FLAME: A Program to Flexibly Align Molecules

Sung Jin Cho* and Yaxiong Sun*

Molecular Structure, Amgen Inc., MS 29-M-B, 1 Amgen Center Drive, Thousand Oaks, California 91320

Received August 1, 2005

Herein, we describe a method to flexibly align molecules (FLAME = FLexibly Align MolEcules). FLAME aligns two molecules by first finding maximum common pharmacophores between them using a genetic algorithm. The resulting alignments are then subjected to simultaneous optimizations of their internal energies and an alignment score. The utility of the method in pairwise alignment, multiple molecule flexible alignment, and database searching was examined. For pairwise alignment, two carboxypeptidase ligands (Protein Data Bank codes 1cbx and 3cpa), two estrogen receptor ligands (1err and 3ert), and two thrombin ligands (1dwe and 1dwd) were used as test sets. Alignments generated by FLAME starting from CONCORD structures compared very well to the X-ray structures (average root-mean-square deviation = 0.36 Å) even without further minimization in the presence of the protein. For multiple flexible alignments, five structurally diverse D3 receptor ligands were used as a test set. The FLAME alignment automatically identified three common pharmacophores: a base, a hydrogen-bond acceptor, and a hydrophobe/aromatic ring. The best alignment was then used to search the MDDR database. The search results were compared to the results using atom pair and Daylight fingerprint similarity. A similar database search comparison was also performed using estrogen receptor modulators. In both cases, hits identified by FLAME were structurally more diverse compared to those from the atom pair and Daylight fingerprint methods.

INTRODUCTION

The binding of a small molecule to its protein target is characterized by its shape and pharmacophore. Mapping of the shape and pharmacophoric requirements for a protein target is critical for optimizing potency as well as properties in drug discovery. In the absence of a ligand−protein complex structure, this information must be indirectly derived from the structure−activity relationship (SAR) information of ligands and their activities. The hypothesis generated can then be used in turn to guide further SAR exploration through medicinal chemistry efforts.

A common computational method for deriving pharmacophoric information is molecular alignment.[1−5] Molecular alignment attempts to find a steric and pharmacophoric overlap of energetically accessible conformations. Pharmacophore types are usually defined as hydrogen-bond donors, hydrogen-bond acceptors, acids, bases, hydrophobes, and sometimes aromatics. Most of the existing methods suffer from the critical limitation that they do not allow simultaneous optimization of both internal conformational energies and the external alignment function, while in reality, the balance between internal energy and alignment is essential. Furthermore, the alignment scoring is often either too simple and requires specific user instruction or too rigid and does not permit easy manual control.

In 2001, a method was described by Labute and Williams, which performed a simultaneous optimization of the alignment score and internal energy, with encouraging results.[5] Here, we describe a program we have developed, based on the same idea of the simultaneous optimization of energies and alignment, but with improvements in two areas. First, the scoring function can be easily adjusted, from completely automatic to completely manual at the atomic level. This allows the inclusion of existing partial SAR information, as well as any user bias. Second, instead of using a random starting alignment, a maximum common pharmacophore method (MCP) coupled with a genetic algorithm (GA) optimization is used to generate initial alignments. This significantly improves the search speed and, when performed in parallel over many processors (such as with a Linux cluster), allows for flexible alignment-based searching for a database of several hundred thousand compounds in less than 1 day.

METHODS

I. Initial Alignment Based On Maximum Common Pharmacophores. The first step in FLAME (FLAME = FLexibly Align MolEcules) is the generation of initial alignments using a genetic algorithm to optimize the maximum common pharmacophores (GA/MCP).[6−8] As with any GA-based optimization method, an encoding scheme and a fitting function are two key components, shown as implemented in FLAME in Figures 1 and 2. By default, one molecule is designated as the probe/reference compound and the other as the target/query compound, with only the query compound subject to optimization. This is the method of choice when the probe compound is that of a known binding mode, such as from an X-ray crystallographic structure. In this instance, the first step is the generation of a specified number of parents for the target compound. To generate unique parents, all noncyclic rotatable bonds in a compound are randomly assigned a discrete value, with the available angle values defined by a modifiable angle increment (5°

* Corresponding authors. E-mail: sung.jin.cho@amgen.com (S.J.C.), yax.sun@amgen.com (Y.S.).

10.1021/ci0503107 CCC: $33.50
Published on Web 10/29/2005

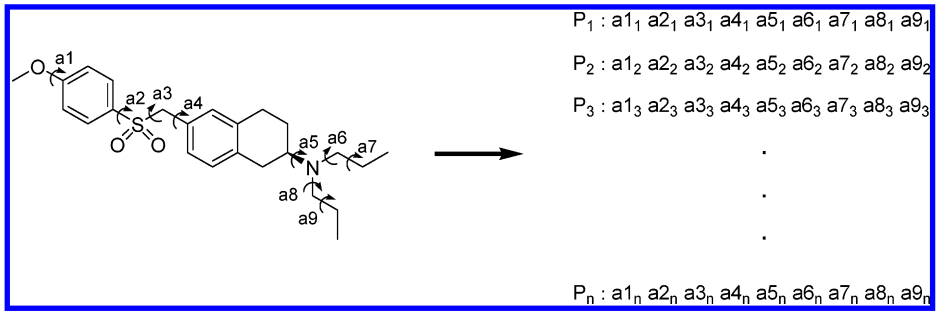

Figure 1. Genetic algorithm encoding scheme. All noncyclic rotatable bonds in a compound are randomly assigned a discrete value ($a1_n$−$a9_n$), with the available angle values defined by a modifiable angle increment (5° default, leading to 360/5 = 72 available discrete values). In this example, there are nine noncyclic rotatable bonds, and the first and last dihedral angles of *n*th parent are denoted as $a1_n$ and $a9_n$, respectively.

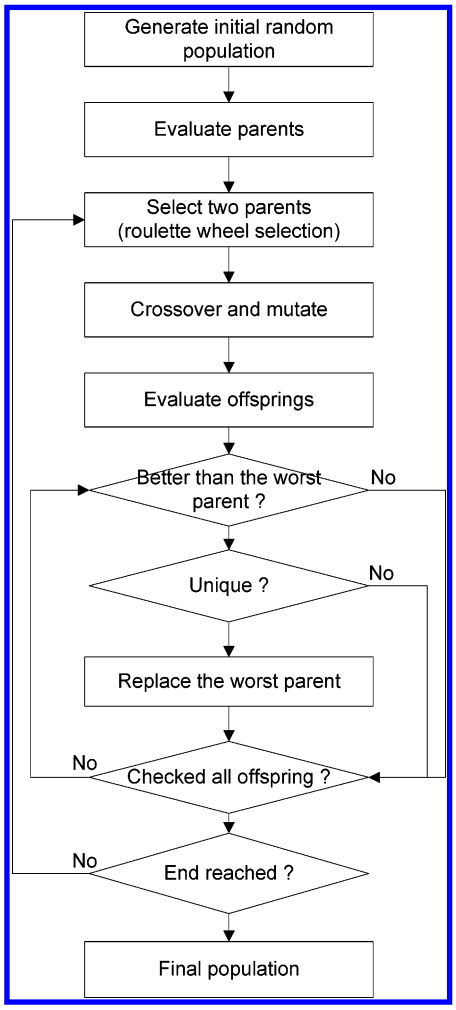

Figure 2. Genetic algorithm steps in FLAME.

default, leading to 360/5 = 72 available discrete values). Once a specified number of parents (i.e., conformations) have been generated, the MCP between the probe compound and each parent of the target compound is evaluated using a clique-detection method.[9−11] The MCP detection algorithm is the extension of the maximum common substructure detection algorithm described by Sheridan and Miller.[10] Briefly, a clique is a set of paired pharmacophores between two molecules A and B. A clique is formed when all intrapharmacophoric distances of the selected pharmacophores in A have matching counterparts in B. The level of matching is controlled by a parameter called matching distance cutoff (MDC); an MDC of 0.1 means a match is formed if the difference between distances in A and B is less than 10% of the distance in A. The clique with the maximum number of pharmacophores is selected as the match for any pairwise comparison, and the fitness score is the number of common pharmacophores. Thus, each parent will have a specific fitness score. In GA evolution, two parents from the initial population are randomly picked using the roulette wheel selection[6] method, with the fitness scores as the sizes of the roulette pockets. Crossover and mutational operations are then performed to generate two new sets of dihedral angles, resulting in the GA-evolved conformations. If any of the resulting offspring are better than any of the worst parents of the entire initial parent population, they replace that parent, always leaving the population size constant. This process is repeated for a specified number of iterations. In the end, the population with the best MCP matches are determined and used to superimpose the probe and the target molecules to generate the initial alignments for further optimization.

II. Optimization of Initial Alignment and Scoring Function. To carry out the simultaneous optimization of internal energy and alignment, a proper scoring function is essential. This function must allow the calculation of derivatives and must have parameters which are intuitive and highly adjustable to provide the fine control that is required. The overall score (F) in FLAME is defined as

$$F = U_{\mathrm{P}} + U_{\mathrm{T}} - TA$$

where U_{P} and U_{T} are the internal energies of probe and target molecules, respectively. A is the alignment score. A temperature factor, T, is used to control the relative weighting of the internal energies and the alignment. It controls how much influence the alignment score has on the internal energies. A high T value can lead to a high internal energy structure. This might be necessary if two molecules are not in a good alignment and need to overcome a high energy barrier. This, however, is not necessary because the GA optimization is performed initially to align them. The rule of thumb is to maintain a similar order of magnitude between the internal energy and the alignment score. By default, T is set to 1.0 and was sufficient for our test cases. The alignment score, A, is defined as

$$A = \sum_{i=1}^{\text{probe}} S_i + \sum_{j=1}^{\text{target}} S_j$$

where the sum is over all individual probe and target atoms. For each probe or target atom, the individual contribution

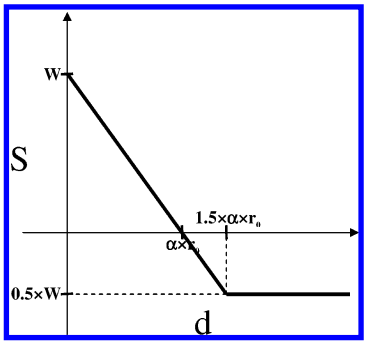

Figure 3. FLAME atomic scoring function *S*.

to the alignment (which applies to both S_i and S_j—only the formula for S_i is shown here for simplicity) is calculated by first finding the closest target or probe atom on the basis of distance, denoted as $j_{\min}$, followed by use of the following formula (Figure 3):

$$S_i = W_{i,j\min}[1 - d_{i,j\min}/(\alpha r_{0j\min})], \quad \text{if } d_{i,j\min} < 1.5\alpha r_{0j\min}$$

$$S_i = -W_{i,j\min}/2, \quad \text{if } d_{i,j\min} > 1.5\alpha r_{0j\min}$$

where the weight term, $W_{i,j\min}$, is the product of chemical matching weight (CMW), probe weight (PW), probe pharmacophore weight (PPW), and target pharmacophore weight (TPW):

$$W_{i,j\min} = \mathrm{CMW}_{i,j\min}\mathrm{PW}_{i,j\min}\mathrm{PPW}_{i,j\min}\mathrm{TPW}_{i,j\min}$$

$d_{i,j\min}$ is the minimum distance to a target atom, α is a scaling parameter to control the slope as a function of the distance, and $r_{0j\min}$ represents the size of atom $j_{\min}$. By default, α and $r_{0j\min}$ are set to 1.0 and 1.5, respectively. Together, the αr_0 term controls how tightly probe and target atoms should be matched. CMW describes which pharmacophores can be matched and how important those matches are. PW allows a user to set the relative importance of individual atoms, while PPW and TPW set the weighting at the pharmacophore level. Through these parameters, the user maintains fine control over the alignment process, if desired and needed. For example, if chemical matching of all H-bond donors in a target compound is important, one could simply increase CMW. However, if such weighting is important only for a particular atom in a probe molecule, adjusting PPW instead proves more useful.

III. Final Selection of Alignments. The entire GA-optimized population is subjected to the second optimization step. The resulting alignment scores of all parents in the population are sorted, and the best scored alignments are saved.

IV. Implementation. FLAME has been implemented as a C++ program using OpenEye toolkits, OEChem and CASE.[12] The MMFF94[13] force field is used to calculate the internal energies of probe and target molecules, and the Broyden−Fletcher−Goldfarb−Shanno (BFGS) algorithm[14] is used to minimize the total score (*F*). Both the MMFF94 force field and the BFGS algorithm are implemented using the CASE toolkit. The graphical user interface (GUI) portion of the program is written in Qt.[15]

Pairwise Alignment with a Static Probe. Pairwise alignment with a static probe molecule is performed by identifying the MCP between the probe and target molecules, and a user-specified number of best matches are then subjected to the minimization with respect to the coordinates of all target atoms. Only the conformation of the target molecule is changed during GA/MCP and energy minimization. The final alignments are sorted by the total score, and the top scoring parents are saved for visual inspection.

Pairwise Alignment with a Flexible Probe. The pairwise alignment described in the previous section works well when the biologically relevant conformation of the probe compound is known. Most of the time, however, this is not the case, and the conformational space of both the probe and target molecules must be searched in order to obtain the best alignment. In FLAME, this is carried out by including both the probe and target molecules in the GA/MCP searches and subsequent minimization.

Multiple Flexible Alignment. In principle, the procedures used for pairwise alignment can be extended to flexible alignment for multiple numbers of molecules. The size of the conformational space which must be searched by the GA algorithm, however, increases exponentially as the number of molecules increases. As a balance between computational efficiency and obtaining reasonable solutions, the following steps are used for the multiple flexible alignment process.

1. CONCORD[16] is used to generate a single conformation for each molecule in the data set. They are combined as the target molecules.

2. For each molecule in the data set, a specified number of diverse low-energy conformations are generated using the FLAME Generate Conformation option. They are combined as the probe molecules.

3. Each conformation of the probe molecules is used as a static probe to align all target molecules. In other words, multiple conformations of all molecules are used as fixed probes, one at a time, to align all other molecules.

4. For each resulting alignment, the total score (energies and alignment scores of all molecules) is minimized with respect to the coordinates of all atoms of all molecules.

5. All alignments are saved and sorted by the total scores for visual inspection.

The FLAME Graphical User Interface. Although FLAME can be run using reasonable default parameters, a large number of parameters can be modified when necessary or desired. For this purpose, a graphical user interface has been created (Figure 4). The interface provides a 3D structure viewer, which can be used to rotate, translate, and scale a molecule, and a FLAME control window. The structure viewer displays atom ID numbers that are needed for the fine control of various probe weighting factors. The control window displays all adjustable parameters with their initial default values. From this window, a user can generate chemical matching and probe weight files, create a FLAME command incorporating all selected parameters, and perform the alignment.

V. Data Sets. *Peptidase Data Set.* The crystal structures of two carboxypeptidase−ligand complexes, with L-benzyl-succinate (1CBX) and glycyl-L-tyrosine (3CPA), were downloaded from the Protein Data Bank (PDB).[17] Proper SYBYL[18] mol2 atom types and bonds were assigned, and both ligands [designated as 1cbx (**1**) and 3cpa (**2**) in Figure 5] were saved in SYBYL mol2 format. No attempts were made to align 1CBX and 3CPA since both complexes were already aligned.

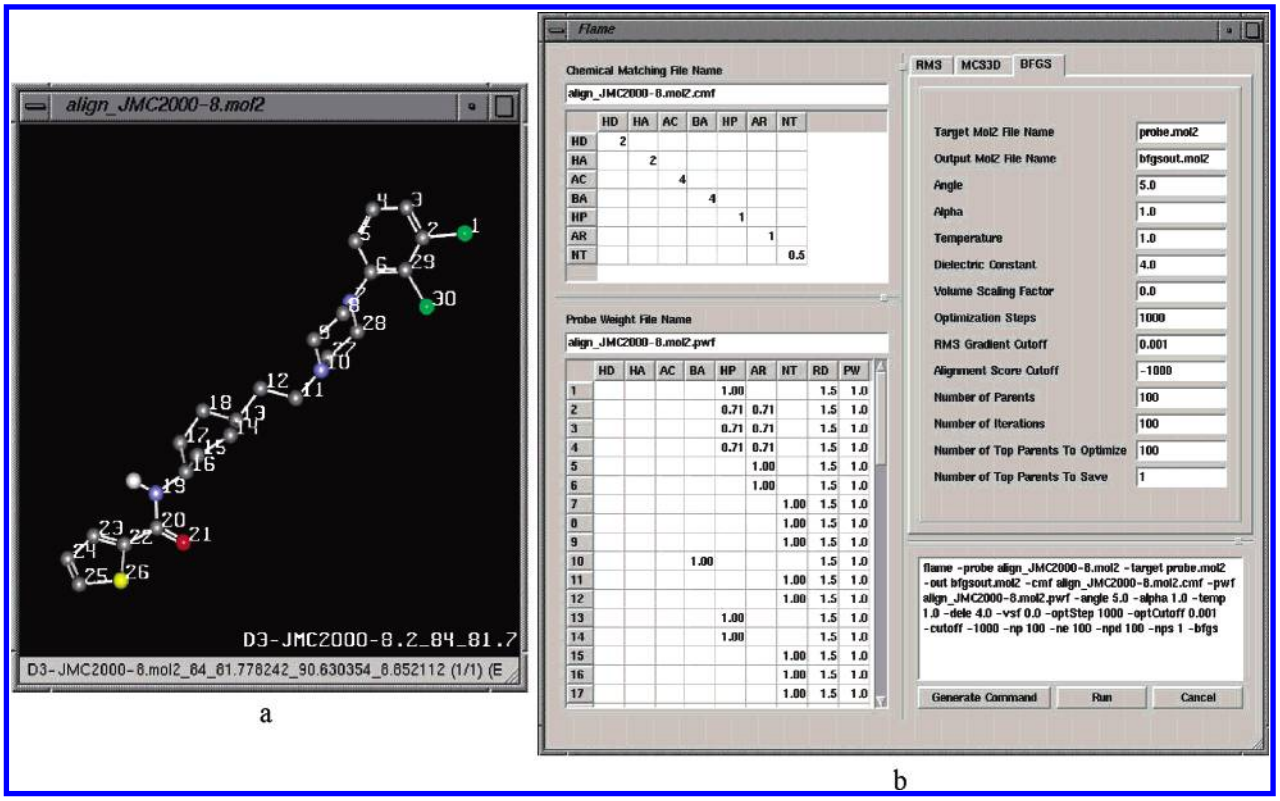

Figure 4. FLAME graphical user interface. (a) FLAME 3D structure viewer; (b) FLAME control window.

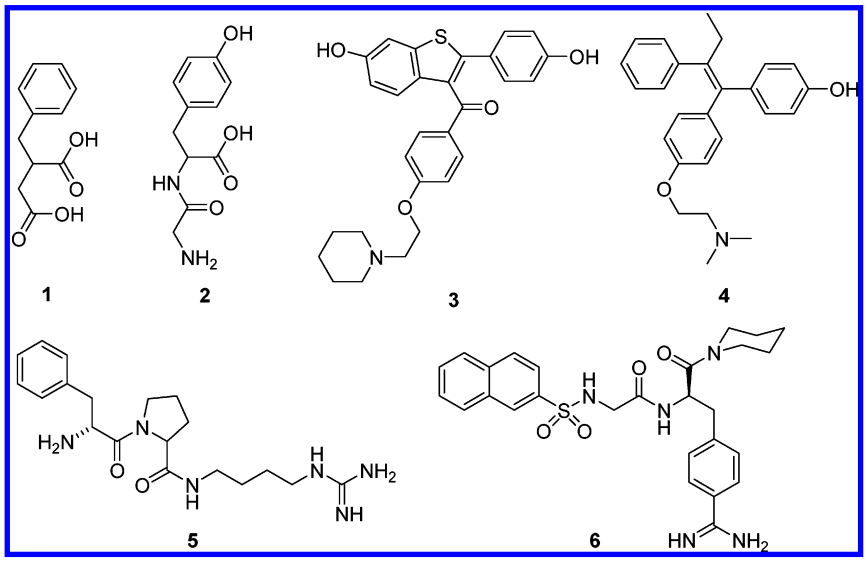

Figure 5. Chemical structures of ligands used in pairwise alignments: **1**, 1cbx; **2**, 3cpa; **3**, 1err; **4**, 3ert; **5**, 1dwe; and **6**, 1dwd.

Estrogen Data Set. The crystal structures of two estrogen receptor complexes, with raloxifene (1ERR) and 4-hydroxytamoxifen (3ERT), were downloaded from the PDB. Two ligands [designated as 1err (**3**) and 3ert (**4**) in Figure 5] were overlaid by aligning the backbone atoms of all 153 residues (from 307 to 459). Proper SYBYL mol2 atom types and bonds were assigned, and both ligands were saved in SYBYL mol2 format.

Thrombin Data Set. The crystal structures of two thrombin ligand complexes, 1DWE and 1DWD, were downloaded from the PDB. Two ligands [designated as 1dwe (**5**) and 1dwd (**6**) are shown in Figure 5] were extracted. Proper SYBYL mol2 atom types and bonds were assigned, and both ligands were saved in SYBYL mol2 file format. No attempts were made to align 1DWE and 1DWD since the two complexes were already aligned.

D3 Data Set. Five representative D3 receptor antagonists (**7**−**11**) were selected from the literature (Figure 6).[19,20] CONCORD was used to generate their three-dimensional structures. No further modification was made to their conformations.

Database Search. To test the FLAME database searching function, molecule **7** (Figure 6) was used to search the MDL Drug Data Report (MDDR)[21] database 122 729 total compounds) for D3 receptor antagonists (activity index = 07703; 254 compounds), and molecule **12** (Figure 7) was used to search MDDR for estrogen receptor modulators (activity index = 41 300; 133 compounds). The compounds in the

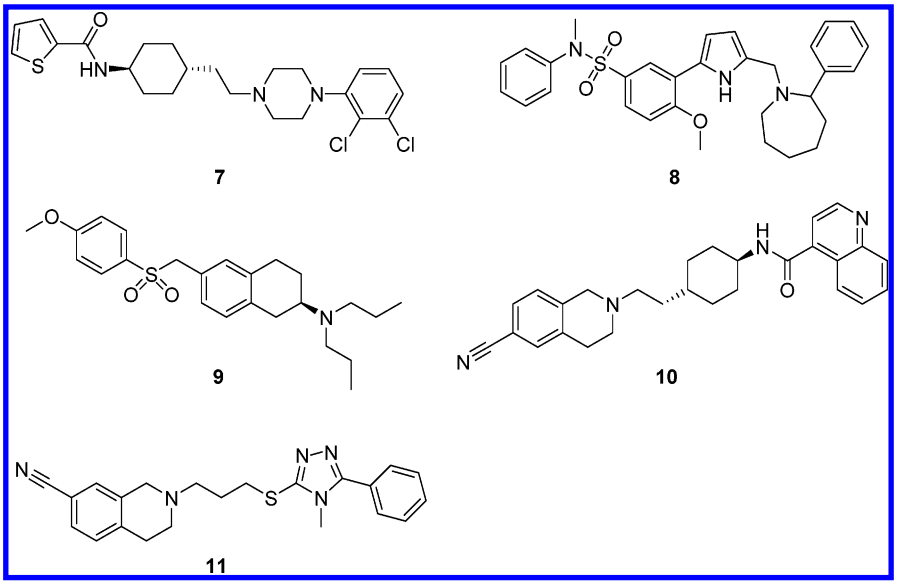

Figure 6. Chemical structures of five D3 receptor antagonists used in multiple flexible alignment.

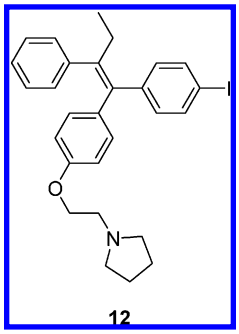

Figure 7. Probe used in the estrogen receptor modulator search.

MDDR database were first filtered by *c* log *P* (≤10), molecular weight (≤700), and the number of rotatable bonds (≤10), followed by conformation generation with CONCORD. Atom pairs[22] and Daylight[23] fingerprints were also generated for the filtered MDDR compounds (89 266 compounds).

RESULTS AND DISCUSSION

Pairwise Alignment. To test the performance of FLAME to align two molecules, we adapted an approach used by Klebe et al.[24] Two ligands bound to the same protein were indirectly aligned by aligning their protein backbones. The ligands were then extracted from the complex as the X-ray alignment, with one of the molecules fixed as the probe and the other allowed to move as the target. The goal is to compare the computationally generated alignment, starting from the CONCORD conformation, to the observed X-ray alignment. For comparison and control, the X-ray alignments were also optimized (minimized) using the same FLAME scoring function. The root-mean-square deviation (RMSD) values were obtained by comparing the X-ray structures of target compounds with the FLAME-minimized structures of target compounds. Three such ligand pairs for peptidase, estrogen, and thrombin were used, and the results were compared with the X-ray alignments (Table 1). For peptidase, the RMSD values of 1.28 and 1.76 Å were obtained starting from the X-ray alignment and CONCORD structure of 3cpa (compared to the X-ray structure of 3cpa), respectively. Figure 8 shows the X-ray structures (Figure 8a) and the alignment obtained after the FLAME minimization of the CONCORD structure of 3cpa (Figure 8b). For estrogen ligands, the RMSD value of 5.10 Å (Table 1) was initially obtained for the alignment generated from the CONCORD structure of 3ert (Figure 9b). In this alignment, the phenol

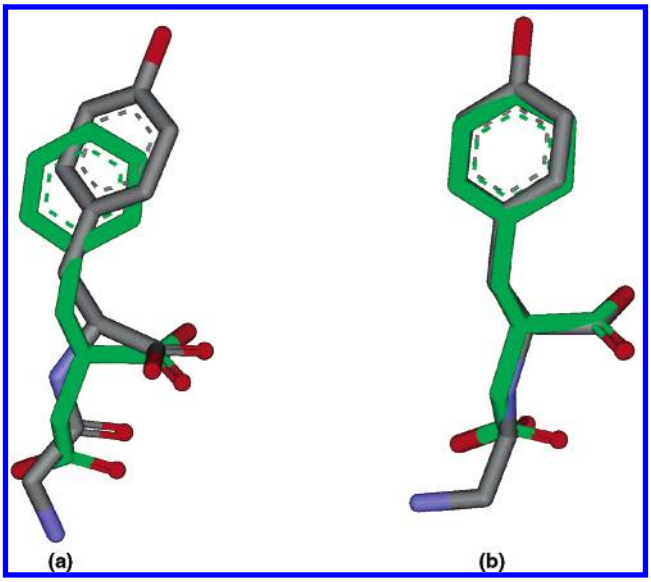

Figure 8. Alignments of peptidase ligands 1cbx (**1**, green) and 3cpa (**2**, gray): (a) X-ray structures of **1** and **2** and (b) the X-ray structure of **1** and the FLAME-generated structure of **2**.

Table 1. RMSD of Pairwise Alignments Compared with X-ray Structures

ligand pairs			CONCORD structure[a]	
probe	target	X-ray structure[b]	before adjustment	after adjustment
1cbx	3cpa	1.28	1.76	
1err	3ert	1.25	5.10	1.68[c]
1dwe	1dwd	1.42	3.38	1.25[d]

[a] RMSD between the X-ray and FLAME-generated structures of a target ligand with the GA optimization. [b] RMSD between the X-ray structure of a target ligand before and after the FLAME minimization without the GA optimization. [c] Probe weight of benzo[*b*]thiophen-6-ol oxygen atom was increased from 1 to 3. [d] Probe weights of three carbon atoms in the proline ring were increased from 1 to 4.

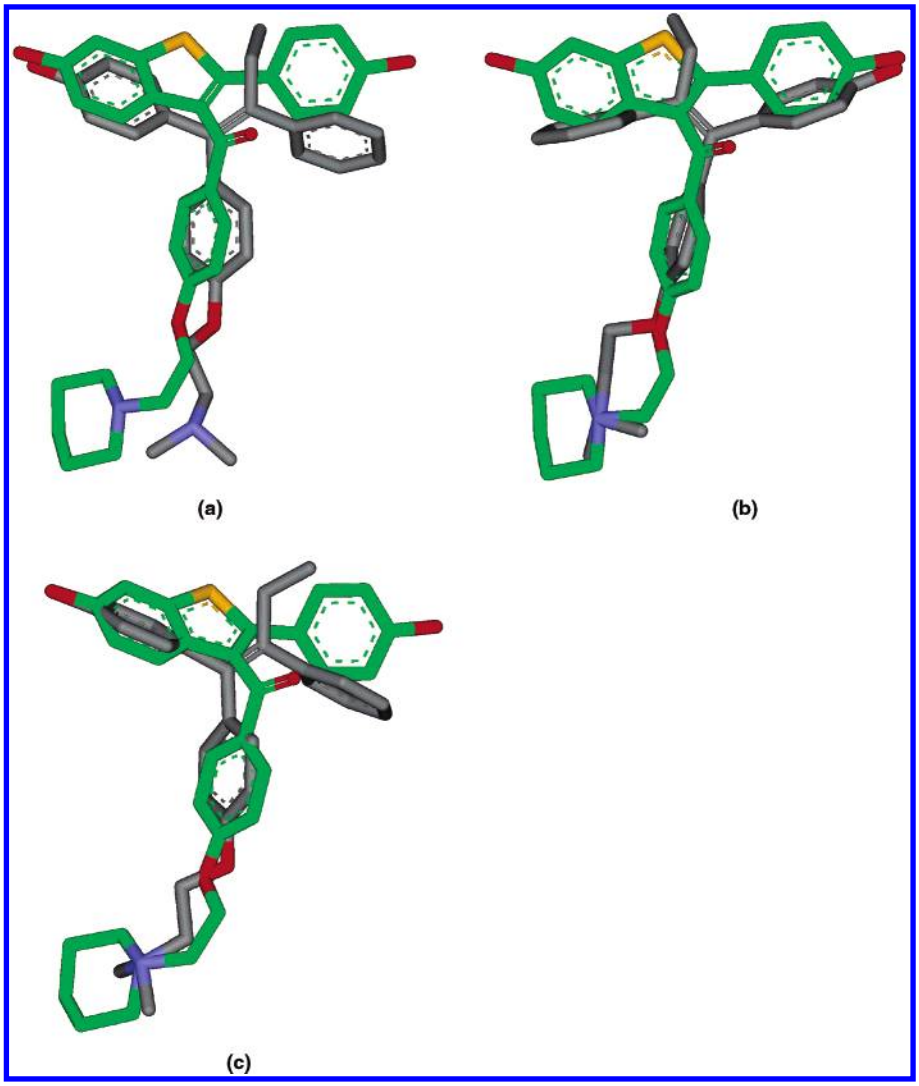

Figure 9. Alignments of estrogen ligands 1err (**3**, green) and 3ert (**4**, gray): (a) X-ray structures of **3** and **4**, (b) the X-ray structure of **3** and the FLAME-generated structure of **4**, and (c) the X-ray structure of **3** and the FLAME-generated structure of **4** after parameter adjustment (see Table 1).

hydroxyl oxygen atom of 3ert (**4**) was matched to the phenol hydroxyl oxygen atom of 1err (**3**), given the almost symmetrical shape of the ligand. In the crystal structure of the 1err/estrogen complex, the hydroxyl of benzo[*b*]thiophen-6-ol of 1err interacts with both GLU353 and ARG394 and is clearly the more important of the two hydroxyls. When a weighting factor of 3 (default: 1) was assigned to the hydroxyl oxygen atom of the benzo[*b*]thiophen-6-ol, FLAME generated an alignment with a RMSD of 1.68 Å (Figure 9c). For the thrombin pair, the alignment generated using the CONCORD structure of 1dwd yielded the RMSD value of 3.38 Å (Table 1). The major discrepancy between RMSD values stems from the inability of the piperidine ring of 1dwd to properly align against the proline-occupied hydrophobic area of 1dwe. To increase the importance of proline matching, the probe weights of three carbon atoms in the proline ring were increased from 1 (default value) to 4. The resulting alignment reduced the RMSD value from 3.38 to 1.25 Å. Figure 10a shows the X-ray structures of 1dwe and 1dwd, and parts b and c of Figure 10 show the predicted alignments before and after FLAME parameter adjustments were made, respectively. Both estrogen and thrombin test cases illustrate that very often there are many possible ways to align two structurally dissimilar ligands. In the absence of receptor structures, it is not easy (and, at times, impossible) to conclude how ligands should align. FLAME's ability to fine-tune the alignment process allows users to derive multiple alignment models. This is especially useful if there is SAR information available to guide and prioritize all possibilities.

Multiple Flexible Alignment. Six published D3 compounds were selected as a test set (Figure 6).[19,20] For each compound, 100 different conformations were generated using FLAME with a minimum RMS cutoff of 0.3, meaning that no two conformers are less than 0.3 Å apart in their RMSD. When the procedure described in the Methods section was used, the multiple alignments were generated and sorted by the total alignment score. Figure 11 shows the final overlay of the best-scoring alignment. Atoms with high individual alignment scores indicate good matching and could potentially be important pharmacophores. Atoms with scores greater that 0.5 were colored green in Figure 11. Three main pharmacophores were identified: a base, a H-bond acceptor, and a hydrophobe/aromatic ring. This is in agreement with the observation made by Varady et al.[25]

Database Searching. D3. Database searching can be based on a known bound conformation from crystal structures of complexes or from a hypothesized active conformation from

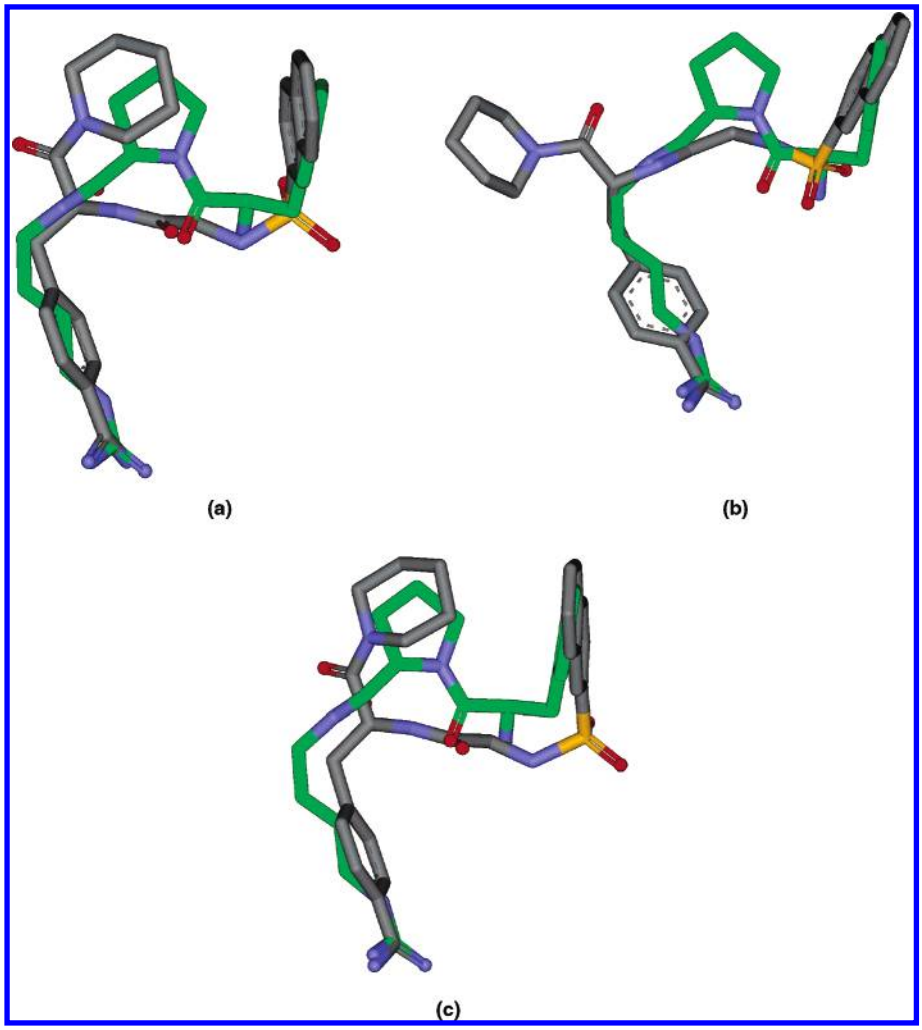

Figure 10. Alignments of thrombin ligands 1dwe (**5**, green) and 1dwd (**6**, gray): (a) X-ray structures of **5** and **6**, (b) the X-ray structure of **5** and the FLAME-generated structure of **6**, and (c) the X-ray structure of **5** and the FLAME-generated structure of **6** after parameter adjustment (see Table 1).

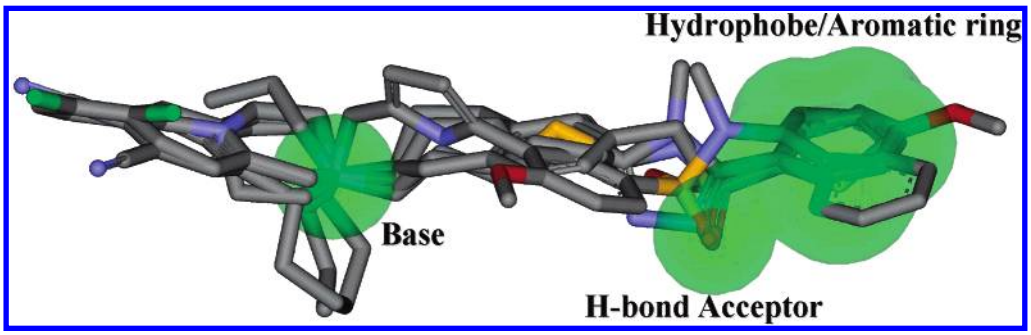

Figure 11. Multiple flexible alignment of five D3 receptor antagonists. Green spheres encompass atoms with an alignment score greater than or equal to 0.5. Three main pharmacophores, a base, a H-bond acceptor, and a hydrophobe/aromatic ring were identified.

a multiple-ligand alignment. In the D3 receptor antagonist example, a reasonable active conformation was derived from multiple flexible alignment and used for subsequent database searching. Specifically, compound **7** was used, since it aligned well to the rest of the four D3 compounds. To expedite the search process, the number of parents and the number of evolution parameters were lowered from 100 to 20 for both. The α value was increased from 1 to 1.5 to relax the stringency and tightness of the fit. The activity index for D3 antagonists in the MDDR database is 07703, and there are 254 compounds with this activity index. When a Linux cluster was used, searching the filtered MDDR database (89 266 compounds) required about 9.5 h using 100 1.0 GHz processors. Figure 12 shows the enrichment plot. The *x* axis represents the percent of the MDDR database searched, and the *y* axis represents the percent of D3 actives found. The top 5% of the database search results contain ~46% of D3 actives (Table 2). In contrast, the same search using atom pairs and Daylight fingerprints identified only ~31% and ~21% of D3 actives, respectively (Table 2).

Estrogen. For our second test case, compound **12** (Figure 7) was randomly selected from MDDR estrogen modulators as a probe for the database search. The search required 9.9 h with 100 1.0 GHz processors, similar to the D3 results.

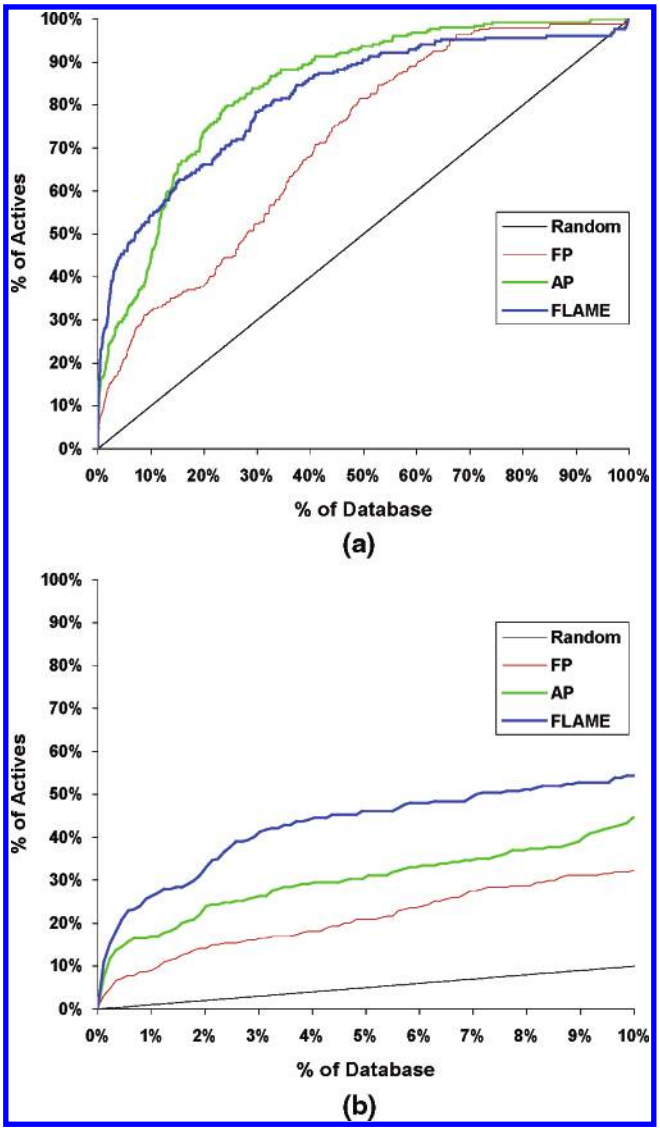

Figure 12. Enrichment plots of FLAME, atom pair (AP), and Daylight fingerprint (FP) searches using a D3 receptor antagonist probe, **7**.

Table 2. Enrichment Results from FLAME and Daylight Fingerprint Searches

	% of actives in top 5% of database search results		
	FLAME	AP[a]	FP[a]
D3	46.1	31.1	20.9
estrogen	88.7	92.5	66.2

[a] AP = atom pair. [b] FP = Daylight fingerprint.

Figure 13 shows the enrichment plot. The top 5% of the database search results contain ~89% of actives (Table 2). The same search using atom pairs and Daylight fingerprints identified ~93% and ~66% of actives, respectively (Table 2). The results reflect the structurally similar nature of actives found in the estrogen test case (Table 3).

Diversity Analysis. Maximum dissimilarity selection[26−28] identifies a set of maximally dissimilar molecules in a database such that any two compounds selected would have similarity below a cutoff value. For D3, there were, in total, 254 active and 84 "diverse" active compounds, as defined using maximum dissimilarity selection with a 0.7 cutoff. In the top 5% of the database results ranked by FLAME, there were 53 "diverse" active compounds. In the top 5% of the database results ranked by atom pairs and Daylight fingerprints, there were 36 and 22 "diverse" active compounds, respectively (Table 3). Similar results were observed for the estrogen hits and are summarized in Table 3. It is clear that FLAME identified more diverse hits.

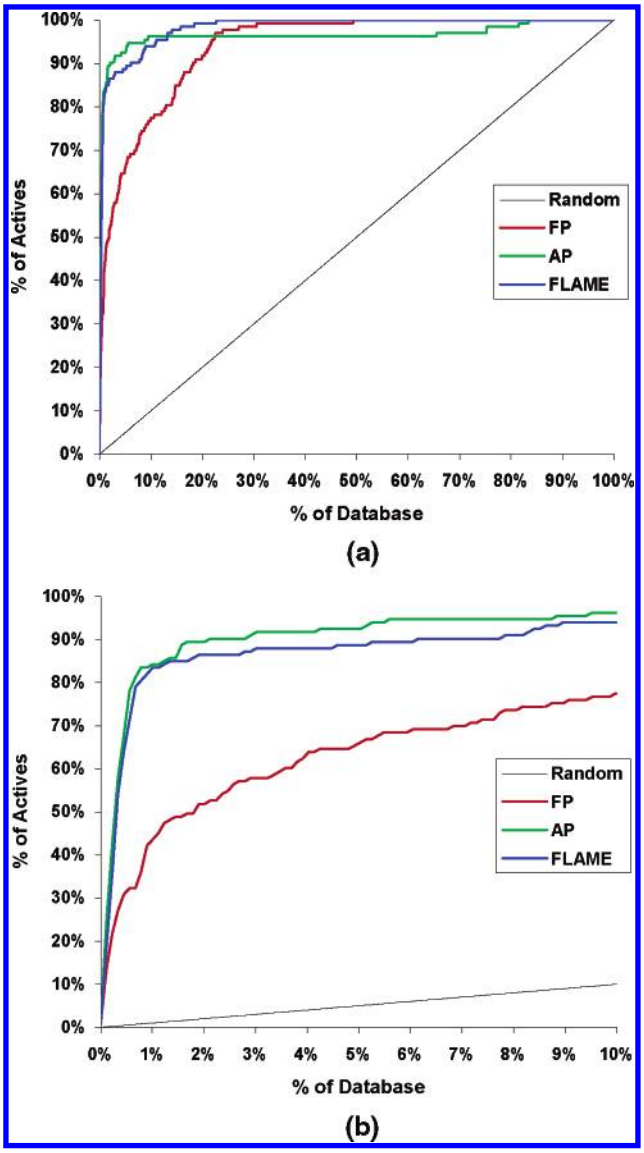

Figure 13. Enrichment plots of FLAME, atom pair (AP), and Daylight fingerprint (FP) searches using an estrogen receptor modulator probe, **12**.

Table 3. Structural Diversity of Actives Found

	total MDDR Actives		FLAME	AP	FP
	all	diverse set[a]	diverse set[a]	diverse set[a]	diverse set[a]
D3	254	84	53	36	22
estrogen	133	38	35	28	23

[a] Diverse set is based on the maximum dissimilarity selection algorithm with a 0.7 similarity cutoff.

CONCLUSION

In the absence of structural information for protein−ligand complexes, a good alignment of molecules can reveal important pharmacophoric information and shed light on

possible receptor−ligand interactions. In this paper, we report a flexible alignment method that performs a simultaneous optimization of internal energies and the alignment score of two or more molecules. Although FLAME is designed to provide reasonable results using default settings, detailed attention was given to FLAME in providing fine control over the alignment process, to allow the incorporation of as much known information and user input as possible. Ligand-based drug design is a trial and error process requiring constant adjustment, and having fine control over the alignment process is a very important feature, as illustrated in the pairwise alignment processes of estrogen and thrombin data sets carried out in this work. In addition, the ability of FLAME to perform multiple flexible alignment and database searching has been investigated. Using a Linux cluster, FLAME was able to perform multiple flexible alignments of a structurally diverse set of five flexible D3 receptor ligands in about 45 min, identifying three common pharmacophores. The use of parallel processing also expedites FLAME database searching (screening 89 260 MDDR compounds in about 9.5 h using 100 1.0 GHz processors). The enrichment observed using FLAME was consistently better than the enrichment observed using Daylight fingerprints and was comparable to the enrichment observed using atom pairs. The top ranked hits selected by FLAME were structurally more diverse than hits selected by the atom pair and Daylight fingerprint approaches.

ACKNOWLEDGMENT

We are grateful for helpful discussions and feedback from other members of the modeling group at Amgen. We especially thank Xiaoyang Xia for providing the D3 data set and Michael Bartberger and Matthew Lee for their critical reading of the manuscript.

REFERENCES AND NOTES

(1) Miller, M. D.; Sheridan, R. P.; Kearsley, S. K. SQ: A Program for Rapidly Producing Pharmacophorically Relevant Molecular Superpositions. *J. Med. Chem.* **1999**, *42*, 1505−1514.

(2) Lemmen, C.; Lengauer, T. Computational Methods for the Structural Alignment of Molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215−232.

(3) Kramer, A.; Horn, H. W.; Rice, J. E. Fast 3D Molecular Superposition and Similarity Search in Databases of Flexible Molecules. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 13−38.

(4) Dror, O.; Shulman-Peleg, A.; Nussinov, R.; Wolfson, H. J. Predicting Molecular Interactions *in silico*: I. A Guide to Pharmacophore Identification and its Applications to Drug Design. *Curr. Med. Chem.* **2004**, *11*, 71−90.

(5) Labute, P.; Williams, C. Flexible Alignment of Small Molecules. *J. Med. Chem.* **2001**, *44*, 1483−1490.

(6) Goldberg, D. E. *Genetic Algorithm in Search, Optimization, and Machine Learning*. Addison-Wesley: Reading, MA, 1989.

(7) Holland, J. H. Genetic Algorithms. *Sci. Am.* **1992**, *267*, 66−72.

(8) Forrest, S. Genetic Algorithms: Principles of Natural Selection Applied to Computation. *Science* **1993**, *261*, 872−878.

(9) Bron, C.; Kerbosch, J. Algorithm 457 - - - Finding All Cliques of an Undirected Graph. *Commun. ACM* **1973**, *16*, 575−577.

(10) Sheridan, R. P.; Miller, M. D. A Method for Visualizing Recurrent Topological Substructures in Sets of Active Molecules. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 915−924.

(11) Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. FLOG: A System to Select Quasi-Flexible Ligands Complementary to a Receptor of Known Three-Dimensional Structure. *J. Comput-Aided. Mol. Des.* **1994**, *8*, 153−174.

(12) The OEChem and CASE toolkits are available from OpenEye Scientific Software, Inc. at http://www.eyesopen.com.

(13) Halgren, T. A. The Merck Molecular Force Field. I. Basis, Form, Scope, Parametrization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(14) Fetcher, R. *Practical Methods of Optimization*, 2nd ed.; Wiley: New York, 1987.

(15) The Qt class library and tools are available from Trolltech Inc. at http://www.trolltech.com.

(16) CONCORD and UNITY are available from Tripos Inc. at http://www.tripos.com.

(17) 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.

(18) *SYBYL*, version 6.9; Tripos Inc.: St. Louis, MO.

(19) Stemp, G.; Ashmeade, T.; Branch, C. L.; Hadley, M. S.; Hunter, A. J.; Johnson, C. N.; Nash, D. J.; Thewlis, K. M.; Vong, A. K. K.; Austin, N. E.; Jeffrey, P.; Avenell, K. Y.; Boyfield, I.; Hagan, J. J.; Middlemiss, D. N.; Reavill, C.; Riley, G. J.; Routledge, C.; Wood, M. Design and Synthesis of *trans*-*N*-[4-[2-(6-Cyano-1,2,3,4-tetrahydroisoquinolin-2-yl)ethyl]cyclohexyl]-4-quinolinecarboxamide (SB-277011): A Potent and Selective Dopamine D_3 Receptor Antagonist with High Oral Bioavailability and CNS Penetration in the Rat. *J. Med. Chem.* **2000**, *43*, 1878−1885.

(20) Macdonald, G. J.; Branch, C. L.; Hadley, M. S.; Johnson, C. N.; Nash, D. J.; Smith, A. B.; Stemp, G.; Thewlis, K. M.; Vong, A. K. K.; Austin, N. E.; Jeffrey, P.; Winborn, K. Y.; Boyfield, I.; Hagan, J. J.; Middlemiss, D. N.; Reavill, C.; Riley, G. J.; Watson, J. M.; Wood, M.; Parker, S. G.; Ashby, C. R., Jr. Design and Synthesis of *trans*-3-(2-(4-((3-(3-(5-Methyl-1,2,4-oxadiazolyl))- phenyl)carboxamido)-cyclohexyl)ethyl)-7-methylsulfonyl-2,3,4,5-tetrahydro-1*H*-3-benzazepine (SB-414796): A Potent and Selective Dopamine D_3 Receptor Antagonist. *J. Med. Chem.* **2003**, *46*, 4952−4964.

(21) *Molecular Design Drug Data Report*, version 2002.2; Molecular Design Ltd.: San Leandro, CA.

(22) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure−Activity Studies: Definition and Applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64−73.

(23) Available from Daylight Chemical Information Systems Inc., 27401 Los Altos, Suite #360, Mission Viejo, CA 92691.

(24) Klebe, G.; Meitzner, 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.

(25) Varady, J.; Wu, X.; Fang, X.; Min, J.; Hu, Z.; Levant, B.; Wang, S. Molecular Modeling of the Three-Dimensional Structure of Dopamine 3 (D_3) Subtype Receptor: Discovery of Novel and Potent D_3 Ligands through a Hybrid Pharmacophore- and Structure-Based Database Searching Approach. *J. Med. Chem.* **2003**, *46*, 4377−4392.

(26) Holliday, J. D.; Ranade, S. S.; Willett, P. A Fast Algorithm for Selecting Sets of Dissimilar Molecules from Large Chemical Databases. *Quant. Struct.-Act. Relat.* **1995**, *14*, 501−506.

(27) Higgs, R. E.; Bemis, K. G.; Watson, I. A.; Wikel, J. H. Experimental Designs for Selecting Molecules from Large Chemical Databases. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 861−870.

(28) Willett, P. Dissimilarity-Based Algorithms for Selecting Structurally Diverse Sets of Compounds. *J. Comput. Biol.* **1999**, *6*, 447−457.

CI0503107