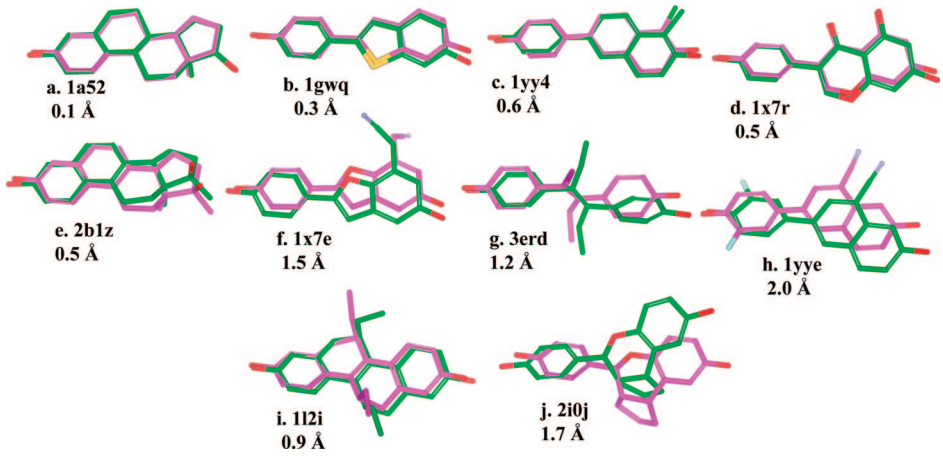

Figure 11. A comparison of the 3D alignment to the X-ray alignment of the 10 estrogen receptor agonists. This figure was created by choosing the rigid body transformation that best superimposed the coordinates of the 1a52 ligand from the 3D alignment to those of the X-ray alignment. This rigid body transformation was then applied to the entire 3D alignment (magenta) to bring it into the same coordinate frame as the X-ray alignment (green). **A.** 1a52. **B.** 1gwq. **C.** 1yy4. **D.** 1x7r. **E.** 2b1z. **F.** 1x7e. **G.** 3erd. **H.** 1yye. **I.** 1l2i. **J.** 2i0j.

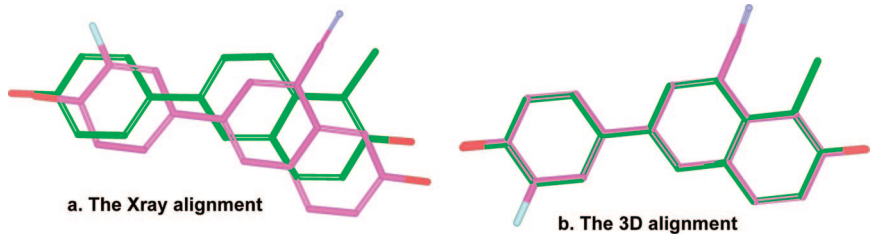

Figure 12. The alignments of the 1yye (purple) and 1yy4 (green) ligands. **A.** The alignment taken from the X-ray alignment. **B.** The alignment taken from the 3D alignment of the 10 estrogen receptor agonists shown in Figure 9.

1l2i. Again, this indicates that the orientation of these ethyl groups is not sufficiently constrained by the remainder of the ligands in the alignment.

The large rms deviations for the final two ligands, 1yye and 2i0j, arise largely because the 3D alignment strives to align the hydroxyls depicted on the right-hand side of Figure 9. As an example, consider the 1yye and 1yy4 ligands. These two ligands have very similar structures: their maximal common substructure covers all but one atom of the 1yy4 ligand and three atoms of the 1yye ligand. As shown in Figure 12a , these two ligands deviate significantly in the X-ray alignment, whereas in the 3D alignment the common scaffolds of these two ligands are aligned directly on top of one another: the rms deviations for the maximal common substructure of these two ligands are 0.06 Å in the 3D alignment and 1.0 Å in the X-ray alignment. The reason the two ligands differ quite markedly in the 1yye and 1yy4 cocrystal structures despite their very similar structures is that if 1yye ligand were placed directly in the 1yy4 ligand binding mode the additional cyano and fluorine would both lead to steric clashes with the estrogen receptors. The nitrogen of the cyano would come within approximately 2.2 Å of the Oγ of ILE376, and the fluorine would come within approximately 2.0 Å of the backbone carbonyl oxygen of LEU339. In the absence of crystallographic data it would be difficult to exclude the possibility of the 3D alignment of these two ligands in favor of that in the X-ray alignment.

For the estrogen receptor agonists both PHASE and the MOE flexible alignment protocol found reasonable alignments for the 10 ligands. The top ranked PHASE pharmacophore model, shown in Figure 13a, consists of three features: a hydrogen bond donor, a hydrogen bond acceptor, and an aromatic ring. The resulting alignment is shown in Figure 13b. As above the 1a52 ligand was used as the reference ligand to place the PHASE alignment in the same frame of reference as the X-ray alignment. The resulting rms deviations are 1a52 - 1.0 Å, 1gwq - 1.4 Å, 1yy4 - 1.1 Å, 1x7r - 3.1 Å, 2b1z - 2.7 Å, 1z7e - 7.3 Å, 3erd - 3.1 Å, 1yye - 2.7 Å, 1l2i - 2.5 Å, and 2i0j - 1.7 Å. The ligands with rms deviations <2.0 Å are all oriented essentially correctly. The ligands with rms deviations between 2 Å and 3.0 Å are primarily flipped vertically. The ligand with a large rms deviation, 1z7e, is flipped horizontally. In essence the model is a valid pharmacophore model. In all likelihood most of the ligands would match this model in several orientations including one that is consistent with the X-ray alignment.

The alignment from the MOE flexible alignment protocol is shown in Figure 13c. As with the alignment from the PHASE pharmacophore model this alignment is essentially correct. The alignment of the hydroxyls and the aromatic rings on both ends is evident. The rms deviations between the MOE alignment and the X-ray alignment are, however, quite a bit higher in this case. Again using the 1a52 ligand to bring the MOE alignment into the same frame of reference as the X-ray alignment four of the 10 have rms deviations >7 Å, i.e., are flipped horizontally. This highlights how difficult the full three-dimensional search is even with relatively rigid molecules.

Example 4 – The Alignment of CXCR3 Agonist to Peptides from IP-10. During a high throughput screening campaign for CXCR3 receptor antagonists, a series of peptidic CXCR3 agonists were discovered.[53] The natural ligand for CXCR3 is IP-10, a small protein consisting of 77 amino acids. Since there are three publicly available crystal struc-

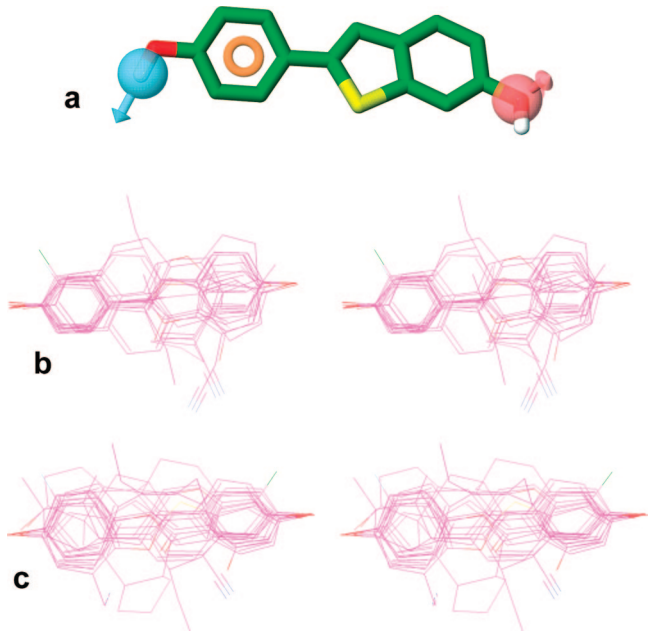

Figure 13. Alignments of the 10 estrogen receptor agonists with the other methodologies. **A.** The top scoring PHASE pharmacophore model. **B.** A stereoview of the resulting alignment from the pharmacophore model in **A**. **C.** A stereoview of the alignment from the MOE flexible alignment protocol.

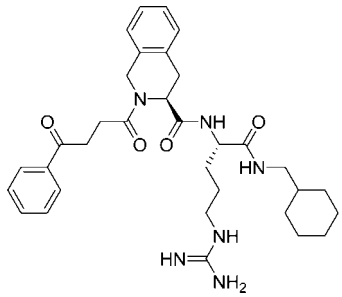

Figure 14. PS372424, a CXCR3 antagonist.

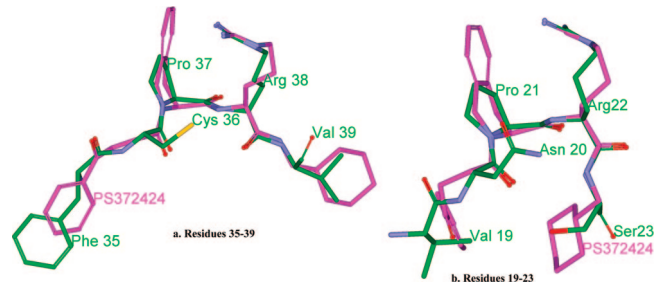

Figure 15. The top scoring three-dimensional alignments of PS372424 to pentapeptides of IP-10. **A.** Residues 35–39 resulted in the top alignment score with PS372424. **B.** Residues 19–23 resulted in the second best alignment score with PS372424.

tures of IP-10, (1o7y, 1o7z, and 1o80[54]), the question arose whether these small molecule agonists mimic any portion of IP-10 and ultimately whether such alignments could be used for developing hypotheses as to how to simplify these molecules. We emphasize that there is no direct evidence that any small molecule agonist binds to CXCR3 in the same manner as any portion of IP-10. Rather we present this example to show how three-dimensional alignment can be used in a nonstandard manner to generate reasonable testable hypotheses.

To develop a hypothesis for the binding of the CXCR3 agonists found in this study, we selected PS372424 (Figure 14) as a representative small molecule agonist from the high throughput screen. Starting from the initial assumption that the agonists mimic short lengths of the polypeptide chains from the binding domain of IP-10 the most complete of the three IP-10 structures, 1o80, was systematically split into pentapeptides, i.e., into residues 1–5, residues 2–6, residues 3–7, etc. The small molecule agonists were then aligned to each pentapeptide with knowledge-based constraints added to ensure that the coordinates of the IP-10 penta-peptides would adopt those found in the X-ray structure 1o80.

There are two peptides that based on the score of their alignment to the PS372424 stand out from the rest. These correspond to residues 35–39 which is the best scoring peptide and residues 19–23 which is the second best scoring peptide. The resulting alignments are shown in Figure 15. In both cases backbone carbon, oxygen, and nitrogen atoms of both of the IP-10 sequences are well aligned to corresponding atoms of PS372424. In addition, both peptides contain a central Pro-Arg motif that aligns with the tetrahydroisoquinoline-arginine motif of PS372424.

There is experimental evidence that implicates both of these portions of IP-10 as being important for interacting with or activating CXCR3. Booth and co-workers[55,56] investigated changes in NMR spectra of IP-10 alone compared with IP-10 with an N-terminal portion of CXCR3 believed to be important for interactions with IP-10. These data suggest that Val7, Arg8, Gln17, Val19, Arg38, and Thr44 of IP-10 all undergo spectral changes upon introduction of the N-terminal portion of CXCR3. Since Val7 and Arg8 of IP-10 comprise a portion of a common surface with Pro37-Arg38, the spectral changes of Val7, Arg8, and Arg38 upon introduction of the N-terminus of CXCR3 suggest residues 35–39 and the surrounding area are directly involved in interacting with this portion of CXCR3. Campanella and co-workers[57] further studied the binding of IP-10 to CXCR3 via systematic mutations of IP-10. Their results indicate that Arg8, Arg22, Lys46, and Lys47 all had significant effects on either IP-10/CXCR3 binding or IP-10 induced CXCR3

signaling. Since Lys46 and Lys47 share a common surface with Pro21-Arg22, these data implicate the pentapeptide 19–23 as being a key determinant in IP-10/CXCR3 binding and subsequent CXCR3 activation.

The chief difference in the two peptides is that residues 35–39 adopt an extended conformation, whereas residues 19–23 adopt a β-turn. This difference could be tested experimentally by developing various constraints that either force the extended conformation of PS372424 thereby mimicking residues 35–39 or constraints that force the β-turn conformation of PS372424 thereby better mimicking residues 19–22. Of particular note is that the alignment to residues 19–23 suggests the hypothesis that the hydrophobic collapse of the phenyl and the cyclohexane rings at opposite ends of the PS372424 potentially function to hold PS372424 in a pseudobeta turn conformation, and therefore these two groups might not be forming critical direct interactions with CXCR3. This hypothesis suggests that the lipophilic groups of PS372424 could be replaced by appropriate cyclic constraints. Doing so would both reduce significantly the molecular weight and diminish the peptidic character of this series of molecules both of which would be necessary to progress this series toward useful biological tools.

CONCLUSIONS

We have presented a method for the fully flexible simultaneous superposition of several molecules. Because our approach addresses much of the alignment problem in one dimension it is able to handle data sets with a large amount of conformational flexibility: the example with the 10 hERG channel blockers has a total of 72 rotatable bonds which together with the 54 relative orientational and translational degrees of freedom makes a total of 126 degrees of freedom. Further we have shown, with the estrogen receptor agonist and antagonist cases, that with no structural information whatsoever the method is capable of reproducing to reasonable accuracy alignments derived from crystallographic data. It is particularly noteworthy that even with a combined 34 rotatable bonds as in the estrogen receptor antagonist case, our approach to three-dimensional alignment is able to produce alignments comparable to the X-ray alignments.

This method is fundamentally dependent on the ability to derive reasonable one-dimensional multiple alignments. For certain classes of molecules, such as those with large macrocyles, the one-dimensional representation is not appropriate, and it is likely this method will not provide useful three-dimensional alignments for these types of molecules. The distortion between the topological pairwise distances in the corresponding distances in the one-dimensional representation is a good indicator of when the one-dimensional representation is no longer appropriate for a class of molecules. We have shown in previous work[29] that >90% of molecules relevant to drug discovery work exhibit a distortion less than 10% between their topological and one-dimensional distances. This implies that this method will be applicable to a wide range of relevant alignment problems. The second class of problems where this technique is less likely to work when compared to explicit three-dimensional methods are those where two molecules bind through a common pharmacophore using very different topology to present the binding elements.

Caveats abound for reproducing X-ray alignments. For example, cases where ligands occupy different areas of a binding site or where ligands interact in fundamentally different ways will continue to provide difficult challenges. Beyond radically different binding modes, subtle steric effects are difficult to predict without knowledge of the binding site. The example with the estrogen receptor agonists with the ligands from structures 1yye and 1yy4, Figure 12 12, is particularly enlightening in this regard. In this case, the two ligands have very similar structures, yet the rms deviation between their common scaffolds from the X-ray alignment is >1.0 Å due largely to the steric constraints of the binding site. Without knowledge of the steric environment of the ligands, alignment algorithms will invariably superimpose their common scaffold. Thus it is important to bear in mind that ligand superpositions are simply hypotheses as to how the ligands align in the appropriate binding site and should be used to derive testable hypotheses for different and improved ligands. For this reason, the CXCR3 example and the hERG example were presented to demonstrate the utility of the method in developing reasonable hypotheses within ligand data sets. Testable hypotheses of this nature can be directly addressed through iterative traditional medicinal chemistry and model development such as deriving conformational constraints or exchanging of the aligned groups. By unifying model development and experimental work through hypothesis generation, a deeper understanding of the key interactions with the target protein can be achieved.

REFERENCES AND NOTES

(1) Patny, A.; Desai, P. V.; Avery, M. A. Homology modeling of G-protein-coupled receptors and implications in drug design. *Curr. Med. Chem.* **2006**, *13*, 1667–1691.
(2) Becker, O. M.; Shacham, S.; Marantz, Y.; Noiman, S. Modeling the 3D structure of GPCRs: advances and application to drug discovery. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 353–361.
(3) Archer, E.; Maigret, B.; Escrieut, C.; Pradayrol, L.; Fourmy, D. Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **2003**, *24*, 36–40.
(4) Ballester, P. J.; Richards, W. G. Ultrafast shape recognition to search compound databases for similar molecular shapes. *J. Comput. Chem.* **2007**, *28*, 1711–1723.
(5) Cho, S. J.; Sun, Y. FLAME: a program to flexibly align molecules. *J. Chem. Inf. Model.* **2006**, *46*, 298–306.
(6) Cosgrove, D. A.; Bayada, D. M.; Johnson, A. P. A novel method of aligning molecules by local surface shape similarity. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 573–591.
(7) Daeyaert, F.; de Jonge, M.; Heeres, J.; Koymans, L.; Lewi, P.; van den Broeck, W.; Vinkers, M. Pareto optimal flexible alignment of molecules using a non-dominated sorting genetic algorithm. *Chemom. Intell. Lab. Syst.* **2005**, *77*, 232–237.
(8) Goldman, B. B.; Wipke, W. T. Quadratic shape descriptors. 1. Rapid superposition of dissimilar molecules using geometrically invariant surface descriptors. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 644–658.
(9) Jewell, N. E.; Turner, D. B.; Willett, P.; Sexton, G. J. Automatic generation of alignments for 3D QSAR analyses. *J. Mol. Graphics Modell.* **2001**, *20*, 111–121.
(10) Kearsley, S. K.; Smith, G. M. An alternative method for the alignment of molecular structures: maximizing electrostatic and steric overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615–633.
(11) Klebe, G.; Mietzner, T.; Weber, F. Different approaches toward an automatic structural alignment of drug molecules: applications to sterol mimics, thrombin and thermolysin inhibitors. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 751–778.
(12) Klebe, G.; Mietzner, T.; Weber, F. Methodological developments and strategies for a fast flexible superposition of drug-size molecules. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 35–49.
(13) Jain, A. N. Morphological similarity: a 3D molecular similarity method correlated with protein-ligand recognition. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 199–213.

(14) Jain, A. N. Ligand-based structural hypotheses for virtual screening. *J. Med. Chem.* **2004**, *47*, 947–961.
(15) Labute, P.; Williams, C.; Feher, M.; Sourial, E.; Schmidt, J. M. Flexible alignment of small molecules. *J. Med. Chem.* **2001**, *44*, 1483–1490.
(16) Lemmen, C.; Lengauer, T. Time-efficient flexible superposition of medium-sized molecules. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 357–368.
(17) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: a method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41*, 4502–4520.
(18) Lemmen, C.; Hiller, C.; Lengauer, T. RigFit: a new approach to superimposing ligand molecules. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 491–502.
(19) Lemmen, C.; Lengauer, T. Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215–232.
(20) Marialke, J.; Korner, R.; Tietze, S.; Apostolakis, J. Graph-based molecular alignment (GMA). *J. Chem. Inf. Model.* **2007**, *47*, 591–601.
(21) Mestres, J.; Rohrer, D. C.; Maggiora, G. M. A molecular field-based similarity approach to pharmacophoric pattern recognition. *J. Mol. Graphics Modell.* **1997**, *15*, 114–121103–116.
(22) Miller, M. D.; Sheridan, R. P.; Kearsley, S. K. SQ: a program for rapidly producing pharmacophorically relevent molecular superpositions. *J. Med. Chem.* **1999**, *42*, 1505–1514.
(23) Richmond, N. J.; Willett, P.; Clark, R. D. Alignment of three-dimensional molecules using an image recognition algorithm. *J. Mol. Graphics Modell.* **2004**, *23*, 199–209.
(24) Robinson, D. D.; Lyne, P. D.; Richards, W. G. Partial molecular alignment via local structure analysis. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 503–512.
(25) Rush, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **2005**, *48*, 1489–1495.
(26) Todorov, N. P.; Alberts, I. L.; Esch, I. J.; Dean, P. M. QUASI: A Novel Method for Simultaneous Superposition of Multiple Flexible Ligands and Virtual Screening Using Partial Similarity. *J. Chem. Inf. Model.* **2007**, *47*, 1007–1020.
(27) Grant, J. A.; Gallardo, M. A.; Pickup, B. T. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1996**, *17*, 1653–1666.
(28) Dixon, S. L.; Merz, K. M. One-dimensional molecular representations and similarity calculations: methodology and validation. *J. Med. Chem.* **2001**, *44*, 3795–3809.
(29) Wang, N.; DeLisle, R. K.; Diller, D. J. Fast small molecule similarity searching with multiple alignment profiles of molecules represented in one-dimension. *J. Med. Chem.* **2005**, *48*, 6980–6990.
(30) Anghelescu, A. V.; Muchnik, I. B. Combinatorial PCA and SVM methods for feature selection in learning classifications (applications to text categorization). *Int. Conf. Integr. Knowledge Intensive Multi-Agent Syst.* **2003**, *2003*, 491–496.
(31) Xu, H.; Izrailev, S.; Agrafiotis, D. K. Conformational sampling by self-organization. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1186–1191.
(32) Agrafiotis, D. K.; Xu, H. A self-organizing principle for learning nonlinear manifolds. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15869–15872.
(33) Agrafiotis, D. K. Stochastic proximity embedding. *J. Comput. Chem.* **2003**, *24*, 1215–1221.
(34) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94*, 8897–8909.
(35) Phase; 2.5 ed.; Schrodinger, LLC: New York.
(36) Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. 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.
(37) Dixon, S. L.; Smondyrev, A. M.; Rao, S. N. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem. Biol. Drug Des.* **2006**, *67*, 370–372.
(38) Molecular Operating Environment, 2006.08 ed.; Chemical Computing Group, Inc.: Montreal, Quebec, Canada.
(39) Cavalli, A.; Poluzzi, E.; De Ponti, F.; Recanatini, M. Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K(+) channel blockers. *J. Med. Chem.* **2002**, *45*, 3844–3853.
(40) Brzozowski, A. M.; Pike, A. C.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engstrom, O.; Ohman, L.; Greene, G. L.; Gustafsson, J. A.; Carlquist, M. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **1997**, *389*, 753–758.
(41) Renaud, J.; Bischoff, S. F.; Buhl, T.; Floersheim, P.; Fournier, B.; Halleux, C.; Kallen, J.; Keller, H.; Schlaeppi, J. M.; Stark, W. Estrogen receptor modulators: identification and structure-activity relationships of potent ERalpha-selective tetrahydroisoquinoline ligands. *J. Med. Chem.* **2003**, *46*, 2945–2957.
(42) Kong, E. H.; Heldring, N.; Gustafsson, J. A.; Treuter, E.; Hubbard, R. E.; Pike, A. C. Delineation of a unique protein-protein interaction site on the surface of the estrogen receptor. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3593–3598.
(43) Kim, S.; Wu, J. Y.; Birzin, E. T.; Frisch, K.; Chan, W.; Pai, L. Y.; Yang, Y. T.; Mosley, R. T.; Fitzgerald, P. M.; Sharma, N.; Dahllund, J.; Thorsell, A. G.; DiNinno, F.; Rohrer, S. P.; Schaeffer, J. M.; Hammond, M. L. Estrogen receptor ligands. II. Discovery of benzoxathiins as potent, selective estrogen receptor alpha modulators. *J. Med. Chem.* **2004**, *47*, 2171–2175.
(44) Hummel, C. W.; Geiser, A. G.; Bryant, H. U.; Cohen, I. R.; Dally, R. D.; Fong, K. C.; Frank, S. A.; Hinklin, R.; Jones, S. A.; Lewis, G.; McCann, D. J.; Rudmann, D. G.; Shepherd, T. A.; Tian, H.; Wallace, O. B.; Wang, M.; Wang, Y.; Dodge, J. A. A selective estrogen receptor modulator designed for the treatment of uterine leiomyoma with unique tissue specificity for uterus and ovaries in rats. *J. Med. Chem.* **2005**, *48*, 6772–6775.
(45) Tanenbaum, D. M.; Wang, Y.; Williams, S. P.; Sigler, P. B. Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5998–6003.
(46) Warnmark, A.; Treuter, E.; Gustafsson, J. A.; Hubbard, R. E.; Brzozowski, A. M.; Pike, A. C. Interaction of transcriptional intermediary factor 2 nuclear receptor box peptides with the coactivator binding site of estrogen receptor alpha. *J. Biol. Chem.* **2002**, *277*, 21862–21868.
(47) Mewshaw, R. E.; Edsall, R. J., Jr.; Yang, C.; Manas, E. S.; Xu, Z. B.; Henderson, R. A.; Keith, J. C., Jr; Harris, H. A. ERbeta ligands. 3. Exploiting two binding orientations of the 2-phenylnaphthalene scaffold to achieve ERbeta selectivity. *J. Med. Chem.* **2005**, *48*, 3953–3979.
(48) Manas, E. S.; Xu, Z. B.; Unwalla, R. J.; Somers, W. S. Understanding the selectivity of genistein for human estrogen receptor-beta using X-ray crystallography and computational methods. *Structure* **2004**, *12*, 2197–2207.
(49) Manas, E. S.; Unwalla, R. J.; Xu, Z. B.; Malamas, M. S.; Miller, C. P.; Harris, H. A.; Hsiao, C.; Akopian, T.; Hum, W. T.; Malakian, K.; Wolfrom, S.; Bapat, A.; Bhat, R. A.; Stahl, M. L.; Somers, W. S.; Alvarez, J. C. Structure-based design of estrogen receptor-beta selective ligands. *J. Am. Chem. Soc.* **2004**, *126*, 15106–15119.
(50) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998**, *95*, 927–937.
(51) Norman, B. H.; Dodge, J. A.; Richardson, T. I.; Borromeo, P. S.; Lugar, C. W.; Jones, S. A.; Chen, K.; Wang, Y.; Durst, G. L.; Barr, R. J.; Montrose-Rafizadeh, C.; Osborne, H. E.; Amos, R. M.; Guo, S.; Boodhoo, A.; Krishnan, V. Benzopyrans are selective estrogen receptor beta agonists with novel activity in models of benign prostatic hyperplasia. *J. Med. Chem.* **2006**, *49*, 6155–6157.
(52) Shiau, A. K.; Barstad, D.; Radek, J. T.; Meyers, M. J.; Nettles, K. W.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A.; Agard, D. A.; Greene, G. L. Structural characterization of a subtype-selective ligand reveals a novel mode of estrogen receptor antagonism. *Nat. Struct. Biol.* **2002**, *9*, 359–364.
(53) Stroke, I. L.; Cole, A. G.; Simhadri, S.; Brescia, M. R.; Desai, M.; Zhang, J. J.; Merritt, J. R.; Appell, K. C.; Henderson, I.; Webb, M. L. Identification of CXCR3 receptor agonists in combinatorial small-molecule libraries. *Biochem. Biophys. Res. Commun.* **2006**, *349*, 221–228.
(54) Swaminathan, G. J.; Holloway, D. E.; Colvin, R. A.; Campanella, G. K.; Papageorgiou, A. C.; Luster, A. D.; Acharya, K. R. Crystal structures of oligomeric forms of the IP-10/CXCL10 chemokine. *Structure* **2003**, *11*, 521–532.
(55) Booth, V.; Clark-Lewis, I.; Sykes, B. D. NMR structure of CXCR3 binding chemokine CXCL11 (ITAC). *Protein Sci.* **2004**, *13*, 2022–2028.
(56) Booth, V.; Keizer, D. W.; Kamphuis, M. B.; Clark-Lewis, I.; Sykes, B. D. The CXCR3 binding chemokine IP-10/CXCL10: structure and receptor interactions. *Biochemistry* **2002**, *41*, 10418–10425.
(57) Campanella, G. S.; Lee, E. M.; Sun, J.; Luster, A. D. CXCR3 and heparin binding sites of the chemokine IP-10 (CXCL10). *J. Biol. Chem.* **2003**, *278*, 17066–17074.

CI700395F