

Novel Approach for Efficient Pharmacophore-Based Virtual Screening: Method and Applications

Oranit Dror,^{†,‡} Dina Schneidman-Duhovny,^{†,‡} Yuval Inbar,^{†,‡} Ruth Nussinov,^{||,§} and Haim J. Wolfson^{*,†}

Blavatnik School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel, Sackler Institute of Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel, and Basic Research Program, SAIC-Frederick, Inc. Center for Cancer Research Nanobiology Program NCI-Frederick, Frederick, Maryland 21702

Received July 21, 2009

Virtual screening is emerging as a productive and cost-effective technology in rational drug design for the identification of novel lead compounds. An important model for virtual screening is the pharmacophore. Pharmacophore is the spatial configuration of essential features that enable a ligand molecule to interact with a specific target receptor. In the absence of a known receptor structure, a pharmacophore can be identified from a set of ligands that have been observed to interact with the target receptor. Here, we present a novel computational method for pharmacophore detection and virtual screening. The pharmacophore detection module is able to (i) align multiple flexible ligands in a deterministic manner without exhaustive enumeration of the conformational space, (ii) detect subsets of input ligands that may bind to different binding sites or have different binding modes, (iii) address cases where the input ligands have different affinities by defining weighted pharmacophores based on the number of ligands that share them, and (iv) automatically select the most appropriate pharmacophore candidates for virtual screening. The algorithm is highly efficient, allowing a fast exploration of the chemical space by virtual screening of huge compound databases. The performance of PharmaGist was successfully evaluated on a commonly used data set of G-Protein Coupled Receptor alpha1A. Additionally, a large-scale evaluation using the DUD (directory of useful decoys) data set was performed. DUD contains 2950 active ligands for 40 different receptors, with 36 decoy compounds for each active ligand. PharmaGist enrichment rates are comparable with other state-of-the-art tools for virtual screening.

INTRODUCTION

Virtual Screening (VS) is a computational approach to drug discovery that successfully complements High Throughput Screening (HTS) for hit detection.¹ The objective is to use a computational approach for rapid cost-effective evaluation of large virtual databases of chemical compounds in order to identify a set of candidates to be synthesized and examined experimentally for their biological activity. Unlike HTS, virtual screening is not based on a brute-force search but on some starting information on the receptor under inspection or on its active ligands.

Virtual screening methods can be divided into two broad categories: structure-based and ligand-based. In case the three-dimensional (3D) structure of the target receptor or of its binding site is available, docking is a highly effective technique for virtual screening.^{2,3} In the absence of structural information on the receptor, virtual screening approaches are mainly based on structural similarity between known and potential active ligands. The rationale is that molecules that share some structural similarity may have a similar activity. Some methods use a known active ligand as a query to

extract structurally similar compounds from large databases.⁴ Among these are the FlexS^{5,6} and fFlash⁷ methods. These methods perform a 3D structural alignment between a pair of compounds, a query ligand assumed to be rigid and each database compound treated as flexible.

When a set of active ligands is available, it is possible to compute their shared pharmacophore. A *pharmacophore* is defined as the 3D arrangement of features that is crucial for a ligand molecule in order to interact with a target receptor in a specific binding site. Once identified, a pharmacophore can serve as an important model for virtual screening, especially in the case where the 3D structure of the receptor is unknown and docking techniques are not applicable. The strength of pharmacophore-based screening compared to other ligand similarity screening approaches lies in the ability to detect a diverse set of putative active compounds with totally different chemical scaffolds. This increases the chances that some of the detected compounds will pass all the stages of drug development. Besides screening, pharmacophore is a powerful model also in other applications of drug development, like *de novo* design, lead optimization, ADME/Tox studies, and Chemogenomics.^{8,9}

Many *in silico* methods are available for identifying pharmacophore models from a set of ligands that have been observed to interact with the same receptor.^{10–12} Generally, these methods search for the largest 3D pattern of features

* Corresponding author e-mail: wolfson@tau.ac.il.

[†] Blavatnik School of Computer Science, Raymond, Tel Aviv University.

[‡] These authors contributed equally to the work.

^{||} Beverly Sackler Faculty of Exact Sciences, Tel Aviv University.

[§] NCI-Frederick.

responsible for binding that is shared by all or most input ligands. From the computational standpoint, this task is challenging with respect to both the number of input ligands and their flexibility. The various approaches mainly differ in three aspects: (i) the chosen feature descriptors and structure representation, (ii) the technique for addressing the flexibility of the ligands, and (iii) the definition of the searched common pattern and the algorithm employed for identifying it.^{10,13}

All ligand-based methods for pharmacophore detection represent the input ligands by their features. The selection of feature descriptions is mainly based on the desired level of resolution. At the highest level of resolution, a feature is the 3D position of an atom associated with the atom type.^{14–16} At a coarser resolution level, adjacent atoms in space are grouped into topological features, such as phenyl ring and carbonyl group.¹⁷ Finally, at the lowest level of resolution, atoms are grouped into functional features based on physicochemical properties that are important for ligand–receptor binding, like charge, aromaticity, hydrogen bonding, and hydrophobicity.^{18–21} Using the feature descriptors, the structures of the input ligands are represented mainly as 3D point sets,^{16,22} distance matrices,^{23,24} graphs,^{24,25} or trees.²⁶

Druglike molecules may adopt many possible conformations. The specific conformations that the input ligands adopt in the active site of the receptor are usually unknown. Additionally, they cannot be assumed to be the ones with the lowest energy.²⁷ Therefore, all the feasible conformations of the ligands should be considered during the search for the common pharmacophore. Most methods perform the conformational search in a separate initial stage. Examples for such methods include RAPID,¹⁶ MPHIL,¹⁴ DISCO,^{28,29} PHASE,³⁰ Catalyst/HipHop,^{18–20,31} Catalyst/HypoGen,^{18,21,31} and others.^{32–34} These methods generate a discrete set of conformations for each ligand with the goal of covering its whole conformational space. The main limitation of this approach is that many conformations (about 100) might be required for each ligand.²⁷ An alternative strategy is to incorporate the conformational search within the pattern detection process. This way the search space is not restricted to a predefined discrete number of conformations. However, currently the methods that adopt this approach are non-deterministic, predominantly based on a random search.^{15,17,35,36} Among them are SCAMPI,¹⁷ GAMMA,^{15,37,38} GASP,^{35,39} GALAHAD,^{40,41} and the MOGA-based method of Cottrell et al.³⁶ Another approach for treating the flexibility of the ligands within the pattern detection process is employed by the MTree method.²⁶ This method aligns feature-trees derived from the 2D structures of the input ligands.

Given a set of ligands, there are different ways to define the common pattern that we are looking for. The traditional approach is to search for the *Maximal Common Substructure (MCS)*, namely the largest set of features shared by all input ligands.^{24,42} However, since some features are less frequent than others, the largest common pattern can be sometimes misleading as a pharmacophore. A better approach is to assign weights to the features based on their infrequency of appearance and to look for the highest weighted sum of common features. Another drawback of the MCS approach is that it is based on the assumption that there is a single pharmacophore and all ligands possess all its features. This

is not always the case. The input ligands can bind to different binding sites, they may have different binding affinities, and in a noisy input some of them might be outliers. Several methods overcome this limitation by relaxing the requirement that all ligands should possess all the pharmacophoric features. In MPHIL, a pharmacophore model is the smallest set of features with which each ligand has at least a predefined number of features in common. In Catalyst/HipHop the user can define how many ligands must map completely or partially to the pharmacophore hypothesis.¹¹ Besides adopting different approaches for defining the pharmacophore pattern, the methods also differ in the algorithm employed for detecting it. The most common algorithms are clique-detection,^{14,28,34,43} Gibbs sampling,³² exhaustive search,^{17,18} and genetic algorithms.^{14,15,35,36}

Here, we present an application of *PharmaGist*, a new ligand-based method for pharmacophore detection,^{44,45} for virtual screening. The main innovation of the method lies in the deterministic consideration of the flexibility of the input ligands in the pattern detection process. PharmaGist is robust, capable of handling a diverse and noisy input. The input ligands can bind to different binding sites, they may have different binding affinities, and some of them might be outliers. In this work we address the problem of selecting the most appropriate pharmacophore candidate for virtual screening. In addition, to address cases where the input ligands have different affinities, we suggest a scheme for weighting of pharmacophore candidates. In a *weighted pharmacophore* each feature is weighted based on the number of ligands that possess it. This means that some of the pharmacophoric features do not necessarily appear in all active ligands.

PharmaGist was tested successfully on the FlexS data set.⁴⁵ This benchmark data set consists of 74 ligands that are classified into 12 cases according to the protein receptor they bind to.⁶ The results, which were achieved using a set of standard default parameters, were consistent with reference pharmacophores derived from the bound ligand–receptor complexes. Here, we examine the applicability of PharmaGist for virtual screening. We estimate its performance on a commonly used data set of G-Protein Coupled Receptor alpha1A. Additionally, we perform a large-scale evaluation using the DUD (directory of useful decoys) data set.⁴⁶ This data set contains 2950 active ligands and 95316 decoys for 40 different receptors. The results show the merit of the method in comparison to other methods for pharmacophore-based screening, Catalyst^{18–20,31} and MTree,^{26,47} as well as to receptor-based screening with the DOCK method.^{48,49}

METHOD

PharmaGist consists of two modules. The first module is for detecting pharmacophore candidates from a set of input ligands. The other module is for pharmacophore-based screening.

Pharmacophore Detection. The input is a collection of ligands. The goal is to detect pharmacophore candidates, namely the highest scoring 3D configurations of features responsible for binding that are common to a significant number of input ligands. We give an outline of the algorithm (Figure 1).

The input ligands can be highly flexible. They might have many possible conformations distributed over the entire range

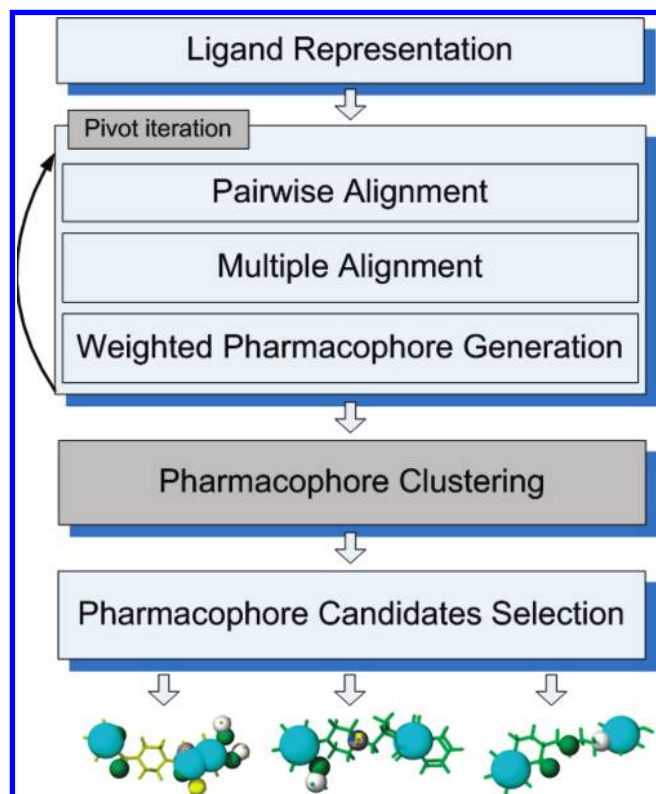


Figure 1. Pharmacophore detection flow.

of reachable energy and conformational space.²⁷ A common approach to deal with the flexibility of the ligands is to generate in an initial stage a large number of possible conformations for each one of them (about 100 for a sufficient coverage²⁷). Instead, our approach assumes that one of the ligands, the *pivot*, is given in its binding conformation and the flexibility of the other ligands is fully and explicitly considered within the search for the common pharmacophore. The advantage of this approach lies in the fact that in the worst case, when there are no clues on the binding conformation of any of the ligands, a set of conformations for only one of them (the *pivot*) needs to be computed. The *pivot* can be selected as the most rigid molecule or as the molecule with the highest affinity toward the target receptor. Unless a *pivot* is specified by the user, the algorithm iteratively tries each input ligand as a *pivot*.

The algorithm identifies pharmacophores by computing multiple flexible alignments between the input ligands. The multiple alignments are generated by combining pairwise alignments between the *pivot* and the other ligands. However, the optimal multiple alignment does not necessarily consist of the optimal pairwise alignments. Therefore, to ensure that we detect the optimal solution, a large number of pairwise alignments are generated and used by the multiple alignment stage. The resulting multiple alignments reveal spatial arrangements of consensus features shared by different subsets of input ligands. The highest-scoring ones are potential pharmacophores.

Here we briefly describe PharmaGist ligand representation as well as pairwise and multiple alignment methods. A more detailed technical description accompanied by a complexity analysis can be found elsewhere.^{44,45} Next, we present our novel contribution in generation of weighted pharmacophores and automated selection of the most appropriate pharmacophores for the virtual screening stage.

Ligand Representation. In this stage, each ligand is partitioned into rigid groups and is assigned with a set of physicochemical features.

A *rigid group* of a ligand is defined as a set of atoms between rotatable bonds, where a bond is considered *rotatable* if it is not (i) double, (ii) a bond on a ring, (iii) a bond connecting a single atom, or (iv) a peptide bond (a bond connecting two amino acids and is formed between the C atom in the carboxyl group of the first amino acid and the N atom in the amino group of the second one).

A *feature* is defined as a set of atoms in the same rigid group with a physicochemical property important for binding. Namely, a feature is (i-ii) a hydrogen-bond acceptor/donor atom; (iii-iv) an anion/cation atom; (v) a set of atoms of an aromatic ring; or (vi) a pair of adjacent hydrophobic atoms (Table 1). In addition, users can define other feature types of their own.

Since a druglike molecule can be represented by a graph, linear time techniques from graph theory are applied to find the rigid groups and the aromatic rings of each ligand.

Pairwise Alignment. The input for this stage is a pair of ligands, a *pivot* ligand and a single *target* ligand. The *pivot* is considered to be rigid, whereas the *target* ligand is treated as flexible. The goal is to generate the *K* highest-scoring pairwise alignments, where a *pairwise alignment* is a superposition of a feasible conformation of the *target* ligand onto the *pivot*. We score a pairwise alignment based on the *pivot* features matched by the superimposed conformation of the *target* ligand. Two features, one from the *pivot* and one from a conformation of the *target* ligand, can be *matched* if they are of the same type and their distance is below a predefined threshold. By default, the distance threshold is defined as 1.8 Å for centers of aromatic rings and 1.4 Å for the centers of the other feature types. Overall, the *score* of a pairwise alignment with a subset $\{f_i\}$ of matched *pivot* features is $\sum_i s(f_i)$, where $s(f_i)$ is the score for matching the *pivot* feature f_i (by default, it is 0.3 for hydrophobic groups, 3.0 for aromatic rings, and 1 for the rest).

The task is tackled by a two-tier procedure. First, for each rigid group of the *target* ligand, a set of rigid transformations for superimposing it on the *pivot* is computed. Then, the resulting candidate new poses of the *target* rigid groups are assembled to form the *K* highest-scoring pairwise alignments. The procedure is highly efficient. A hybrid technique of Pose-Clustering⁵⁰ and Geometric Hashing⁵¹ is employed for generating candidate new poses for the *target* rigid groups. Their best assemblies are computed by dynamic programming.

Multiple Alignment. The input is a *pivot* ligand, a collection of *M* *target* ligands and a large number (*K*) of pairwise alignments between the *pivot* and each *target* ligand. The goal is to compute the highest scoring multiple alignments. A *multiple alignment* is defined by a subset of *target* ligands and exactly one pairwise alignment for each one of them. The score of a multiple alignment is a function of the number of participating *target* ligands and the consensus subset of *pivot* features matched by them. Our aim is to find multiple alignments with significant subsets of *pivot* features matched by as many *target* ligands as possible. However, there is an inherent trade-off between the two requirements. The larger the number of participating ligands (*m*), the smaller is their consensus subset of *pivot* features ($\{f_i\}$). We have found that this trade-off is well balanced by defining

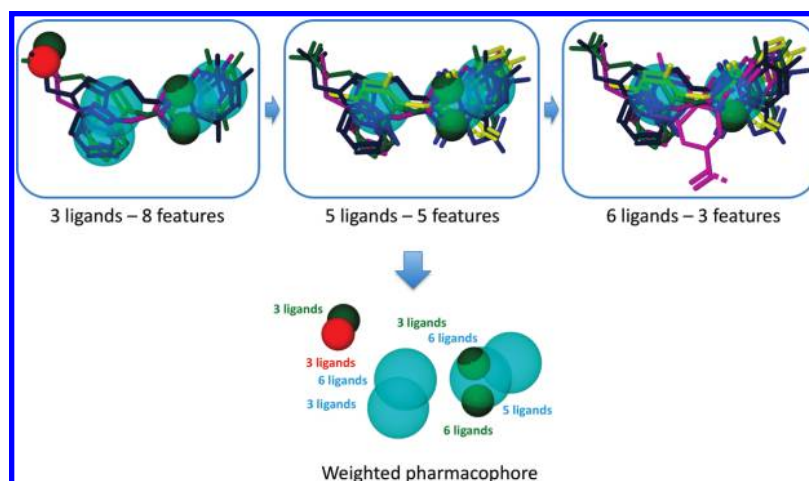


Figure 2. The figure presents a multiple alignment ensemble of aldose reductase (ALR2) ligands as a path with three nodes. The parent multiple alignment (left node) consists of eight features shared by three ligands: four aromatic rings (cyan), three acceptors (green) and one charged group (red). Its direct child node stands for a multiple alignment with two additional molecules and with only five common features: three aromatic rings (cyan) and two acceptors (green). Finally, the last child node (right) corresponds to a multiple alignment with an additional molecule (six in total) and with only three common features: two aromatic rings and one acceptor. The derived weighted pharmacophore is shown at the bottom.

the *score* of a multiple alignment as $(m + 1)^{1/2} \cdot \text{feature_score}$, where *feature_score* is $\sum_i s(f_i)$.

Given K pairwise alignments for each of the M target ligands, there is an exponential number of $(K + 1)^M$ combinations for a multiple alignment. An enumeration over all these combinations is impractical since K is a large number (1500 by default) and M is unbounded. Instead, we apply an exhaustive search on the possible subsets of matched pivot features. If n is the number of pivot features, then there are 2^n subsets to enumerate in the worst case. This enumeration is practical since the number of atoms, and thus the number of features (n), in a typical druglike molecule is small. The algorithm works in a bottom-up manner. It starts with subsets of one matched pivot feature and incrementally extends them until no larger subsets are matched. The implementation is efficient using bitwise representation.

Weighted Pharmacophore Generation. The input is a list of multiple alignments between the pivot and different subsets of target ligands. In principle, pharmacophore models can be derived from the highest scoring multiple alignments. However, due to the trade-off between the number of ligands participating in a multiple alignment and its consensus set of features, the input multiple alignments are not disjoint, and some of them can be combined to gain more information on the candidate pharmacophores. Specifically, we are interested in combining a *parent* multiple alignment with other *children* multiple alignments for which the consensus is partial to the consensus of the parent due to additional target ligands. Figure 2 presents a parent multiple alignment with eight features shared by three ligands and children multiple alignments with less common features compared to their parent but with more input molecules.

The pharmacophore derived from such an ensemble of multiple alignments is called a *weighted pharmacophore* (Figure 2). In a weighted pharmacophore, the features $\{f_i\}$ are the consensus pivot features of the parent multiple alignment assigned with a weight $w(f_i)$ in addition to the feature score $s(f_i)$. The assigned weight $w(f_i)$ is a function of the number of molecules $m(f_i)$ in the ensemble that match f_i , that is $w(f_i) = (m(f_i))^{1/2}$. The overall *score* of a weighted pharmacophore is $\sum_i w(f_i) \cdot s(f_i)$.

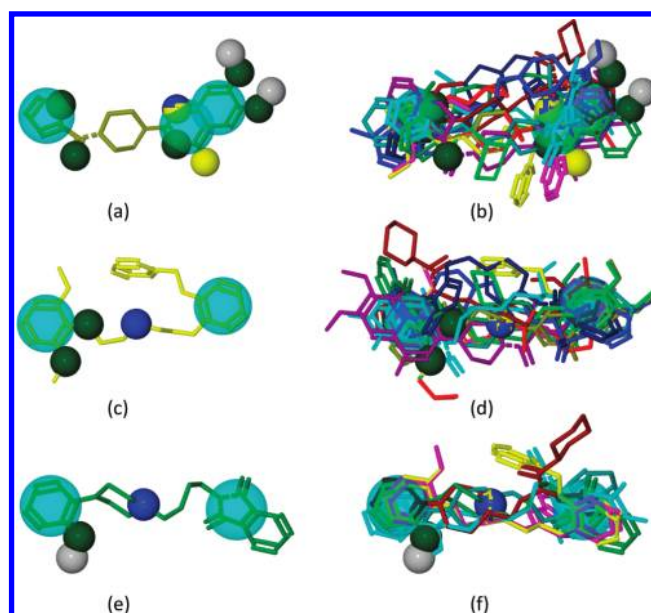


Figure 3. AlphaA1 class I pharmacophore candidates. Three pharmacophore candidates were computed for the 12 alpha1A class I ligands. (a) The first selected pharmacophore candidate consists of 13 features: three aromatic rings (cyan), two hydrophobic groups (gray), two donors (yellow), five acceptors (green), and one charged group (blue). Note that one donor and a charge group overlap. The pivot molecule is colored in yellow. (b) The multiple alignment that corresponds to the pharmacophore candidate depicted in (a). (c–d) The second pharmacophore candidate and its corresponding multiple alignment. This pharmacophore model consists of two aromatic rings (cyan), one donor (yellow), two acceptor groups (green), and one charged group (blue). (e–f) The third pharmacophore candidate and its corresponding multiple alignment. The pharmacophore model contains two aromatic rings (cyan), a hydrophobic group (gray), a donor (yellow), an acceptor (green), and a charged group (blue). Note that two aromatic rings and one acceptor (close to the left aromatic ring) are shared by all three pharmacophore models and to all 12 ligands.

The weighted pharmacophores are computed as follows. A directed acyclic graph (DAG) is constructed. The nodes of the graph stand for the input multiple alignments. A node u is connected to node v by a direct edge if u is a parent of v . Additionally, the nodes of the graph are weighted, where

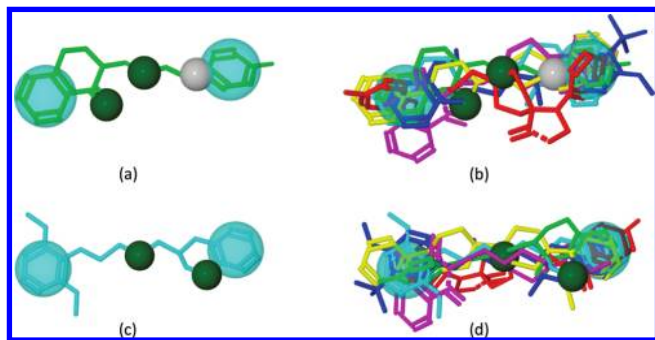


Figure 4. AlphaA1 class II pharmacophore candidates. Two pharmacophore candidates were computed for the six alpha1A class II ligands. (a) The first pharmacophore candidate consists of five features: two aromatic rings (cyan), two acceptors (green), and one hydrophobic feature (gray). (b) The multiple alignment that corresponds to the pharmacophore candidate depicted in (a). (c–d) The second pharmacophore candidate and its corresponding multiple alignment. This pharmacophore model contains two aromatic rings (cyan) and two acceptor groups (green).

the weight of a node is the score of the corresponding multiple alignment. Paths in this graph stand for parent-children ensembles of multiple alignments, and our aim is to find the highest-scoring ones. The highest-scoring paths are computed by dynamic programming.

The result of this stage is a set of *multiple alignment ensembles*, where each one of them is assigned with the corresponding weighted pharmacophore. Figure 2 presents a multiple alignment ensemble as a path with three nodes. The parent node stands for a multiple alignment between three molecules, which share eight common features. Its direct child node stands for a multiple alignment with two additional molecules and with only five common features. Finally, the last child node corresponds to a multiple alignment with an additional molecule (six in total) sharing only three common features.

Pharmacophore Clustering. The three stages described above (Pairwise Alignment, Multiple Alignment, and Weighted Pharmacophore Generation) are repeated for each selected pivot. Different pivots may lead to the same or similar weighted pharmacophores. Therefore, after iterating over all the possible pivots, we cluster weighted pharmacophores with a similar 3D pattern of features.

Pharmacophore Candidates Selection. The algorithm produces a large set of weighted pharmacophores, and the task of selecting the most appropriate ones for virtual screening is not trivial. The input molecules may bind different binding sites or belong to different classes of molecules. Based on this observation, we have developed a strategy for selecting a small set of weighted pharmacophores that cover as much as possible input molecules. The weighted pharmacophores are selected iteratively as follows. The first selected weighted pharmacophore is the top-scoring one. The ligands of the parent multiple alignment of this weighted pharmacophore, and these only, are marked as covered. Next, we iterate over the remaining weighted pharmacophores and select the next best scoring weighted pharmacophore that has no more than two molecules in common with the set of already covered ones. The process is repeated until all molecules are covered or no more weighted pharmacophores can be selected. This results in a ranked list of weighted pharmacophores associated with different classes of binding ligands.

Virtual Screening. The input to this module is a cover set of weighted pharmacophores and a 3D database (DB) of druglike compounds. The goal is to rank the database compounds by their similarity to the weighted pharmacophores. The similarity is computed by a pairwise alignment between the DB molecule and the pivot molecule of each pharmacophore model. In addition to the pivot, it is possible to align the DB molecule with other molecules in the pharmacophore model. This can improve the alignment quality but the running time increases.

The pairwise alignments are scored by $\sum_i w(f_i) \cdot s(f_i)$, where $w(f_i)$ and $s(f_i)$ are the weight and score of the matched feature in the weighted pharmacophore, respectively. In the case of several pharmacophore models, the maximal score can be selected.

This stage employs the same pairwise alignment algorithm as in the pharmacophore detection module. The difference is that the number of pairwise alignments (K) generated in this stage is much lower compared to the pharmacophore detection module, since here we are only interested in the highest-scoring alignment. This leads to a highly efficient pairwise alignment, which is suitable for screening large compound databases with explicit consideration of full compound flexibility.

RESULTS

The capability of PharmaGist to identify correct pharmacophores was successfully tested on the FlexS data set.⁴⁵ This benchmark data set consists of 74 ligands that are classified into 12 cases according to the protein receptor they bind to.⁶ Detailed description of the results is available elsewhere.⁴⁵ Generally, the results, which were achieved using a set of standard default parameters, were consistent with reference pharmacophores derived from the bound ligand–receptor complexes. Below, we provide a description of additional tests for examining the applicability of PharmaGist not only for pharmacophore detection but also mainly for virtual screening.

A key criterion for measuring the success of screening is the enrichment of annotated ligands among top ranking compounds. The higher the percentage of annotated ligands among the top ranking compounds, the better the performance of the screening. The enrichment criterion is evaluated by a numerical factor defined as

$$EF_{subset} = \frac{\{ligands_{subset}/compounds_{subset}\}}{\{ligands_{total}/compounds_{total}\}}$$

where $ligands_{total}$ and $compounds_{total}$ are the total number of ligands and compounds in the database, respectively, and $ligands_{subset}$ and $compounds_{subset}$ are the number of ligands and compounds found in the top-ranking subset of compounds, respectively. This enrichment factor reflects the capability of a screening application to detect active ligands (true positives) compared to random selection. Thus, its value should always be greater than 1, and the higher it is, the better the enrichment performance of the virtual screening.

Running times of the pharmacophore detection stage range from several seconds to several minutes.⁴⁵ In the screening application the running time of the method is 0.5–1.0 s

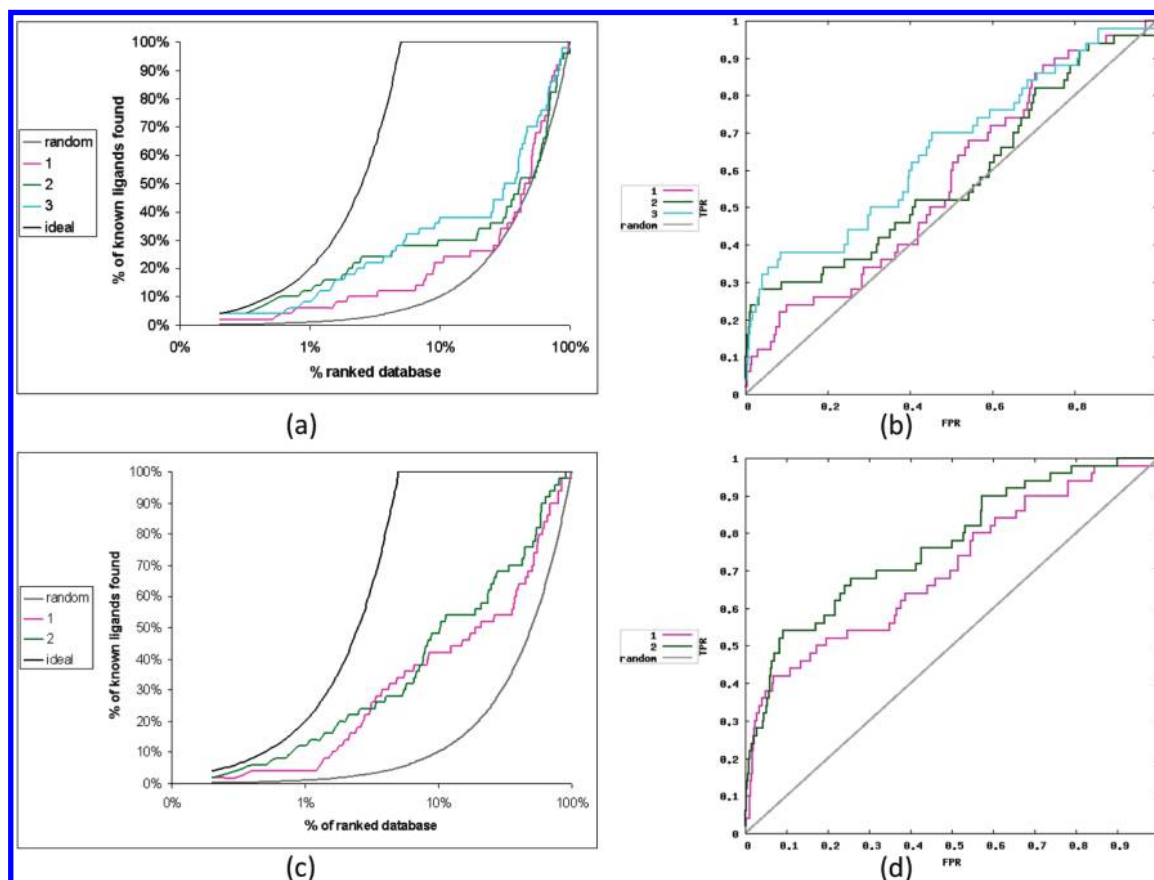


Figure 5. Enrichment plots and ROC curves for the GPCR data set. In the enrichment plots the percentage of ranked database (x-axis) is plotted against the percentage of known ligands (y-axis) found at any given percentage of ranked database. The gray line is the enrichment plot for a random predictor, and the black line is the plot of an ideal predictor. The ROC curves plot the fraction of true positives (TPR = true positive rate) vs the fraction of false positives (FPR = false positive rate). (a–b) Enrichment plots and ROC curves for the three pharmacophore candidates of PharmaGist for class I compounds (the magenta, green, and cyan lines). (c–d) Enrichment plots and ROC curves for the two pharmacophore candidates of PharmaGist for class II compounds (the magenta and green lines).

(Intel(R) Xeon(TM) CPU 3.20 GHz) per database compound, depending on the compound's size.

Virtual Screening for G-Protein Coupled Receptors.

G-Protein Coupled Receptors (GPCRs) is an important family of drug targets. GPCRs are membrane proteins that are hard to crystallize. Since many of them lack crystal structures, GPCRs have long been the ultimate targets for pharmacophore-based virtual screening.⁴⁷ Here, we describe the performance of PharmaGist in identifying known antagonists of the $\alpha 1A$ target receptor in comparison to two other pharmacophore-based tools, Catalyst^{18–20,31} and Mtree.²⁶

Data and Experiment Description. We used a set of 18 known $\alpha 1A$ antagonists for pharmacophore detection. The compounds were divided into two classes based on their structures: 12 class I molecules and six class II molecules.⁵² A cover set of weighted pharmacophores was computed for each class of ligands separately, and then each of these weighted pharmacophore candidates was used to screen a large data set consisting of 50 active compounds and 950 decoys from the MDL Drug Data Report (MDDR) database.⁵²

Class I Pharmacophore Candidates. PharmaGist automatically selected three pharmacophore candidates (Figure 3). Two aromatic rings and one acceptor (close to the left aromatic ring) are shared by all the three pharmacophore models and by all the 12 input ligands. Other features as well as the pivot molecule vary between the three phar-

cophore models. The first pharmacophore candidate is the largest one. It consists of 13 features: three aromatic rings, two hydrophobic groups, two donors, five acceptors, and one charged group (Figure 3a,b). These 13 features are shared by only three input molecules. The second pharmacophore candidate consists of six features: two aromatic rings, one donor, two acceptors, and one charged group (Figure 3c,d). The third pharmacophore candidate also consists of six features: two aromatic rings, hydrophobic group, donor, acceptor, and charged group (Figure 3e,f). The three pharmacophore candidates were found to be similar to a pharmacophore hypothesis computed by Catalyst.⁵² This pharmacophore hypothesis consists of hydrophobic groups instead of the common aromatic rings detected by PharmaGist, the acceptor that is shared by all molecules and the charge feature in the middle.

Class I Virtual Screening. The screening results with each pharmacophore candidate are summarized in Table 2 and Figure 5a. The enrichment factor at the top 1% of the ranked database (EF_1) is 6.0, 14.0, and 10.0 for the first, second, and third pharmacophore candidates, respectively. The second pharmacophore candidate is much better in differentiating between actives and decoys compared to the first (and larger) pharmacophore candidate. The reason is probably due to the fact that the 13 features in the largest pharmacophore are shared by only three molecules. This artifact can be handled by an *a priori* clustering of the input molecules. The enrichment factor EF_1 for this pharmacophore

Table 1. Feature Definitions^a

feature	center	direction
<i>Hydrogen Bond Donor</i>		
O.3 S.3: < 2 > (1 H) (*)	atom	(*) → atom
N.4 N.3 N.2 N.am N.pl3: (H)	atom	atom → (H)
N.ar: < 3 > (H)	atom	atom → (H)
<i>Hydrogen Bond Acceptor</i>		
N.3 P.3: < 3 > (3 *)	atom	centroid of (3 *) → atom
O.3 S.3: < 2 > (2 *) [H]	atom	centroid of (2 *) → atom
N.2 N.ar: < 2 > (2 *)	atom	centroid of (2 *) → atom
O.2 S.2: < 1 > (*: < 3 >)	atom	(*: < 3 >) → atom
O.2: < 1 > (P S: < 4 >)	atom	(P S: < 4 >) → atom
N.1: < 1 > (*)	atom	(*) → atom
O.3 S.3: < 2 > (1 H) (*)	atom	(*) → atom
C.2: < 3 > (2 O.co2) (*)	2 × O.co2	(*) → atom
O.3: < 1 > (N.4: < 4 >)	atom	(N.4: < 4 >) → atom
<i>Acid</i>		
C.2: < 3 > (2 O.co2)	atom	—
P.3 S.3: (>2 O.co2)	atom	—
O.3: (N.4)	atom	—
<i>Base</i>		
C.cat N.4 N.ar: < 3 >	atom	—
<i>Aromatic Ring</i>		
A ring satisfying Definition 1 ^b	ring center	ring normal
<i>Hydrophobic</i>		
An atom pair satisfying Definition 2 ^c	pair center	—

^a The different features are listed in italics in the 1st column. *Hydrogen Bond Donor/Acceptor*, *Acid* and *Base* are single-atom features. Their definitions follow the ones of fFLASH⁷ and appear in the format: *atom types: bonds*. The *atom types* field is a list of SYBYL atom types separated by the | OR operator. The *bonds* field can specify the number of bonds of the atom in < >, the bonded atoms in () and atoms that should not be bonded in [] (* stands for any atom type). For example, O.3 | S.3: < 2 > (1 H) (*) defines an O.3 or S.3 atom attached to exactly one hydrogen and to additional atom. Each feature is represented in 3D space by its center (appears in the 2nd column). Note that for a carboxyl group, C.2: < 3 > (2 O.co2) (*), two feature points are defined. The 3rd column specifies the direction vector for some of the features. ^b Definition 1. Aromatic Ring. A planar ring of five or six atoms is aromatic if one of the following holds: (i) it has 4n+2 Pi electrons, where n is a non-negative integer (Hückel rule); (ii) it has five sp² atoms and one sp³ atom; or (iii) it has four sp² atoms and two sp³ atoms. ^c Definition 2. Hydrophobic Pair. A pair of nonpolar carbon atoms that are at least two bonds away from a heteroatom or a charged atom.

Table 2. Enrichment Factors for the GPCR Data Set^a

enrichment factor	PharmaGist 1	PharmaGist 2	PharmaGist 3	MTree	Catalyst
class I					
EF ₁	6.0	14.0	10.0	12.0	14.0
EF ₅	2.4	5.6	5.6	6.8	6.4
EF ₁₀	2.2	3.0	3.8	4.8	3.6
class II					
EF ₁	4.0	14.0	—	16.0	12.0
EF ₅	6.8	5.6	—	9.2	10.0
EF ₁₀	4.2	4.8	—	5.4	5.4

^a For each class of alpha1A compounds, enrichment factors at the top 1%, 5%, and 10% of the ranked GPCR database (EF₁, EF₅, and EF₁₀) are provided for the selected pharmacophore candidates of PharmaGist in comparison to the enrichment factors obtained by the pharmacophore models suggested by Catalyst^{18–20,31} and MTree.^{26,47}

the maximal enrichment is achieved for the top 1% of the ranked database.

Class II Pharmacophore Candidates. Two pharmacophore candidates were computed as suitable for virtual screening by PharmaGist for the six class II ligands (Figure 4). Similarly to the class I pharmacophore candidates, two aromatic rings and one acceptor are shared by the two pharmacophore candidates and the six input ligands. However, the common acceptor for this class is between the aromatic rings. In addition to these three common features, the first pharmacophore candidate includes an additional acceptor and a hydrophobic group (Figure 4a,b), which are shared by only three input molecules. The second pharmacophore candidate has an additional acceptor, namely it consists of four features overall: two aromatic rings and two acceptors (Figure 4c,d)). The additional acceptor is shared by five out of six input molecules. The pharmacophore models suggested by Catalyst and MTree⁴⁷ are similar to the pharmacophore candidates of PharmaGist. They all include the same two aromatic rings, and there is a small variance in their other features. Additionally, the additional acceptor of the second selected pharmacophore candidate of PharmaGist is also observed in the pharmacophore model of MTree.

Class II Virtual Screening. Table 2 and Figure 5b summarize the enrichment results obtained for screening with each of the two pharmacophore candidates. The enrichment factor at the top 1% of the ranked database (EF₁) is 4.0 and 14.0 for the first and second selected pharmacophore candidates, respectively (Table 2). The enrichment factor for the first pharmacophore increases to 6.8 at the top 5% of the ranked database. Similarly to the results for class I, the second pharmacophore candidate is much better in differentiating between actives and decoys than the first (and larger) pharmacophore candidates. The EF₁ enrichment factor for this pharmacophore candidate (14.0) is comparable with the enrichment factors of the pharmacophore models suggested by MTree (16.0) and Catalyst (12.0).⁴⁷

Large-Scale Evaluation. We evaluated the performance of PharmaGist also on the DUD (directory of useful decoys) benchmark data set.⁴⁶ DUD contains 2950 annotated active ligands for 40 different protein receptors. In addition, each ligand has about 36 decoy molecules that are topologically distinct from the particular ligand but physically similar to it in attributes, such as molecular weight and number of hydrogen bonds. The generation of DUD is based on the observation that the enrichment factor greatly depends on the physical properties of the background set of decoys. Specifically, misleading good enrichment might result if the physical properties of the decoys can easily be distinguished from those of the active ligands. Overall, the DUD database contains 2950 annotated ligands and 95316 decoys for 40 different receptors.

Experiment Description. The experiment for measuring the performances of PharmaGist for virtual screening on DUD was carried out as follows. For each of the 40 target receptors in DUD, we computed a set of weighted pharmacophores. This was performed by five runs of PharmaGist on random subsets (i.e., training sets) with five native ligands and additional five runs on random subsets with ten native ligands. PharmaGist selected only one weighted pharmacophore for each input with five ligands and at most three

candidate (14.0) is comparable with the enrichment factors of MTree (12.0) and Catalyst (14.0).⁴⁷ It is also notable that

Table 3. DUD Screening Benchmark^a

receptor	data		EF_1				DOCK
	no. of ligands	no. of decoys	5 ₁	10 ₁	10 ₁₋₂	10 ₁₋₃	
ACE	49	1728	12.6	12.0	12.6	12.1	40.4
ACHE	105	3732	28.6	36.0	36.0	35.8	1.9
ADA	23	822	17.0 (29.4)	19.1 (30.3)	17.4 (29.7)	17.4 (29.5)	12.9
ALR2	26	920	22.3 (18.1)	26.5 (26.3)	25.6 (25.4)	23.4 (23.4)	38.1
AmpC	21	734	56.4	67.0	67.0	64.2	17.1
AR	74	2630	27.0	19.3	19.1	18.3	33.5
CDK2	50	1780	17.3	11.5	13.1	13.1	13.9
COMT	12	430	41.1	—	—	—	0.0
COX-1	25	850	8.3	4.1	3.6	3.3	4.0
COX-2	349	12491	28.7	30.1	30.1	30.1	20.0
DHFR	201	7150	21.2	20.0	21.3	21.3	21.7
EGFr	416	14914	32.8	29.7	34.7	34.4	2.1
ER _{agonist}	67	2361	28.7	21.6	20.8	20.7	19.2
ER_{antagonist}	39	1399	25.0 (31.2)	27.2 (39.3)	27.2 (57.2)	27.2 (57.2)	12.7
FGFr1	118	4216	33.0	30.8	37.0	37.0	0.0
FXa	142	5102	35.0	43.8	43.8	42.7	14.6
GART	21	753	15.7	23.6	23.6	23.6	42.4
GPB	52	1851	38.1	41.6	43.9	42.5	22.8
GR	78	2804	23.4	36.4	36.7	35.6	8.9
HIVPR	53	1888	30.5	39.4	39.3	38.0	3.7
HIVRT	40	1439	2.0	7.1	7.0	6.9	5.0
HMGR	35	1242	34.9	36.9	47.1	44.6	33.9
HSP90	24	861	33.5	41.0	41.0	39.6	8.6
InhA	85	3043	35.3 (33.1)	34.9 (42.6)	35.3 (42.0)	34.7 (42.0)	0.0
MR	15	535	58.1	42.7	68.4	68.4	46.2
NA	49	1745	45.0	34.1	43.9	43.9	20.2
P38 MAP	234	8399	20.6 (49.1)	19.1 (46.7)	19.1 (45.2)	19.0 (45.0)	2.1
PARP	33	1178	40.8	44.8	41.1	0.0	6.0
PDE5	51	1810	16.1	17.6	19.3	19.3	11.8
PDGFRb	157	5625	20.4	23.9	25.4	24.6	0.0
PNP	25	884	26.2	29.3	29.4	28.4	31.7
PPARg	81	2910	40.1	42.2	43.2	43.2	0.0
PR	27	967	35.1	49.8	51.4	45.7	0.0
RXRa	20	708	57.4	81.7	81.7	81.7	24.8
SAHH	33	1159	32.9	32.3	32.3	31.7	78.0
SRC	162	5801	25.0	24.6	36.6	35.9	1.2
thrombin	65	2294	29.5	41.6	41.6	41.6	13.7
TK	22	785	15.4 (53.0)	15.0 (33.3)	26.8 (53.3)	26.8 (53.3)	54.0
trypsin	43	1545	33.3	39.0	44.7	44.5	22.5
VEGFR2	74	2647	10.3	10.5	11.2	11.1	1.3
average			28.9 (35.7)	31.0 (36.4)	33.3 (42.1)	32.4 (41.7)	17.3

^a The table mainly presents the enrichment factors of PharmaGist against the 'own decoys' database of DUD in comparison to the enrichment factors of DOCK against the 'entire' DUD database.⁴⁶ Note that screening against the 'own decoys' database is more challenging than against the 'entire' DUD database. Nevertheless, PharmaGist outperformed DOCK. For six representative cases (in bold) the enrichment factors of PharmaGist against the 'entire' DUD database are also provided in brackets. The data appearing in the columns are as follows: (i-iii) the name of the protein receptor, the number of its annotated ligands, and the total number of their decoys; (iv) the average enrichment factor at the top 1% of the ranked database (EF_1) for the five runs of PharmaGist on random subsets with five ligands; (v-vii) the average EF_1 factor for the three screening ways using pharmacophore candidates selected by PharmaGist for five runs on random subsets with ten ligands (see the text); and (viii) the EF_1 factor of DOCK. Note that the COMT receptor has only 12 annotated ligands. Thus, screens with selected pharmacophore candidates for ten input ligands were not performed (10₁, 10₁₋₂, and 10₁₋₃).

weighted pharmacophores for each input with ten ligands. Then, the selected weighted pharmacophores were used to screen a data set consisting of all the native ligands of the receptor without the ligands in the training set (i.e., test set) and their corresponding decoys (called the 'own decoys' database in DUD⁴⁶). For runs on ten input ligands and for which more than one weighted pharmacophore was selected by the method, we ranked the compounds in the database in three ways: (i) based on the score obtained for their similarity with only the first selected weighted pharmacophore (10₁), (ii) based on the highest score obtained for their similarity with the first and second selected weighted pharmacophore (10₁₋₂), and (iii) the same as in (ii) but with all the three selected weighted pharmacophores (10₁₋₃).

Overall Enrichment Results. The screening results were evaluated by an enrichment factor at the top 1% of the ranked database (EF_1) and are summarized in Table 3. Significant enrichment was observed for most receptors. Additionally, the results are competitive with the ones obtained by screening with the docking method, DOCK.^{48,49} Screening with DOCK was performed both against the 'own decoys' database as well as against the 'entire' DUD database.⁴⁶ The 'own decoys' database is more challenging for screening. Thus, as reported by Huang et al., the results of DOCK are poorer against this database than against the 'entire' data set, and dramatic differences in enrichment were observed for some receptors.⁴⁶ In contrast, the enrichment results of PharmaGist against the 'own decoys' database were far above

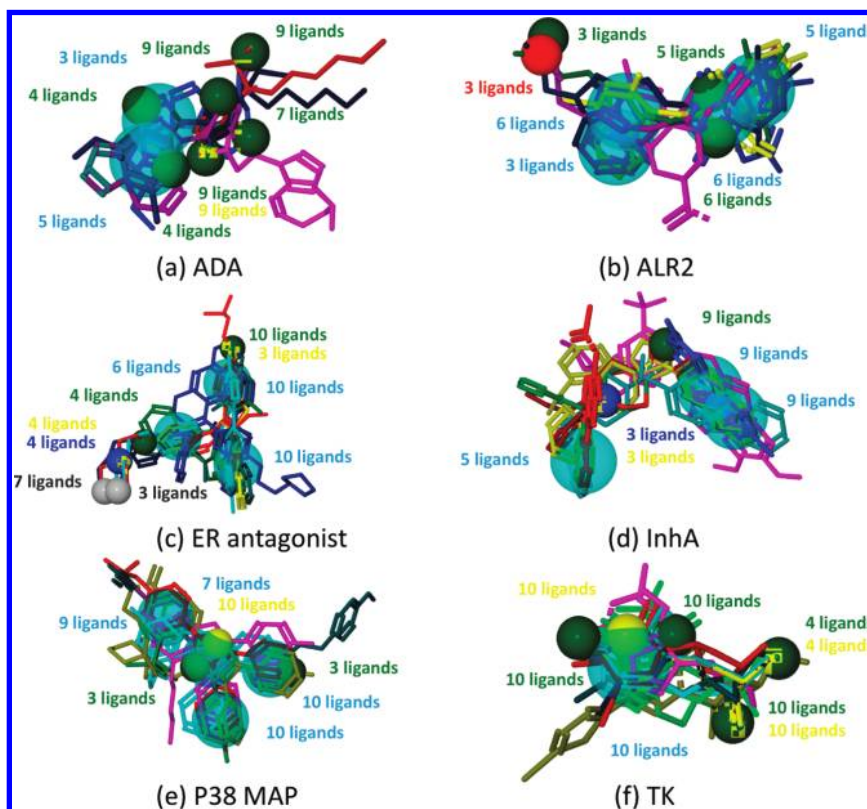


Figure 6. Pharmacophore candidates and the corresponding multiple alignments of ligands for the six representative receptors of DUD.⁴⁶ These were computed by a run of PharmaGist on a random set of ten annotated ligands. Note that each feature in these pharmacophore candidates is weighted by the number of aligned ligands that possess it.

random. On average, about 31% of the annotated ligands were found at the top 1% of the ranked database (corresponding to the average *EF*), compared to 17.3% obtained for DOCK against the ‘entire database’, which is the less challenging database (Table 3). For most of the 40 receptors, PharmaGist achieved a more significant enrichment than DOCK, where for only eight receptors the situation was reversed. However, it should be noted that for at least three out of these eight cases (TK, PNP, and SAHH), the reason for the differences in the enrichment obtained for the two methods results from the fact that the ‘own decoys’ database is more difficult for screening than the ‘entire’ database.⁴⁶ It is also notable that the maximum enrichment factor of PharmaGist was obtained for 93% of the screens at the top 1% of the ranked database.

Representative Targets. We examined the six representative receptors of Huang et al.⁴⁶ in greater details. For these receptors, screening with PharmaGist was performed also on the ‘entire’ DUD database.⁴⁶ As expected, the obtained enrichment factors for these receptors were much higher than the ones obtained against the ‘own decoys’ database, where, on average, about 39% of the annotated ligands were found at the top 1% of the ranked database (corresponding to the average EF_1) (highlighted in Table 3 and Figure 7). Pharmacophore candidates and the corresponding multiple alignments of ligands for the six representative receptors are presented in Figure 6. These pharmacophore candidates were computed by a run of PharmaGist on a random set of ten annotated ligands. Note that each feature in these pharmacophore candidates is weighted by the number of aligned ligands that possess it.

CONCLUSIONS AND FUTURE WORK

We have described PharmaGist, a fully automated indirect method for both pharmacophore elucidation and virtual screening. PharmaGist has two modules. Given a collection of druglike molecules, the first module of PharmaGist computes a set of pharmacophore candidates by multiple flexible alignments of input molecules. The second module uses the pharmacophore candidates for virtual screening in order to identify new compounds that bind to the target receptor.

From the computational standpoint, PharmaGist consists of two main algorithms for pairwise and multiple alignment of flexible druglike molecules. The computational problems addressed by the two algorithms are NP-hard. Nevertheless, PharmaGist is highly efficient. Specifically, the pairwise alignment algorithm provides a fast heuristic solution with polynomial time complexity. The multiple alignment algorithm is also fast. Its time complexity is exponential only in the number of features (which is relatively low) and polynomial in the number of input ligands and their degrees of freedom.

The capability of PharmaGist to identify correct pharmacophores was successfully tested on the FlexS data set.⁴⁵ Here, we have performed additional tests for examining the applicability of PharmaGist not only for pharmacophore detection but mainly for virtual screening. We have evaluated its performance in identifying known antagonists of the alpha1A G-Protein Coupled Receptor from a large virtual library. In addition, we have performed a large-scale evaluation using the DUD data set, which contains 2950 annotated ligands and 95316 decoys for 40 different receptors. The results show the success of the method in comparison to other methods for pharma-

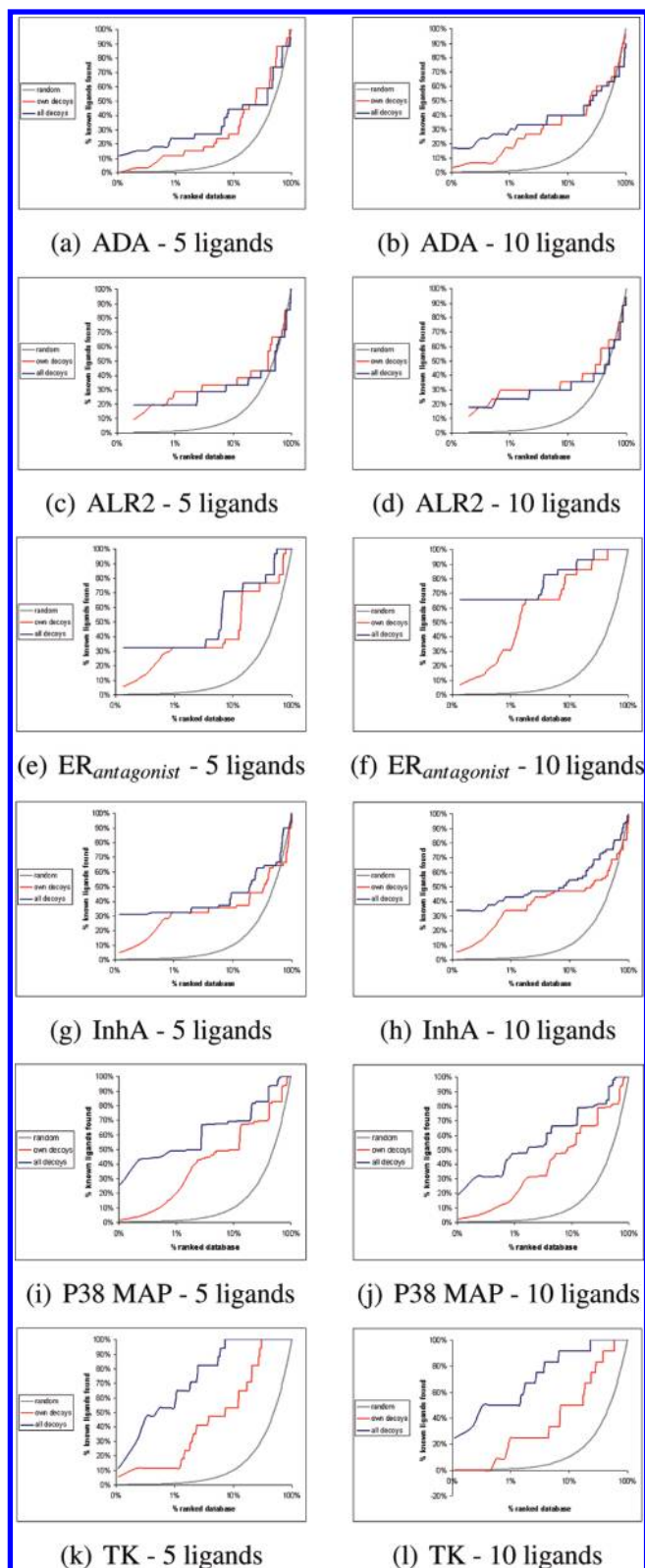


Figure 7. Enrichment plots for the six selected cases from DUD data set. The percentage of ranked database (x-axis) is plotted against the percentage of known ligands (y-axis) found at any given percentage of ranked database. The gray line is the enrichment plot for a random predictor, the red line represents screening results with 'own decoys' database, and the blue line is for the 'entire' DUD database.

The results also demonstrate the advantages of the method. PharmaGist is able to deal with different types of target receptors. The flexibility of the ligands is fully considered in a deterministic manner, an attribute that is unique to PharmaGist. This allows the user to screen large databases without *a priori* conformational search of the compounds in the database. Another important attribute of the method is the capability of detecting pharmacophore candidates that are shared by nonpredefined subsets of input ligands. This makes PharmaGist tolerant to the presence of outlier ligands and may aid in distinguishing between pharmacophores of different binding sites. There is a trade-off between the number of aligned molecules and the number of their common features. This trade-off is addressed in PharmaGist by introducing the notion of a weighted pharmacophore. In such a model for a pharmacophore, each feature is weighted based on the number of ligands that possess it. This means that some of the features do not necessarily appear in all active ligands. In addition, we have proposed a strategy for selecting a small cover set of weighted pharmacophores for virtual screening.

The software for pharmacophore detection and virtual screening is available for download. In addition, a user-friendly Web interface for pharmacophore detection is available at <http://bioinfo3d.cs.tau.ac.il/PharmaGist>. The software allows fully automated application of the method. In addition, it allows full user control of the program parameters through a configuration file. In the pharmacophore detection stage, the user can set a pivot molecule, change feature weights, and define a new feature type as well as change more advanced method parameters. The user can also manually select which of the pharmacophores produced by the method should be used in the virtual screening stage.

In future work we intend to improve the method so it will be able to detect pharmacophores that are present on *nonadjacent* rigid groups of the ligands. PharmaGist can also be strengthened by considering ring conformations, deriving a pharmacophore shape as the negative image of the receptor binding site and introducing the notion of excluded volumes.

ACKNOWLEDGMENT

The authors thank D. Fishlovitch, A. Oron, and H. Senderowitz for their useful advice and A. Evers for providing the GPCR data set. The research of O. Dror and Y. Inbar has been supported by the Eshkol Fellowship funded by the Israeli Ministry of Science. The research of H. J. Wolfson has been supported in part by the Israel Science Foundation (grant no. 281/05) and by the Hermann Minkowski-Minerva Center for Geometry at TAU. The research of H. J. Wolfson and R. Nussinov has been supported by the NIAID, NIH (grant No. 1UC1AI067231), and by the Binational US-Israel Science Foundation (BSF). This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number N01-CO-12400. The content of this publication does not necessarily reflect the view of the policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government. This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

cophore-based screening, Catalyst^{18–20,31} and MTree,^{26,47} as well as to receptor-based screening with the DOCK method.^{48,49}

REFERENCES AND NOTES

- (1) Sun, H. Pharmacophore-based Virtual Screening. *Curr. Med. Chem.* **2008**, *15*, 1018–24.
- (2) Shoichet, B. K. Virtual Screening of Chemical Libraries. *Nature* **2004**, *432*, 862–865.
- (3) Schneidman-Duhovny, D.; Nussinov, R.; Wolfson, H. J. Predicting Molecular Interactions in silico II: Protein-protein and Protein-drug Docking. *Front. Med. Chem.* **2006**, *3*, 585–614.
- (4) Lengauer, T.; Lemmen, C.; Rarey, M.; Zimmermann, M. Novel Technologies for Virtual Screening. *Drug Discovery Today* **2004**, *9*, 27–34.
- (5) Lemmen, C.; Lengauer, T. Time-efficient Flexible Superposition of Medium-sized Molecules. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 357–368.
- (6) Lemmen, C.; Lengauer, T.; Klebe, G. FlexS: A Method for Fast Flexible Ligand Superposition. *J. Med. Chem.* **1998**, *41*, 4502–4520.
- (7) Krämer, A.; Horn, H.; Rice, J. Fast 3D Molecular Superposition and Similarity Search in Databases of Flexible Molecules. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 13–38.
- (8) Rognan, D. Chemogenomic Approaches to Rational Drug Design. *Br. J. Pharmacol.* **2007**, *152*, 38–52.
- (9) Klabunde, T. Chemogenomic Approaches to Drug Discovery: Similar Receptors Bind Similar Ligands. *Br. J. Pharmacol.* **2007**, *152*, 5–7.
- (10) Dror, O.; Shulman-Peleg, A.; Nussinov, R.; Wolfson, H. J. Predicting Molecular Interactions In Silico: I. An Updated Guide to Pharmacophore Identification and its Applications to Drug Design. *Front. Med. Chem.* **2006**, *3*, 551–584.
- (11) *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: La Jolla, CA, U.S.A., 2000.
- (12) Wolber, G.; Seidel, T.; Bendix, F.; Langer, T. Molecule-pharmacophore Superpositioning and Pattern Matching in Computational Drug Design. *Drug Discovery Today* **2008**, *13*, 23–9.
- (13) 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.
- (14) Holliday, J.; Willett, P. Using a Genetic Algorithm to Identify Common Structural Features in Sets of Ligands. *J. Mol. Graphics Modell.* **1997**, *15*, 203–253.
- (15) Handschuh, S.; Wagener, M.; Gasteiger, J. The Search for the Spatial and Electronic Requirements of a Drug. *J. Mol. Model.* **2000**, *6*, 358–378.
- (16) Finn, P. W.; Kavrakli, L. E.; Latombe, J.-C.; Motwani, R.; Shelton, C.; Venkatasubramanian, S.; Yao, A. RAPID: Randomized Pharmacophore Identification for Drug Design. *Comput. Geom.: Theory Appl.* **1998**, *10*, 263–272.
- (17) Chen, X.; Rusinko, A., III.; Tropsha, A.; Young, S. Automated Pharmacophore Identification for Large Chemical Data Sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 887–896.
- (18) Güner, O. F.; Clement, O.; Kurogi, Y. Pharmacophore Modeling and Three Dimensional Database Searching for Drug Design Using Catalyst: Recent Advances. *Curr. Med. Chem.* **2004**, *11*, 2991–3005.
- (19) Clement, O. A.; Mehl, A. T. In *Pharmacophore Perception, Development, and Use in Drug Design*; International University Line: La Jolla, CA, U.S.A., 2000; Chapter HipHop: Pharmacophores Based on Multiple Common-Feature Alignments, pp 69–84.
- (20) Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. Identification of Common Functional Configurations among Molecules. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563–571.
- (21) Li, H.; Sutter, J.; Hoffmann, R. In *Pharmacophore Perception, Development, and Use in Drug Design*; International University Line: La Jolla, CA, U.S.A., 2000; Chapter HypGen: An Automated System for Generating 3D Predictive Pharmacophore Models, pp 171–189.
- (22) Wolber, G.; Dornhofer, A.; Langer, T. Efficient Overlay of Small Organic Molecules Using 3D Pharmacophores. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 773–88.
- (23) Crandell, C.; Smith, D. Computer-assisted examination of compounds for common three-dimensional substructures. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 186–197.
- (24) Brint, A.; Willett, P. Algorithms for the identification of three-dimensional maximal common substructures. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 152–158.
- (25) Takahashi, Y.; Satoh, Y.; Suzuki, H.; Sasaki, S. Recognition of Largest Common Structural Fragment among a Variety of Chemical Structures. *Anal. Sci.* **1987**, *3*, 23–28.
- (26) Hessler, G.; Zimmermann, M.; Matter, H.; Evers, A.; Naumann, T.; Lengauer, T.; Rarey, M. Multiple-Ligand-Based Virtual Screening: Methods and Applications of the MTree Approach. *J. Med. Chem.* **2005**, *48*, 6575–6584.
- (27) Günther, S.; Senger, C.; Michalsky, E.; Goede, A. Preissner, R. Representation of target-bound drugs by computed conformers: implications for conformational libraries. *BMC Bioinf.* **2006**, *7*, 293.
- (28) Martin, Y.; Bures, M.; Dahaner, E.; DeLazzer, J.; Lico, I.; Pavlik, P. A fast new approach to pharmacophore mapping and its Application to Dopaminergic and Benzodiazepine agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83–102.
- (29) Martin, Y. C. In *Pharmacophore Perception, Development, and Use in Drug Design*; International University Line: La Jolla, CA, U.S.A., 2000; Chapter DISCO: What We Did Right and What We Missed, pp 49–68.
- (30) Dixon, S.; Smondyrev, A.; Knoll, E.; Rao, S.; Shaw, D.; Friesner, R. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 647–71.
- (31) Kurogi, Y.; Güner, O. F. Pharmacophore Modeling and Three-Dimensional Database Searching for Drug Design Using Catalyst. *Curr. Med. Chem.* **2001**, *8*, 1035–1055.
- (32) Feng, J.; Sanil, A.; Young, S. PharmID: Pharmacophore Identification Using Gibbs Sampling. *J. Chem. Inf. Model.* **2006**, *46*, 1352–1359.
- (33) Zhu, F.; Agrafiotis, D. Recursive Distance Partitioning Algorithm for Common Pharmacophore Identification. *J. Chem. Inf. Model.* **2007**, *47*, 1619–1625.
- (34) Podolyan, Y.; Karypis, G. Common Pharmacophore Identification Using Frequent Clique Detection Algorithm. *J. Chem. Inf. Model.* **2009**, *49*, 13–21.
- (35) Jones, G.; Willett, P.; Glen, R. A Genetic Algorithm for Flexible Molecular Overlay and Pharmacophore Elucidation. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532–49.
- (36) Cottrell, S. J.; Gillet, V. J.; Taylor, R.; Wilton, D. J. Generation of Multiple Pharmacophore Hypotheses Using Multiobjective Optimization Techniques. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 665–682.
- (37) Handschuh, S.; Wagener, M.; Gasteiger, J. Superimposition of Three-Dimensional Chemical Structures Allowing for Conformational Flexibility by a Hybrid Method. *J. Chem. Inf. Comput. Sci.* **1997**, *38*, 220–232.
- (38) Handschuh, S.; Gasteiger, J. In *Pharmacophore Perception, Development, and Use in Drug Design*; International University Line: La Jolla, CA, U.S.A., 2000; Chapter Pharmacophores Derived from the 3D Substructure Perception, pp 429–454.
- (39) Jones, G.; Willett, P.; Glen, R. In *Pharmacophore Perception, Development, and Use in Drug Design*; International University Line: La Jolla, CA, U.S.A., 2000; Chapter GASP: Genetic Algorithm Superposition Program, pp 85–106.
- (40) Richmond, N.; Abrams, C.; Wolohan, P.; Abrahamian, E.; Willett, P.; Clark, R. GALAHAD: 1. Pharmacophore Identification by Hypermolecular Alignment of Ligands in 3D. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 567–87.
- (41) Shepphird, J. K.; Clark, R. D. A Marriage Made in Torsional Space: Using GALAHAD Models to Drive Pharmacophore Multiplet Searches. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 763–771.
- (42) Shatsky, M.; Shulman-Peleg, A.; Nussinov, R.; Wolfson, H. J. The Multiple Common Point Set Problem and Its Application to Molecule Binding Pattern Detection. *J. Comput. Biol.* **2006**, *13*, 407–42.
- (43) Baum, D. Multiple Semi-flexible 3D Superposition of Drug-sized Molecules. *Computational Life Sciences First International Symposium, CompLife 2005*, Konstanz, Germany, 2005; pp 198–207.
- (44) Inbar, Y.; Schneidman-Duhovny, D.; Dror, O.; Nussinov, R.; Wolfson, H. J. Deterministic Pharmacophore Detection Via Multiple Flexible Alignment of Drug-Like Molecules. *Research in Computational Molecular Biology (RECOMB)*, 11th Annual International Conference, Oakland, CA, U.S.A., 2007; pp 412–429.
- (45) Schneidman-Duhovny, D.; Dror, O.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. Deterministic Pharmacophore Detection via Multiple Flexible Alignment of Drug-Like Molecules. *J. Comput. Biol.* **2008**, *15*, 737–54.
- (46) Huang, N.; Shoichet, B.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.
- (47) Evers, A.; Hessler, G.; Matter, H.; Klubunde, T. Virtual Screening of Biogenic Amine-Binding G-Protein Coupled Receptors: Comparative Evaluation of Protein- and Ligand-Based Virtual Screening Protocols. *J. Med. Chem.* **2005**, *48*, 5448–5465.
- (48) Kuntz, I.; Blaney, J.; Oatley, S.; Langridge, R.; Ferrin, T. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
- (49) Ewing, T.; Makino, S.; Skillman, A.; Kuntz, I. DOCK 4.0: Search Strategies for Automated Molecular Docking of Flexible Molecule Databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411–428.
- (50) Stockman, G. Object Recognition and Localization via Pose Clustering. *J. Comput. Vision, Graphics, Image Process.* **1987**, *40*, 361–387.
- (51) Lamdan, Y.; Wolfson, H. J. Geometric Hashing: A General and Efficient Model-Based Recognition Scheme. *Proceedings of the IEEE Int. Conf. on Computer Vision*, Tampa, Florida, U.S.A., 1988; pp 238–249.
- (52) Klubunde, T.; Evers, A. GPCR Antitarget Modeling: Pharmacophore Models for Biogenic Amine Binding GPCRs to Avoid GPCR-mediated Side Effects. *ChemBioChem* **2005**, *6*, 876–889.