Novel Approach to Structure-Based Pharmacophore Search Using Computational Geometry and Shape Matching Techniques

Jerry Osagie Ebalunode,[†] Zheng Ouyang,[‡] Jie Liang,*,[‡] and Weifan Zheng*,[†]

Department of Pharmaceutical Sciences, BRITE Institute, North Carolina Central University, 1801 Fayetteville Street, Durham, North Carolina 27707, and Bioengineering Department, University of Illinois at Chicago, Chicago, Illinois 60612

Computationally efficient structure-based virtual screening methods have recently been reported that seek to find effective means to utilize experimental structure information without employing detailed molecular docking calculations. These tools can be coupled with efficient experimental screening technologies to improve the probability of identifying hits and leads for drug discovery research. Commercial software ROCS (rapid overlay of chemical structures) from Open Eye Scientific is such an example, which is a shape-based virtual screening method using the 3D structure of a ligand, typically from a bound X-ray costructure, as the query. We report here the development of a new structure-based pharmacophore search method (called *Shape4*) for virtual screening. This method adopts a variant of the ROCS shape technology and expands its use to work with an *empty* crystal structure. It employs a rigorous computational geometry method and a deterministic geometric casting algorithm to derive the negative image (i.e., pseudoligand) of a target binding site. Once the negative image (or pseudoligand) is generated, an efficient shape comparison algorithm in the commercial OE SHAPE Toolkit is adopted to compare and match small organic molecules with the shape of the pseudoligand. We report the detailed computational protocol and its computational validation using known biologically active compounds extracted from the WOMBAT database. Models derived for five selected targets were used to perform the virtual screening experiments to obtain the enrichment data for various virtual screening methods. It was found that our approach afforded similar or better enrichment ratios than other related methods, often with better diversity among the top ranking computational hits.

INTRODUCTION

Drug discovery and development is a lengthy and costly process. It has been widely reported that, on average, over 10 years and 800 million dollars are needed to bring a drug from early discovery to the market.[1,2] Thus, it is not surprising that pharmaceutical companies and biotech startups have invested so much in the development of new enabling technologies to help reduce the cycle time and cost as well as to increase the productivity of drug discovery research.[3] Among the enabling technologies are high-throughput screening (HTS),[4] combinatorial chemistry,[5] high-throughput X-ray crystallography,[6] computer-aided drug design (CADD),[7] and cheminformatics tools.[8] Of special interest to us are the development and application of novel computational methods for lead generation and lead optimization in the drug discovery process. These computational methods are generally categorized as ligand-based methods and (receptor) structure-based methods.

Ligand-based computational methods are often employed when detailed structural information is not available for the target of interest, or the biological target is completely unknown as is the case in many phenotypic assay based discovery.[9] In such cases, similarity search around an initial HTS (high throughput screening) hit based on either 2D or 3D similarity measure is often quite useful. When biological activities of multiple hits are known, a more sophisticated class of computational techniques known as pharmacophore identification methods is often employed to deduce the essential features required for the biological activity.[10] If a significant number of confirmed SAR (structure–activity relationship) data becomes available for a discovery project, one can also apply QSAR (quantitative structure–activity relationship) methods to build statistical models that can be used to predict new active molecules via virtual screening or inverse-QSAR process.[11–13] One key prerequisite for all these methods to be successful is the availability of confirmed chemical and biological data for a series of molecules. In other words, one has to know some active compounds in order to search for other active compounds in the hope that one may obtain either better molecules or perhaps a novel series of compounds (scaffold hopping). These ligand-based methods proved to be especially useful for targets of GPCR and ion channel gene families.[14,15] However, the often cited drawback of the ligand-based methods is that they do not provide detailed structural information to help medicinal chemists design new molecules, which is critical especially during the lead optimization stage of the discovery process. Because of this pitfall, structure-based methods are often employed in this phase.

The structure-based methods are typified by various docking technologies that have been widely adopted by the

* To whom all correspondence should be addressed. Tel.: 919 530 6752 (W.Z.). Fax: 919 530 6600 (W.Z.). E-mail: wzheng@nccu.edu (W.Z.); jliang@uic.edu (J.L.).
[†] North Carolina Central University.
[‡] University of Illinois at Chicago.

pharmaceutical industry for virtual (in silico) screening, library design and one-off molecular design. They are often the computational tools of choice for both lead generation and lead optimization because of the detailed structural information and testable hypotheses provided by these methods. In the past 20 years, docking methods have evolved from the initial development of the DOCK program[16] to many flavors of state-of-the-art commercial and academic docking tools. Widely used docking programs today include, but not limited to, FRED,[17] GOLD,[18] FlexX,[19] Glide,[20] DOCK,[21] AutoDock,[22] ICM,[23] and Surflex-docking,[24] just to name a few. Despite many reports of successful applications of off-the-shelf docking methods, serious issues remain unsolved. For example, docking tools in some cases still can not reliably generate docking poses consistent with X-ray experimental data. Frequently, docking scores are not well correlated with known binding data and should best be used as a guide or filter rather than a quantitative ranking tool.[25] In practice, empirical docking methods have often been used to augment regular docking scores such as SDOCKER[26] and PharmDock[27] to overcome these problems.

Recently, more intuitive and computationally more efficient structure-based pharmacophore methods have been reported that seek to find effective means to utilize experimental structure information without employing detailed docking calculations. These tools can (should) be coupled with efficient, experimental screening technologies to improve the probability of success in the discovery process. For example, LigandScout has been successfully applied in several virtual and experimental HTS projects.[28] They are used as effective virtual screening tools and provide testable hypotheses for medicinal chemists to study experimentally.[29] Other commercial tools that have similar functionality include Catalyst[30] and MOE.[31] However, most of these structure-based pharmacophore methods ignore the geometric or topographic constraints imposed by the receptor binding site. Even though some methods can use *excluded volumes* to impose rough geometric and topographic constraints, they are computationally inefficient. To our best knowledge, no methods have been reported on using accurate shapes derived from rigorous computational geometry analysis of the binding site to construct shape-based pharmacophore queries and use them for large scale virtual screening.

Here, we report the development of a new structure-based pharmacophore search method (called *Shape4*) for virtual screening. This method adopts a variant of the commercial ROCS (rapid overlay of chemical structures) technology from Open Eye Scientific and expands its use to work with an empty crystal structure. It employs a rigorous computational geometry method and a deterministic geometric casting algorithm to derive the negative image (i.e., pseudoligand) of a target binding site. Once the negative image (or pseudoligand) is generated, an efficient shape comparison algorithm in the commercial OE SHAPE Toolkit[32] is adopted to compare and match small organic molecules with the shape of the pseudoligand. We report the detailed computational protocol of our method and its computational validation using known biologically active compounds extracted from the WOMBAT database.[33] Five protein targets (phosphodiesterases, HIV reverse transcriptase, HIV protease, estrogen receptors and thymidine kinase), and their ligands (inhibitors, agonists) have been selected as the

validation sets because of the availability of their crystal structures and their use as validation targets in other related studies.[34] Models derived for these targets are used to perform the virtual screening experiments to obtain the enrichment data for various methods. It is found that our new approach (dubbed Shape4 for shape pharmacophore) affords better or similar enrichment ratios than other methods studied in this work.

MATERIALS AND METHODS

We describe in this section the details of the computational methods used for virtual screening in this work. These methods include our new development, the Shape4 program and two related existing methods, FRED and ROCS, which were used as benchmarks for comparative studies. We also describe in detail the database used for virtual screening and five sets of known biologically active compounds used to validate the computational hits obtained with the aforementioned virtual screening methods.

1. Structure-Based Shape Pharmacophore Program (Shape4). As mentioned earlier, current pharmacophore based virtual screening methods often ignore the intricate details of the binding site shapes and focus only on the key pharmacophore elements as the query.[28,30] Thus, they miss critically important information during the virtual screening process, resulting in more false positives than they should. For example, large molecules with multiple side chains attached to a central scaffold may be selected as (false) positives, simply because their core structures have the required pharmacophore elements. This situation may be alleviated if binding site excluded volumes are considered during the virtual screening process. However, the computational efficiency of such processes is dramatically reduced due to the required frequent checks for ligand clashes with the excluded volumes. Shape4 is thus designed to increase the efficiency of such methods for database searching while taking into account the topographical constraints of the target binding site to help reduce the false positive rate.

1.1. Overall Workflow of Shape4. To address the pitfalls and shortcomings of current structure-based pharmacophore methods, we have been engaged in the study of a new structure-based shape pharmacophore method for virtual screening. It is based on the shape technology developed by Open Eye Scientific. It also employs the computational geometry algorithms (Delauney tessellation/α-shape analysis) to detect the binding site atoms[35] and generate a negative image of the target binding site. This negative image is first represented by a set of spheres of different sizes. A variety of techniques can then be applied to represent the overall surface shape of this set of spheres. Currently, we use the shape representation and comparison functions in the Shape Toolkit provided by OE Scientific[32] due to the well-documented computational efficiency of the Gaussian-based shape algorithms. Other computer vision methods are also being explored to improve the accuracy and efficiency of our method. The overall flow of Shape4 is shown in Figures 1 and 2, which involves the following steps: (a) The **α-shape** based program is first used to detect potential binding site (pocket) atoms and the Delauney tetrahedra formed by these atoms for a given protein structure. (b) **nCast** (section 1.3) is then used to calculate orthogonal centers defined by the
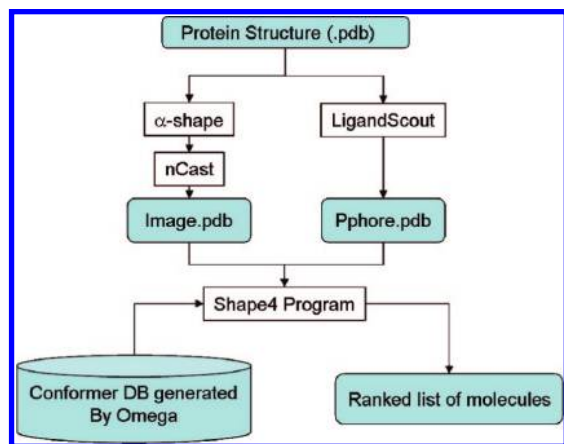
PHARMACOPHORE SEARCH USING MATCHING TECHNIQUES

J. Chem. Inf. Model., Vol. 48, No. 4, 2008 **891**

Figure 1. Overall flow diagram for the virtual screening process with Shape4.
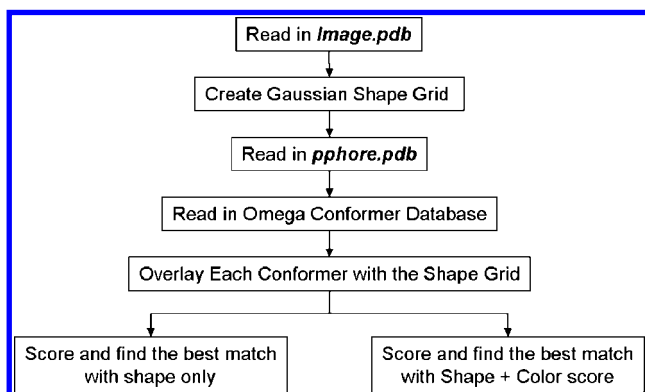

Figure 2. Flowchart for the Shape4 program.

vertices of the above Delauney tetrahedra and (c) generate inner spheres around each orthogonal center. Filters may be applied to eliminate the spheres that are unlikely to be useful for ligand discovery. (d) The overall shape of this collection of spheres (stored as image.pdb) is then represented by Gaussian functions based on the OE SHAPE library. (e) The shape representation is then used by Shape4 to query a database of molecules whose conformers are pregenerated (e.g., using the program Omega).[36] This protocol implements an efficient, basic structure-based shape matching method for virtual screening. To allow for more detailed information to be used in the virtual screening, we have added pharmacophore group information in the shape matching process. We have implemented in Shape4 the combination of Ligand-Scout[37] generated pharmacophore elements, the α-shape derived binding site image, and the OE shape matching function. The overall similarity score between the query and a matching ligand is calculated as the weighted average between the Fit Tversky score and the pharmacophore matching score (the color score).

1.2. Identification of the Binding Site Atoms. The binding site was first identified by the α-shape based program.[38] It was developed from a large set of precomputed protein surfaces. A probabilistic model was used to predict whether a residue located in a surface pocket is functionally important based on the analysis of bias of functionally important key residues in composition, in secondary structure, and in atomic patterns. This method is sequence and fold independent. It is able to identify systematically functional surfaces of

enzyme proteins with 91.2% accuracy. The details are provided in that publication.[38]

1.3. Generation of Binding Site Negative Images (nCast). The protein's binding pockets are first generated by CASTp.[39,40] The topological relationships among atoms in the protein are first established using Delaunay tessellation in 3D space.[41] Alpha shapes are then obtained by removing Delaunay edges and triangles whose corresponding Voronoi edges and vertices are not contained within the molecule. Further analytical measurements are used to identify the shape of a ligand pocket, which is a subset of alpha shape. Here, we define a pocket negative image as a set of circumscribed spheres derived from the discrete set of Delaunay tetrahedra and triangles for a pocket.[42] The generation of those circumscribed spheres proceeds by first calculating the orthogonal center for each alpha tetrahedron, using the following linear equations:

$$C = V_0 + A^{-1}B$$

$$A = V_i - V_0$$

$$B = \frac{1}{2}\left(\|V_i - V_0\|^2 + r_i^2 - r_0^2\right)$$

where C is orthogonal center, V_0 and V_i are tetrahedron vertices, and r_0 and r_i are radii of four vertices. Circumscribed spheres are then generated from the orthogonal center, by initially reducing the orthogonal sphere to a small inner sphere and followed by an approximation process. The center tangent spheres are used as circumscribed spheres for pocket triangles. Overlap checking is performed to prevent circumscribed spheres from overlapping with other pocket atoms. We also set a threshold to remove tiny spheres. The remaining circumscribed spheres thus make up the negative image of the ligand pocket. Although step 1.2 and 1.3 are both computationally automated, the refinement of the final set of spheres representing the negative image of the binding site often needs human intervention. Further development to automate these two steps is needed in our future studies.

Methods for extracting binding site negative image have been developed and used in various docking programs. For example, DOCK[16] uses a process that places spheres into the binding site randomly followed by clustering analysis to identify the best set of spheres to represent the shape of the binding site. Surflex[24] uses a different approach to generate so-called "prototype molecules" to represent the binding site shape and other information. Our approach is unique in that it uses α-shape to deterministically detect the binding site atoms, followed by a geometric casting algorithm to generate the negative image as a collection of spheres.

1.4. Implementation of ShapeOnly and Combo Scoring Options in Shape4. Two different scoring options are implemented in Shape4 based on the functions in the commercial OE SHAPE toolkit. The first scoring option is by Shape overlay only, and the second scoring option uses a combined score calculated from the shape overlay score and the chemical matching (or color) score. This step is completely automated as the commercial software ROCS.

In the Shape4:ShapeOnly method, the negative image, derived from the α-shape and nCast programs, is stored in the PDB file format, in which the spheres are stored as dummy atoms with their coordinates placed on the XYZ coordinate fields of the PDB file, and the sphere radii are
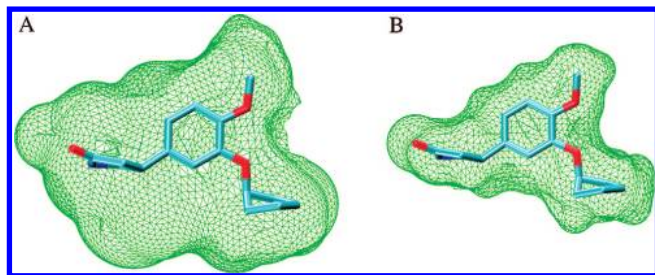
Figure 3. Comparison of the shapes of a structure-based negative image (A) and that of an inhibitor of PDE4B (B).

placed in the occupancy field. This format could easily be recognized by functions in the OE CHEM toolkit, and the shape of the negative image is internally represented as a smooth grid-based function with a grid spacing of 0.5 Å. This grid-based shape representation can be visualized as shown in Figure 3 for the PDE4B binding site with its inhibitor rolapram present. This negative image representation is then used as the query by the Shape4 program to score the 3D shape overlap between the query and the ligands in a 3D molecular database. We found that the Fit Tversky score is preferred to the Tanimoto similarity score because it measures the subshape matching between the fit molecule and the query (which is the binding site image).

In the Shape4:Combo method, the shape overlay between the binding site image and a database molecule is augmented with the matching of detailed pharmacophoric data. We use LigandScout to derive the pharmacophore elements from a target and then add this information to the binding site negative image. Thus, a pseudo molecule is created in which the "atoms" are the dummy atoms representing the shape of the negative image and pharmacophoric atoms are considered according to their respective atom types. This representation is used by the Shape4 program to overlay the negative image and a potential ligand by optimizing both the shape overlay and the chemistry matching. To ensure the computational efficiency of this procedure, the negative image data generated by nCast is further approximated by an even spaced grid, of which each grid point is considered as a dummy atom. The overall score of a particular overlay between the negative image and a database molecule is the weighted sum of the shape Tversky score and the scaled chemical (or color) score.

There are technical differences in the way how shape and pharmacophore matching is done in Shape4 and ROCS. In shape4, the grid representation of the binding site negative image and the pharmcophore constraints are used simultaneously to find the best overlay between the database molecule and the shape pharmacophore query in terms of their shape and color matching. In addition, the radii information for the different spheres making up the negative image is accounted for in the grid approximation. The default approach in ROCS, however, is to assign all the heavy atoms in the reference molecule a single radii value (that for carbon atom). In practice, this approach is good only for organic molecules, but not good for negative image representation of a protein binding site, where the sizes of the spheres vary widely. Thus, it is difficult if not impossible to use the available ROCS version to perform the virtual screening using a combination of active site negative image and externally derived pharmacophores. We note that our Shape4

implementation is greatly facilitated by the functions provided in the Open Eye SHAPE Toolkit.[32]

2. Ligand-based Shape Overlay Program (ROCS). The program ROCS (rapid overlay of chemical structures, OE Scientific) was designed to perform shape-based overlays of conformers of a database molecule to a query molecule in one or more conformations. The overlays can be performed very quickly based on the description of molecules as atom-centered Gaussian functions.[43–45] ROCS maximizes the rigid body overlap of the molecular Gaussian functions and therefore the shared volume between a query molecule and a conformation of a database molecule. Two methods of scoring were employed for this work: they were ROCS: ShapeOnly and ROCS:Combo scoring. The former method scores a database molecule based on the Shape Tanimoto similarity between the query and the database molecule, and the latter calculates a combined score based on Shape Tanimoto similarity and the chemical matching score, where the default chemical force field was employed.

3. Structure-based Molecular Docking Program (FRED). FRED (fast rigid exhaustive docking, OE Scientific) is a protein–ligand docking program, which docks and scores molecules in a multiconformer database against the receptor structure of a biological target. FRED 2.1.2.[46] was used in this work for docking and scoring molecules in the validation database to the binding site of each validation target. The receptor binding site was created from the protein crystal structure and represented as a box defining the extents of the active site. The "-addbox" parameter was set to 4 throughout this work, which extends each side of the binding site box by 4 Angstroms. Even though multiple scoring options exist in FRED, we chose to use only the *ShapeGauss* scoring in this work with the intent to compare the FRED shape scoring method with what is implemented in the Shape4 program.

4. Database Preparation for Virtual Screening Experiments. To validate virtual screening methods, we chose to use the medicinal chemistry database WOMBAT as the reference,[33,47] which contains historic SAR (structure–activity relationship) data about compounds and their biological activities against a variety of targets. This version of the WOMBAT database contains 49 765 unique compounds, in 58 327 chemical entries. These unique molecules were used to create the multiconformer database for virtual screening.

The conformer database was created using OMEGA 2.1 from the SMILES representation of the WOMBAT database using a modified form of the high quality setting (HQS) parameters recommended by Kirchmair.[48] This parameter setting was proved to identify more conformations similar to the known bioactive conformations. We increased the maximum number of allowed conformers from the recommended 500 to 2000. Our own studies indicated that conformer generation with a maximum number of 2000 could enumerate conformers that were much more similar to the experimental ligand conformations in crystal structures when compared to either the default OMEGA setting, or those recommended by Kirchmair.

5. Validation Data Sets. *5.1. Receptor Structures.* In order to analyze the performance of our structure-based shape pharmacophore approach (Shape4) in the context of existing methods, five protein–ligand cocrystal structures were selected for this study (Table I). The structures and their

Pharmacophore Search Using Matching Techniques

J. Chem. Inf. Model., Vol. 48, No. 4, 2008 **893**

Table I. Selected Biological Targets for Method Validation

target	PDB code	family	super family
HIV-1 protease	1PRO	retroviral protease	acid proteases
phosphodiesterase 4B	1RO6	PDEase	HD-domain/PDEase-like
estrogen receptor alpha	3ERT	nuclear receptor–ligand-binding domain	nuclear receptor–ligand-binding domain
HIV-1 reverse transcriptase	2HNY	reverse transcriptase	DNA/RNA polymerases
thymidine kinase	1KIM, 1KI5	nucleotide and nucleoside kinases	P-loop containing nucleoside triphosphate hydrolases

Table II. Profile of Selected Active Compounds from WOMBAT

targets	total number of ligands	molecular weight range	IC$_{50}$
HIV-1 protease	1891	159–1311	3.14–10.09
phosphodiesterase 4B	61	261–454	5.77–9.00
estrogen receptor alpha	64	272–523	4.72–9.69
HIV-1 reverse transcriptase	948	226–1297	3.22–9.22
thymidine kinase	51	227–411	3.31–6.82

corresponding PDB codes are phosphodiesterase 4B (PDE4B, 1ro6), HIV-reverse transcriptase (HIV-RT, 2hny), HIV-protease (1pro), ERα (2ayr), and thymidine kinase (1ki5, 1kim). Crystal water molecules as well as the ligand molecules were removed from the structures before use, since the water molecules are not involved in interesting interactions in these test cases. Metal ions that are necessary for the protein activity were preserved before the structures were used in calculations as they also contribute to the shape of the binding sites.

5.2. Ligands. Five sets of molecules were extracted from the WOMBAT database and used as the basis to validate the virtual screening results in this work. These include inhibitors/ligands for the following targets: PDE4B, estrogen receptor (ERα), HIV-reverse transcriptase (HIV-RT), HIV-protease (HIV-PR), and thymidine kinase (TK). Most of the chemical structures from the WOMBAT database have target relating keywords in the SD File. These keywords were used in identifying the ligands/inhibitors for the selected targets. For instance, molecular entries with the keyword PDE4B under target.name or target.full.name fields were considered inhibitors of the target PDE4B. Ligands with activity toward the other targets considered in this study were selected in a similar manner. The numbers of ligands found for each target are as follows: PDE4B (61), HIV-RT (948), HIV-PR (1891), ERα (64), and thymidine kinase (51) (see Table II).

6. Virtual Screening Experiments. Each of the computational methods was used to search the multiconformer WOMBAT database generated using the Omega[36] conformational analysis program. For each method, the database molecules were scored and ranked according to their scores given by that method. The sorted lists of molecules were the results of virtual screening experiments and were further analyzed to obtain the enrichment data as described below.

6.1. Virtual Screening with FRED:ShapeGauss Scoring. For each validation target, FRED was used to search the multiconformer WOMBAT database. We chose to do the virtual screening experiments with FRED in order to compare the effectiveness of FRED:ShapeGauss option and the shape method implemented in the Shape4 program. In each case, virtual screening with FRED generates a list of molecules that are sorted according to their ShapeGauss scores against the target structure. This sorted list is then used to calculate the enrichment curve for FRED.

6.2. Virtual Screening with Ligand-Based Shape Similarity Method. One of the goals of this work was to test if the structure-based shape method implemented in Shape4 could capture more information than the ligand-based shape similarity methods. Thus, for each of the five validation cases, we performed virtual screening experiments using the ligand structures extracted from the X-ray structures of the targets as the queries for ROCS database search. Database compounds were then ranked by their shape similarity scores generated by ROCS:ShapeOnly method. The sorted list of compounds is analyzed to generate the enrichment curve for each target.

6.3. Virtual Screening with Ligand-Based ROCS:Combo Scoring. To compare the efficiency of virtual screening using the ligand-based ROCS method and the structure-based pharmacophore method Shape4, we conducted virtual screening experiments using ROCS with its combo scoring option. For each of the five validation cases, database compounds were scored and sorted according to the ROCS:Combo scores (which is a weighted sum of both the shape Tanimoto similarity and the scaled chemical matching scores). Enrichment curve is then calculated for each target.

6.4. Virtual Screening with Shape4 Methods. This was the development focus of this work. To validate these methods and demonstrate their efficiency, we conducted virtual screening experiments using both the Shape4:ShapeOnly and Shape4:Combo methods. For each of the five targets, the corresponding negative image derived from each target structure was used by Shape4 as the query for the virtual screening experiments. Database compounds are ranked by either the shape Fit Tversky score alone or the combo score of Tversky similarity and the chemical (color) score. For the Combo score, the pharmacophore elements are derived by the LigandScout program and incorporated into Shape4.

7. Calculation of Enrichment Curves. The virtual screening results obtained as described above are analyzed in terms of the percentage of the WOMBAT database virtually screened versus the percentage of known active compounds retrieved by a method. In the calculation, all known ligands/inhibitors for a particular target are considered as hits for that target if they appear at a said percentage of the database screened. Other compounds that are not labeled as ligands for that target are considered "decoys", i.e. they are considered to be inactive against that target under study. A typical enrichment curve has its *X*-axis representing the percentage of database molecule screened and the *Y*-axis the percentage of known active compounds recovered by a method. In theory, a random screening would give average hit rates along the diagonal of the enrichment plot.

8. Calculation of the Chemical Diversity Plots. To show the diversity of a set of molecules, we calculate the pairwise Tanimoto similarity values among the set of molecules, using the MOE (molecular operating environment) software with

MACCS key fingerprint as the descriptors.[31] Then the histogram, showing the frequency distribution of various bins of similarity values, is generated by counting and then normalizing the number of occurrences in each similarity bin. This calculation is used to characterize the diversity of any given set of molecules. A distribution shifted toward the left (smaller similarity) is more diverse than a one shifted toward the right (bigger similarity).

9. Calculation and Presentation of the Trellis Plot. In order to summarize the performance of multiple computational methods across multiple test cases, and across different relevant screening percentages, a Trellis plot is generated. This plot contains a matrix of small bar graphs, where each row corresponds to a certain screening percentage (0.1%, 0.3%, 1.0%, 3.0%, and 10.0%) and each column corresponds to a test case. Within each column, there are multiple bars, indicating the retrieval rate of known active compounds by different virtual screening protocols as well as the ideal retrieval rate. This graph would show at a single glance which methods worked best on which targets across the practical screening ranges.

RESULTS

As described in Materials and Methods, we aimed to demonstrate the effectiveness of our structure-based shape pharmacophore method (Shape4) in virtual screening experiments in terms of its ability to find known active compounds from a compound database. Although the ultimate validation of any computational screening method would come from experimental testing of the virtual screening hits, we had adopted a standard computational protocol for validating new virtual screening methods. In this protocol, we used a medicinal chemistry database (WOMBAT) with known biological data to test how well a method can recover the known active compounds for a given target. Since the experimental data in the WOMBAT database were NOT used in building the Shape4 models (note: this is different from QSAR methods where multiple known active compounds are used to build the statistical models), these data can be considered as pseudo experimental testing of the computational hits. To this end, five validation sets were extracted from the WOMBAT database. These active molecules were mixed into the WOMBAT database, and a computational method was then used to virtually screen the database to see how well a method recover known active compounds for a given target. For the five validation cases, five sets of enrichment curves were obtained and shown in Figures 4–8. One additional figure (Figure 9) of enrichment curves was also obtained for one of the targets, where the substrate–protein structure was used as the starting point as oppose to the inhibitor–protein structure. The chemical diversity plots (cf. Materials and Methods 8) of the validation sets of molecules are shown in Figure 10. For each validation case, five enrichment curves were calculated that correspond to five different virtual screening methods used in this work. In Figures 4–9, the enrichment curves obtained with FRED are labeled as diamonds; the enrichment curves resulting from the ROCS:ShapeOnly method are labeled as triangles; the enrichment curves resulting from the ROCS:combo method are represented as solid curves. The structures of the ligands used in each ROCS similarity search are shown in Figure 11. Finally, the enrichment curves obtained with
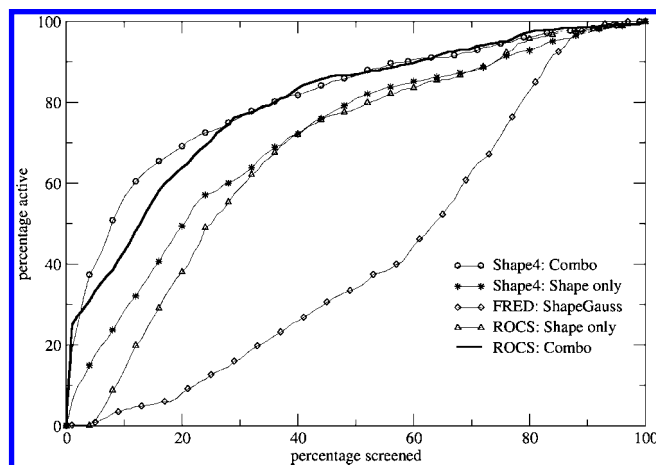

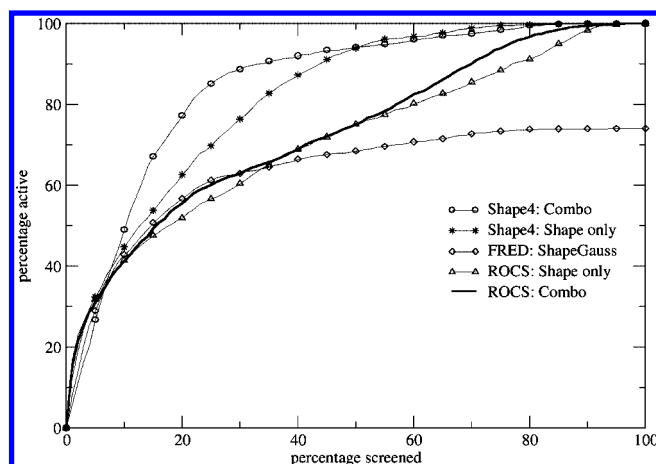**Figure 4.** Enrichment plot for reverse transcriptase virtual screening.

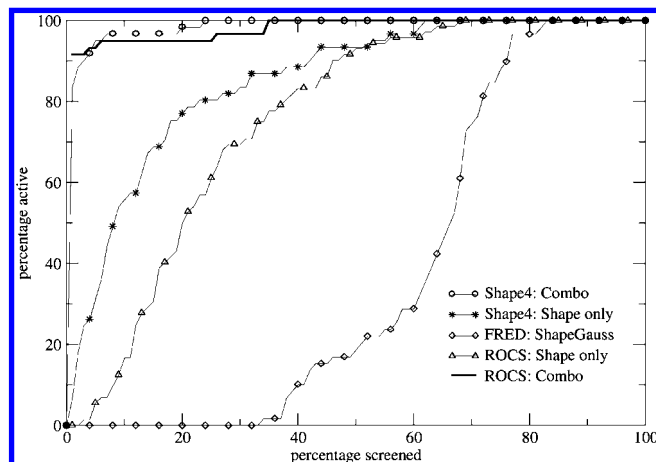**Figure 5.** Enrichment plot for HIV protease virtual screening.

Figure 6. Enrichment plot for PDE4B virtual screening.

the Shape4 method are labeled as asterisks and circles for the ShapeOnly and Combo scoring options, respectively.

In the following test cases, the crystal structure of each protein target was used to create the active site description for the FRED calculations and generate the negative image for Shape4 calculations. The structure of the bound inhibitor or substrate for each target was used as the query for ROCS calculations.

1. Reverse Transcriptase Case Study (Figure 4). The crystal structure of HIV-RT was used for this experiment.
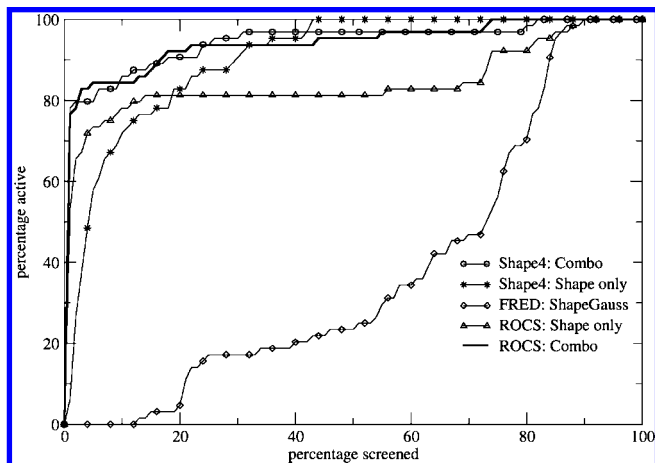
PHARMACOPHORE SEARCH USING MATCHING TECHNIQUES

J. Chem. Inf. Model., Vol. 48, No. 4, 2008 **895**

Figure 7. Enrichment plot for estrogen receptor (ER) virtual screening.

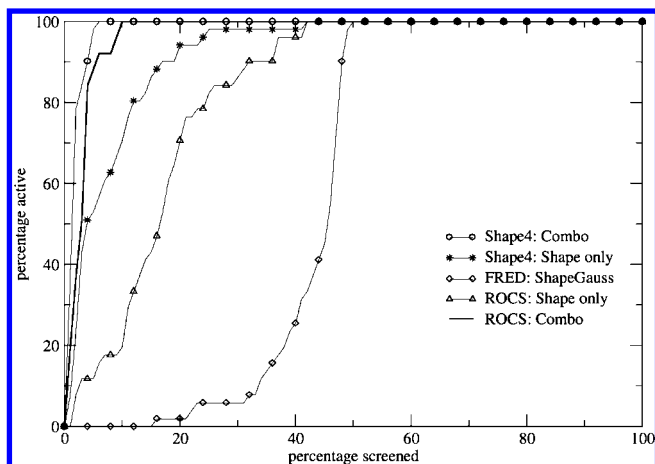**Figure 8.** Enrichment plot for thymidine kinase (1ki5) virtual screening.

Figure 9. Enrichment plot for thymidine kinase (1KIM) virtual screening.

The diversity of the HIV-RT validation set is shown in Figure 10a as a histogram of pairwise similarity values calculated with MOE as MACCS key based Tanimoto similarity (cf. Materials and Methods 8).[31] As indicated by the diversity plot, this set of molecules is chemically diverse compared to other sets used in this work. In this case, the FRED: ShapeGauss method did not perform well as indicated by the enrichment curve even below random. In fact it is true that in almost all the cases reported in this work, the FRED: ShapeGauss method did not perform well compared to other methods. A possible explanation for this is that the Shape-Gauss scoring was originally designed as a filter in the FRED docking process where it provides molecules for further screening and scoring. The results reported in this paper do not suggest that FRED as a docking method is not a good virtual screening tool; rather, it cautions that one needs to be judicious about the use of FRED:ShapeGauss alone as a virtual screening tool. In contrast, screening with ROCS: ShapeOnly performed fairly well. It recovered roughly 20% of the actives by screening 10% of the database and 80% of the actives by screening 50% of the database. Virtual screening with the Shape4:ShapeOnly method obtained roughly 28% of the actives at 10% of the database, and 80% of actives by screening 50% of the database. It appeared that Shape4:ShapeOnly performed better than ROCS:ShapeOnly up to screening 30% of the database. Both the Shape4:Combo and ROCS:Combo methods performed much better, recover-

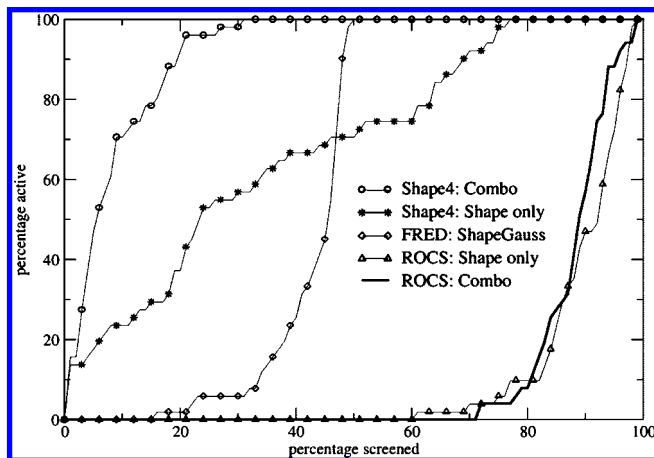ing over 50% and 40% of the actives, respectively, by screening only 10% of the database compounds. Both methods recovered nearly 90% of the actives by screening 50% of the database compounds. Overall, the Shape4:Combo method performed the best among all the methods.

2. Protease Case Study (Figure 5). The crystal structure of HIV-PR (1pro) was used for this experiment. This validation set has 1891 molecules and its chemical diversity is shown in Figure 10b as a histogram of pairwise fingerprint Tanimoto similarity values (cf. Materials and Methods 8). In this case, all five methods performed similarly well by screening the first 10% of the database, with all methods recovering between 40 and 50% of the actives. Shape4: Combo again performed the best overall among the methods. The performance of the FRED:ShapeGauss method reached plateau (about 70% of known actives) after screening 40–50% of the database, while the ROCS:ShapeOnly and ROCS:Combo methods performed equally well throughout the virtual screening process. Both the Shape4:ShapeOnly and Shape4:Combo methods performed well. The reason for the superior performance of Shape4 in this case was probably due to the fact that there were more chemically diverse compounds in this validation set, which can be observed from the diversity plot (Figure 10b). We also note that the molecules found by Shape4 methods were more diverse compared to the query molecule used by ROCS. In addition, Shape4:Combo performed significantly better than Shape4: ShapeOnly, indicating that the use of chemical constraints did further weed out false positives leading to the enrichment of true actives in the top ranking compounds.

3. PDE4B Case Study (Figure 6). The crystal structure of PDE4B (1ro6) was used in this experiment. The validation set has 61 molecules and its chemical diversity is shown in Figure 10c. Similar trends have been observed for this case in that Shape4:ShapeOnly performed better than ROCS: ShapeOnly. By screening 10% of the database, both Shape4: Combo and ROCS:Combo identified about 95% of the known actives. Some example structures of the computational hits obtained by Shape4 and ROCS are shown in Figure 12.

4. Estrogen Receptor Case Study (Figure 7). In the case of estrogen receptor (ER), the crystal structure of ERα (3ert) was used as the basis for all experiments. The validation set has 64 molecules and its diversity plot is shown in Figure 10d. All methods (except FRED:ShapeGauss) performed
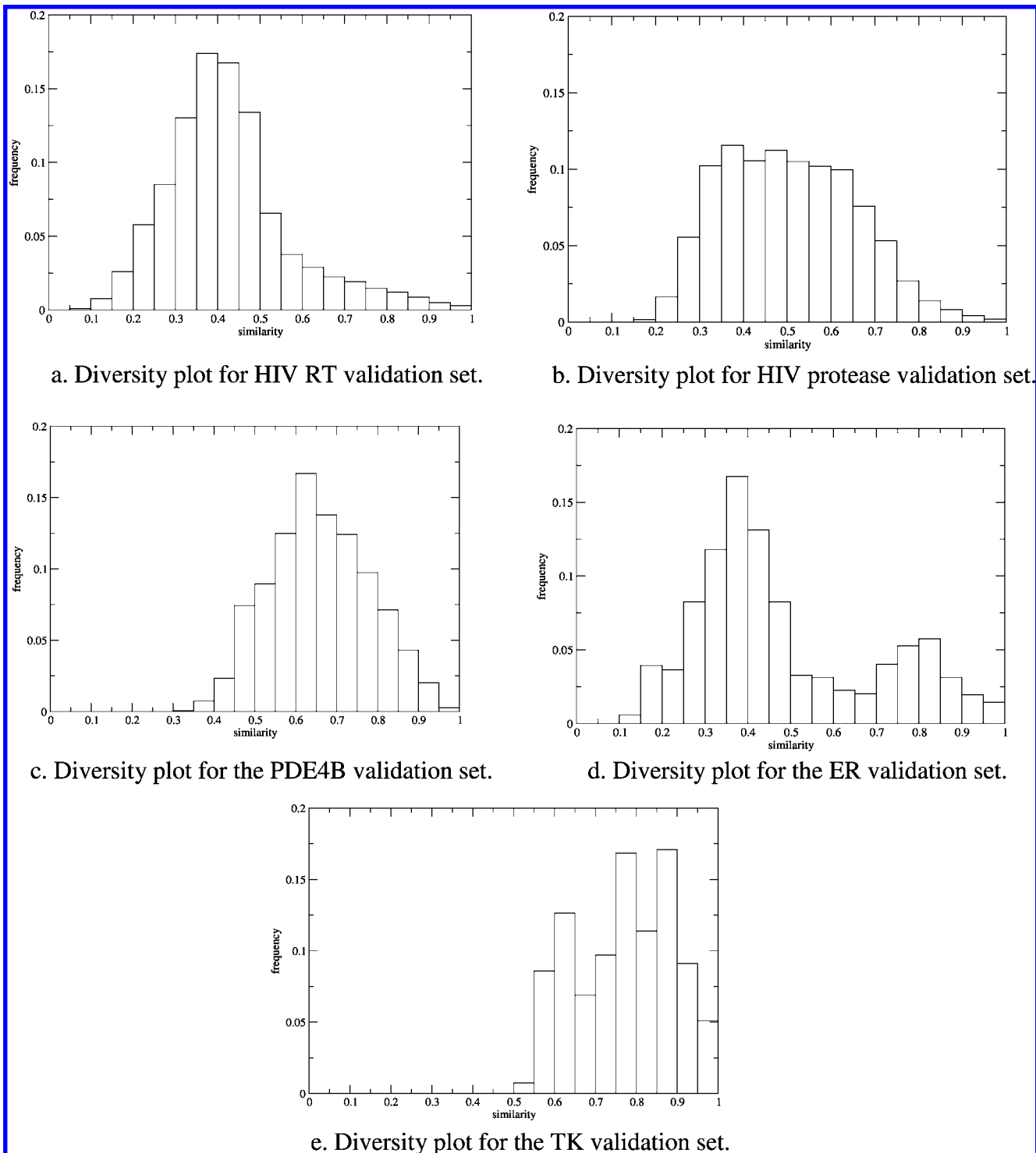
a. Diversity plot for HIV RT validation set.

b. Diversity plot for HIV protease validation set.

c. Diversity plot for the PDE4B validation set.

d. Diversity plot for the ER validation set.

e. Diversity plot for the TK validation set.

Figure 10. Diversity of the ligands in the validation sets.

extremely well reaching above 60–85% of the actives by screening the first 10% of database compounds, probably due to the fact that ER binding pocket is fairly hydrophobic and well-defined, and the shape plays a critical, if not determinant, role in selecting ER ligands. However, some intricate details are worth noting. For example, ROCS: ShapeOnly performed better than Shape4:ShapeOnly in the first 10% of the database screened, but ROCS seemed to have reached a plateau after screening 18% of the database. Meanwhile, Shape4:ShapeOnly continued to pick up ER ligands and reached 100% recovery of active compounds at 45% of the database screened. This is probably due to the

fact that ROCS is ligand-based method, picking up chemically similar molecules very quickly but missing out more chemically diverse molecules. This point can be observed from the diversity plot for this set of ER ligands, where a group of molecules seemed to be very similar while others are quite different in that the histogram almost displays a bimodel distribution. On the other hand, the structure-based shape method, Shape4, tends to provide more space in the binding pocket to capture more chemically diverse classes of molecules (see also discussions on diversity).

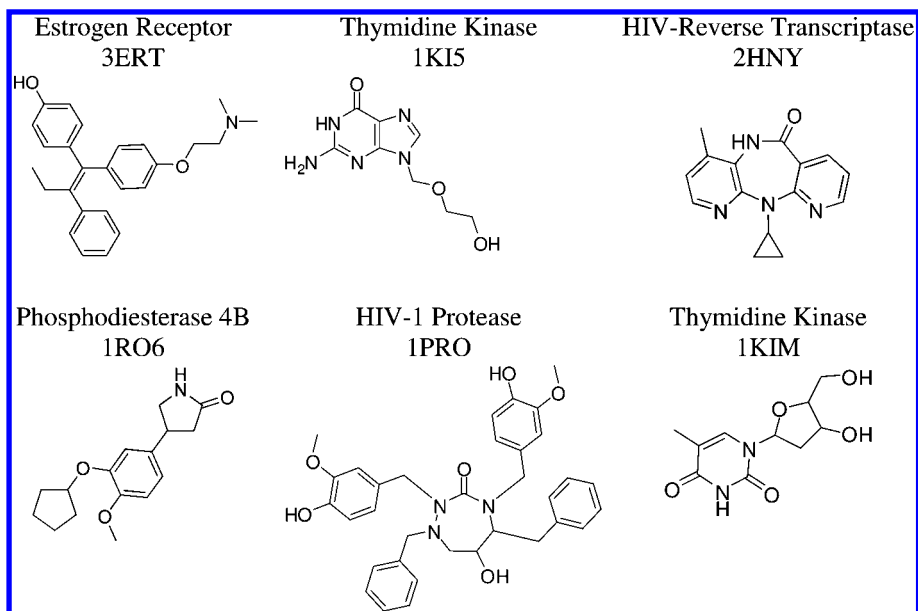**5. Thymidine Kinase Case Study (Figure 8).** Crystal structure 1ki5 was used in this experiment. The validation

Figure 11. Structures of the query molecules used for ROCS-based virtual screening.
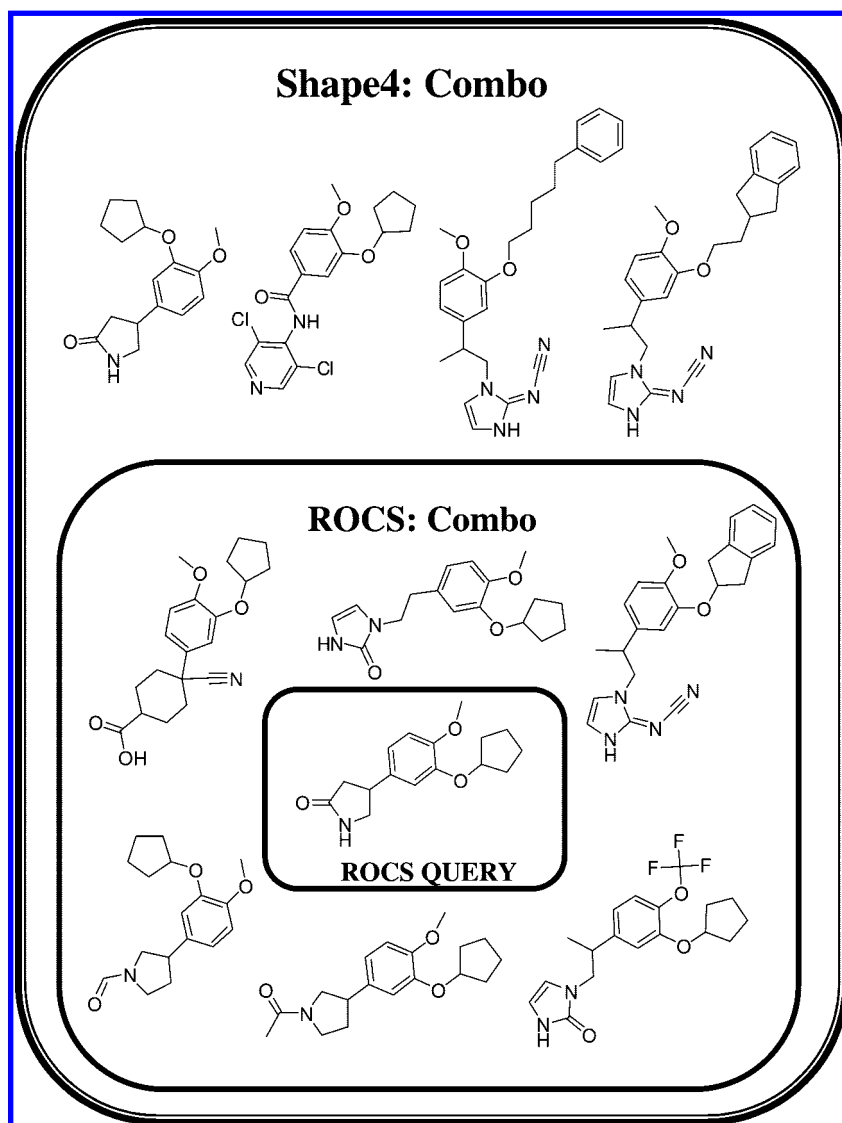

Figure 12. Venn diagram of sample actives recovered after screening 20% of the database for PDE4B inhibitors using the ROCS:Combo and Shape4:Combo methods.
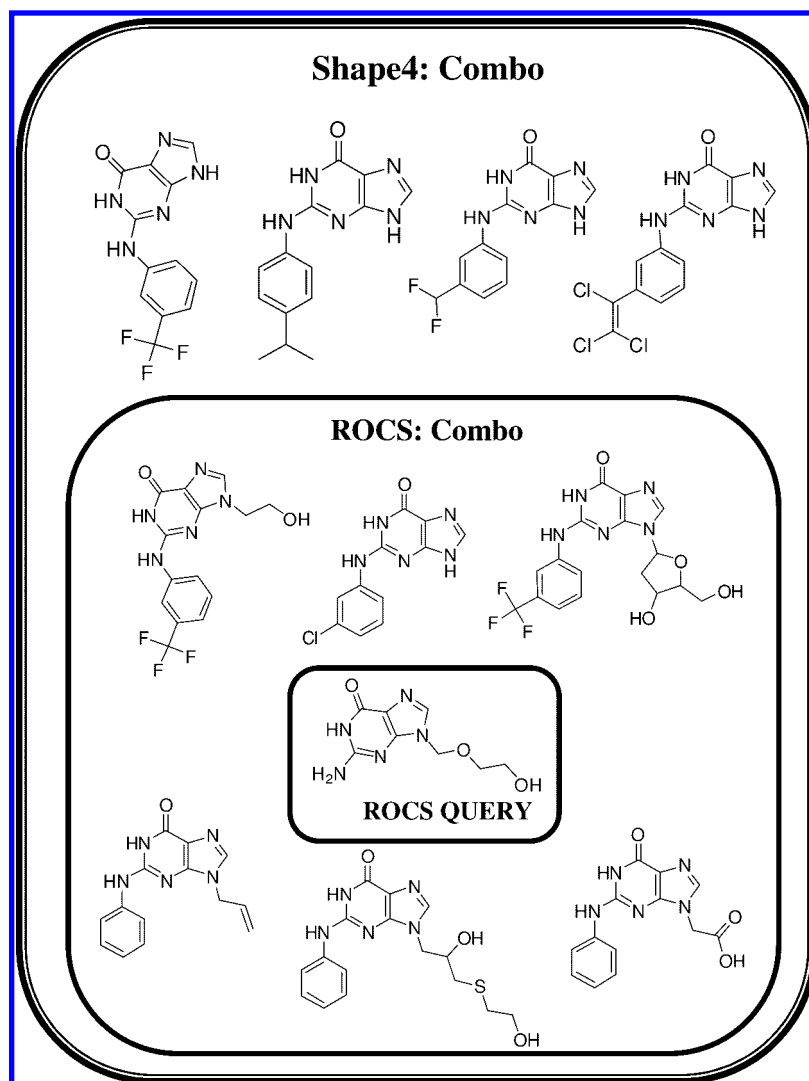
Figure 13. Venn diagram of sample actives recovered after screening 5% of the database. for thymidine kinase inhibitors using ROCS: Combo and Shape4:Combo methods.

set has 51 molecules and its diversity plot is shown in Figure 10e. ROCS:ShapeOnly picked up 20% of the actives by screening only 10% of the database, and nearly 60% of the actives when 20% of the database was screened. Shape4: ShapeOnly performed better than corresponding ROCS method, recovering over 60% of the actives at 10% of the database screened. Both Shape4:Combo and ROCS:Combo recovered 100% of the actives by screening less than 10% of the database. As indicated by the diversity plot, this set of validation molecules is not very diverse, which may explain that all methods performed well. Some example structures of the computational hits obtained by Shape4 and ROCS are shown in Figure 13, which indicates that more different structures were found by Shape4 than by ROCS, even in this chemically similar validation set.

6. Thymidine Kinase Case Study 2 (Figure 9). The last case shown here covered a scenario where there was no ligand-protein complex structure and only the substrate–protein structure was available. In this case, ROCS methods did not perform well due to the fact that there were no known inhibitors that were similar to the substrate. However, the Shape4 methods, which rely only on the binding site image derived from the protein structure, performed very well. This further proved the value of developing and applying the structure-based shape pharmacophore methods as implemented in Shape4 in cases where little is known about the ligands/inhibitors of a target protein.

DISCUSSION

In all test cases, Shape4:Combo method performed extremely well in terms of the enrichment of active compounds in virtual screening experiments. In many cases, it was the most effective method. We note that it is much more effective to use the Fit Tversky similarity measure than Tanimoto similarity measure when comparing the ligand shape to the shape of the binding site negative image, because the binding site shape often is bigger than that of a ligand, and that calculating Tanimoto coefficient would compare the shape of a ligand and that of the whole binding pocket. Fit Tversky measure, on the other hand, allows unmatched space in the binding site shape thus capturing diverse set of molecules that may bind to the receptor in somewhat different binding modes.

ROCS:Combo performed significantly better than ROCS: ShapeOnly method, indicating that chemical constraints did afford better true positive rate by reducing false positives. Similarly, Shape4:Combo further enriched the hits compared
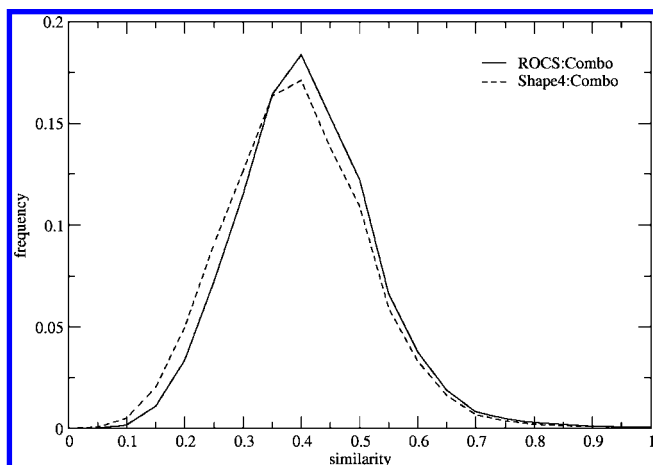
Figure 14. Distribution of pairwise Tanimoto similarity values for top 10% of WOMBAT database compounds screened using ROCS and Shape4. The distribution is the average for the five targets screened.

to Shape4:ShapeOnly method. In most cases, Shape4:Combo performed similarly or better than the ROCS:Combo method, indicating that the structure-based method did afford the ability to discover more diverse classes of compounds. Thus, the fact that the structure-based shape methods performed almost always better than the corresponding ROCS methods supports the use of the structure-based pharmacophore methods in the hope to find novel series of compounds for a target under study.

It is speculated that the Shape4:Combo method may have the advantage over the ROCS:Combo method in finding more diverse set of hits, because Shape4 uses the binding site shape from the empty binding pocket as the query and Tversky measure for similarity, whereas ROCS uses the bound ligand conformation as the query and Tanimoto as similarity measure. To demonstrate this point, we have calculated the chemical diversity plots (cf. Materials and Methods 8) for

the top 10% of the hits obtained with Shape4 and ROCS, respectively (Figure 14). These plots show the histogram distribution of the pairwise similarity values, where the similarity values have been averaged over the results obtained for the five test targets. As can be seen, the diversity obtained with the Shape4 method is better (though only slightly) than that obtained with ROCS. This is indeed a potential advantage afforded by the Shape4 method, which uses the shape of the empty binding pocket as the virtual screening query.

There has been great interest in understanding the early enrichment behavior of computational virtual screening methods. In order to summarize the performance of five computational methods across five test cases, and across different relevant screening percentages, we generated a Trellis plot (Figure 15). This plot contains a matrix of small bar graphs, where each row corresponds to a certain screening percentage (0.1%, 0.3%, 1.0%, 3.0%, and up to 10.0%) and each column corresponds to a test case. Within each column, there are six bars, indicating the retrieval rate of known active compounds by five virtual screening protocols as well as the ideal retrieval rate in each case. This graph clearly demonstrated that both the Shape4:Combo and ROCS:Combo methods performed equally well in early enrichment behavior, from 0.1% to 10%. In some cases, Shape4:Combo performed better, while in other cases ROCS: Combo performed better. In most cases, Shape4:ShapeOnly performed better than ROCS:ShapeOnly.

CONCLUSIONS AND FUTURE DEVELOPMENT

We have developed a new method that uses α-shape analysis to detect the binding site atoms and to construct the negative image of a target binding site. The so-derived negative image is then represented as Gaussian functions using the Shape toolkit (from Open Eye Scientific) and fast shape overlays between the negative image and database molecules can be performed. Such an implementation
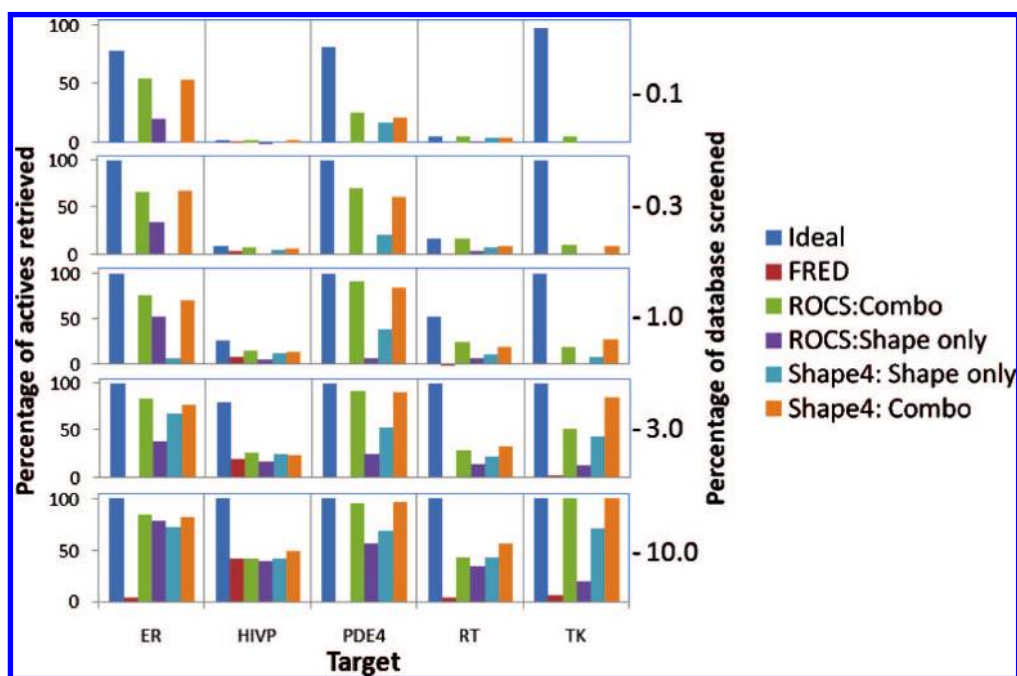

Figure 15. Trellis plot illustrating the performance of Shape4, ROCS, and FRED against five targets: estrogen receptor (ER), HIV protease (HIVP), HIV reverse transcriptase (RT), phosphodiesterase 4B (PDE4), thymidine kinase (TK). PDB codes for the virtual screening targets are 3ERT (ER), 1PRO (HIVP), 2HNY (RT), and 1KI5 (TK).

captures the intricate details of a binding site shape, the effect of which during virtual screening experiments has been demonstrated by five test cases. Using Shape4, a variant of ROCS, one can yield comparable enrichments to ROCS on a real inhibitor at <10% of the database screened. Likewise, screening 20% of the database, except for HIV protease, Shape4 worked better than ROCS. For thymidine kinase, Shape4 worked dramatically better at retrieving known inhibitors than did ROCS using a bound substrate, rather than an inhibitor. Including pharmacophoric features (color) helps the enrichment both for ROCS and Shape4. However, while the enrichment is not much improved by screening <10% of the database, the diversity of the computational hits obtained with Shape4 is better (Figure 14). Together, the ability to use the Shape4 method without a known inhibitor, and the additional diversity of the computational hits, constitute a substantial advantage for drug discovery programs.

Thus, Shape4 offers a fast, effective and intuitive virtual screening alternative in cases where the X-ray crystal structure of the target is known without performing computationally more expensive docking calculations. Shape4 is not meant to be a replacement of docking tools; rather, it is an alternative tool and can be efficient when combined with experimental screening efforts. Shape4 can also be used in conjunction with more sophisticated docking tools as their prescreening filter prior to a more detailed docking calculation (as is done by Miteva et al.).[49] The effectiveness of a Shape4-Docking combination as a virtual screening tool will be examined in our future studies.

Since Shape4 can generate reasonable alignments of a set of molecules from their multiconformer database, one may use it to create structure-based alignment for 3D-QSAR analysis of active compounds. Such an analysis could provide additional insights as to which region of the binding site shape plays an important role and which pharmacophore elements help discriminate active compounds against inactive ones. This "learned" pharmacophore knowledge can then be used by Shape4 as a refined or optimized shape pharmacophore query in further virtual screening experiments. Such an iterative approach may prove effective in lead optimization stage of a discovery project.

Additional applications of the Shape4 method may be its use in the assessment of *biologically relevant diversity* of compound collections. Because of the efficiency to compare small molecule shapes and the binding site shapes, we could use this method to evaluate the fitness of a collection of molecules to a reference panel of protein binding site shapes. For example, we can calculate the scores of compounds in a collection (e.g., by a chemical vendor) and evaluate the coverage of biological shapes derived from the whole genome structures. If a collection of compounds covers more families of target shapes, we can say that the collection is likely to be more diverse and biologically more relevant. We are actively pursuing this aspect of the application and will report the results elsewhere.

REFERENCES AND NOTES

(1) DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The price of innovation: new estimates of drug development costs. *J. Health Econ.* **2003**, *22*, 151–85.

(2) Rawlins, M. D. Cutting the cost of drug development. *Nat. Rev. Drug Discov.* **2004**, *3*, 360–4.

(3) Dickson, M.; Gagnon, J. P. Key factors in the rising cost of new drug discovery and development. *Nat. Rev. Drug Discov.* **2004**, *3*, 417–29.

(4) Sittampalam, G. S.; Kahl, S. D.; Janzen, W. P. High-throughput screening: advances in assay technologies. *Curr. Opin. Chem. Biol.* **1997**, *1*, 384–91.

(5) Nefzi, A.; Dooley, C.; Ostresh, J. M.; Houghten, R. A. Combinatorial chemistry: from peptides and peptidomimetics to small organic and heterocyclic compounds. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2273–8.

(6) Blundell, T. L.; Jhoti, H.; Abell, C. High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov.* **2002**, *1*, 45–54.

(7) Joseph-McCarthy, D.; Baber, J. C.; Feyfant, E.; Thompson, D. C.; Humblet, C. Lead optimization via high-throughput molecular docking. *Curr. Opin. Drug Discov. Dev.* **2007**, *10*, 264–74.

(8) Olsson, T.; Oprea, T. I. Cheminformatics: a tool for decision-makers in drug discovery. *Curr. Opin. Drug Discov. Dev.* **2001**, *4*, 308–13.

(9) Stockwell, B. R. Frontiers in chemical genetics. *Trends Biotechnol.* **2000**, *18*, 449–55.

(10) Khedkar, S. A.; Malde, A. K.; Coutinho, E. C.; Srivastava, S. Pharmacophore modeling in drug discovery and development: an overview. *Med. Chem.* **2007**, *3*, 187–97.

(11) Brown, N.; McKay, B.; Gasteiger, J. A novel workflow for the inverse QSPR problem using multiobjective optimization. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 333–41.

(12) Esposito, E. X.; Hopfinger, A. J.; Madura, J. D. Methods for applying the quantitative structure-activity relationship paradigm. *Methods Mol. Biol.* **2004**, *275*, 131–214.

(13) Shen, M.; Beguin, C.; Golbraikh, A.; Stables, J. P.; Kohn, H.; Tropsha, A. Application of predictive QSAR models to database mining: identification and experimental validation of novel anticonvulsant compounds. *J. Med. Chem.* **2004**, *47*, 2356–64.

(14) Rolland, C.; Gozalbes, R.; Nicolai, E.; Paugam, M. F.; Coussy, L.; Barbosa, F.; Horvath, D.; Revah, F. G-protein-coupled receptor affinity prediction based on the use of a profiling dataset: QSAR design, synthesis, and experimental validation. *J. Med. Chem.* **2005**, *48*, 6563–74.

(15) van Rhee, A. M. Use of recursion forests in the sequential screening process: consensus selection by multiple recursion trees. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 941–8.

(16) Shoichet, B. K.; Kuntz, I. D. Matching chemistry and shape in molecular docking. *Protein Eng.* **1993**, *6*, 723–32.

(17) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68*, 76–90.

(18) Jones, G.; Willett, P.; Glen, R. C. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532–49.

(19) Cross, S. S. Improved FlexX docking using FlexS-determined base fragment placement. *J. Chem. Inf. Model.* **2005**, *45*, 993–1001.

(20) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–49.

(21) Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. Development and validation of a modular, extensible docking program: DOCK 5. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 601–19.

(22) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293–304.

(23) Totrov, M.; Abagyan, R. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins* **1997**, *1*, 215–20.

(24) Jain, A. N. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 281–306.

(25) Coupez, B.; Lewis, R. A. Docking and scoring--theoretically easy, practically impossible. *Curr. Med. Chem.* **2006**, *13*, 2995–3003.

(26) Wu, G.; Vieth, M. SDOCKER: a method utilizing existing X-ray structures to improve docking accuracy. *J. Med. Chem.* **2004**, *47*, 3142–8.

(27) Joseph-McCarthy, D.; Thomas, B. E. t.; Belmarsh, M.; Moustakas, D.; Alvarez, J. C. Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins* **2003**, *51*, 172–88.

(28) Steindl, T. M.; Schuster, D.; Laggner, C.; Langer, T. Parallel screening: a novel concept in pharmacophore modeling and virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 2146–57.

(29) Krovat, E. M.; Fruhwirth, K. H.; Langer, T. Pharmacophore identification, in silico screening, and virtual library design for inhibitors of the human factor Xa. *J. Chem. Inf. Model.* **2005**, *45*, 146–59.

(30) Guner, O.; 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.

(31) *MOE*; Chemical Computing Group: Montreal, Quebec, Canada, 2006.

(32) *OEShape Toolkit*; OpenEye Scientific Software: Santa Fe, NM, 2006.

(33) Olah, M.; Mracec, M.; Ostopovici, L.; Rad, R.; Bora, A.; Hadaruga, N.; Olah, I.; Banda, M.; Simon, Z.; Mracec, M.; Oprea, T. I. WOMBAT: World of Molecular Bioactivity. In *Cheminformatics in Drug Discovery*; Oprea, T. I., Ed.; Wiley-VCH: New York, 2004; pp 223–239.

(34) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74–82.

(35) Liang, J.; Edelsbrunner, H.; Woodward, C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **1998**, *7*, 1884–97.

(36) *OMEGA*; OpenEye Scientific Software: Santa Fe, NM, 2006.

(37) Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2005**, *45*, 1609.

(38) Tseng, Y. Y.; Liang, J. Predicting enzyme functional surfaces and locating key residues automatically from structures. *Ann. Biomed. Eng.* **2007**, *35*, 1037–42.

(39) Binkowski, T. A.; Naghibzadeh, S.; Liang, J. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res.* **2003**, *31*, 3352–5.

(40) Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **2006**, *34*, W116–8.

(41) Delaunay, B. Sur la sphère vide. *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk* **1934**, *7*, 793–800.

(42) Edelsbrunner, H.; Facello, M.; Liang, J. On the definition and the construction of pockets in macromolecules. *Pac. Symp. Biocomput.* **1996**, 272–87.

(43) *ROCS*; OpenEye Scientific Software: Santa Fe, NM, 2006.

(44) Grant, J. A.; M., A. G. B. T. P. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1996**, *17*, 1653–1666.

(45) Rush, T. S., 3rd; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **2005**, *48*, 1489–95.

(46) *FRED*; OpenEye Scientific Software: Santa Fe, NM, 2006.

(47) *WOMBAT*; Sunset Molecular Discovery LLC: Santa Fe, NM, 2005.

(48) Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T. Comparative performance assessment of the conformational model generators omega and catalyst: a large-scale survey on the retrieval of protein-bound ligand conformations. *J. Chem. Inf. Model.* **2006**, *46*, 1848–61.

(49) Miteva, M. A.; Lee, W. H.; Montes, M. O.; Villoutreix, B. O. Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. *J. Med. Chem.* **2005**, *48*, 6012–22.