JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

SHAFTS: A Hybrid Approach for 3D Molecular Similarity Calculation. 1. Method and Assessment of Virtual Screening

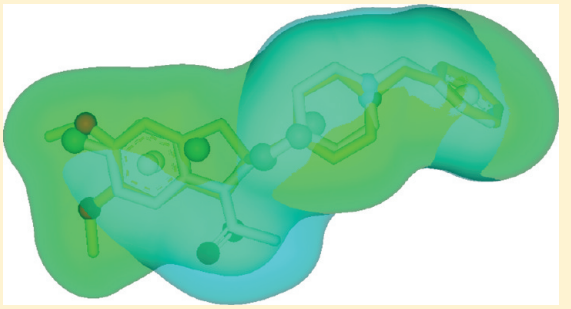
Xiaofeng Liu,[†,‡] Hualiang Jiang,[‡] and Honglin Li*[,†]

[†]State Key Laboratory of Bioreactor Engineering, Shanghai Key Laboratory of Chemical Biology, School of Pharmacy, East China University of Science and Technology, 130 Mei Long Road, Shanghai 200237, China

[‡]Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

Supporting Information

ABSTRACT: We developed a novel approach called SHAFTS (SHApe-FeaTure Similarity) for 3D molecular similarity calculation and ligand-based virtual screening. SHAFTS adopts a hybrid similarity metric combined with molecular shape and colored (labeled) chemistry groups annotated by pharmacophore features for 3D similarity calculation and ranking, which is designed to integrate the strength of pharmacophore matching and volumetric overlay approaches. A feature triplet hashing method is used for fast molecular alignment poses enumeration, and the optimal superposition between the target and the query molecules can be prioritized by calculating corresponding "hybrid similarities". SHAFTS is suitable for large-scale virtual screening with single or multiple bioactive compounds as the query "templates" regardless of whether corresponding experimentally determined conformations are available. Two public test sets (DUD and Jain's sets) including active and decoy molecules from a panel of useful drug targets were adopted to evaluate the virtual screening performance. SHAFTS outperformed several other widely used virtual screening methods in terms of enrichment of known active compounds as well as novel chemotypes, thereby indicating its robustness in hit compounds identification and potential of scaffold hopping in virtual screening.



INTRODUCTION

Molecular similarity is a pivotal concept in drug discovery and the cornerstone of structure−activity relationship (SAR) and structural clustering analysis. The rationale which runs as a red line through the process of drug design behind this approach is that structurally similar molecules should exhibit the same (or similar) bioactivities, which is also called *Similar Property Principle*.[1] Based on this entrenched assumption, if a bioactive query structure is searched for, the top-ranked similar molecules are likely to possess the same activity profiles and may be regarded as the prime candidates for pharmacological tests. Although it is an arbitrary hypothesis and challenged recently,[2] similarity-based database searching and candidates ranking remains to be one of the most powerful tools in the medicinal chemists' toolkit[3,4] and has proven successful in a number of cases of ligand-based virtual screening,[5−9] even in exploring pharmacological effects via chemical-protein relationship.[10,11]

The exclusive preliminary of effective similarity searching is the employment of appropriate similarity measurement methods. Similarity searching programs can be categorized into two classes in general according to whether three-dimensional conformational information is considered: namely 2D similarity and 3D similarity. The most common 2D similarity calculation algorithms utilize molecular fingerprints as the similarity measure and encapsulate the molecular connectivity/substructure/chemical feature information into a string of bits, and then some straightforward analytical metrics can be used to compare the implicit relationship between the two compounds, with the famous being like Tanimoto coefficient.[12] Representative implementations of 2D molecular similarity calculation include Pipeline Pilot's connectivity fingerprint family (like ECFP and FCFP),[13] MACCS Keys,[14] Daylight fingerprint,[15] MOLPRINT 2D,[16,17] maximum common subgraph (MCS),[18,19] and feature trees (FTree).[20] However, although 2D similarity methods are proven efficient for fast neighboring compounds space profiling, the medicinal chemists who use them are exposed to two challenges. First, different 2D similarity definitions target different aspects of the information (substructure key definition) contained in the query molecules and may give different hit lists to some extent for the same query and candidate pool. Second, fingerprint-based 2D similarity searching is often criticized for tending to discover close structural analogues instead of novel hits, a concept referred as "Scaffold Hopping", which is considered as one of the most important criteria to evaluate the performance of virtual screening. As a complementarity, some methods incorporate the results from different similarity searching, including data fusion,[21,22] consensus scoring,[23] and belief theory.[24]

3D similarity measurements usually involve geometrical information of predefined objectives from the 3D molecular

Received: February 8, 2011
Published: August 08, 2011

conformations, which include pharmacophores, molecular shapes, and molecular fields. Pharmacophore based methods enumerate the profile of the distances between the predefined pharmacophore features and compare the profiles between the query and target molecules. Flexophore[25] and MED-SuMoLig[26] use enhanced atom types to establish the reduced graph of the underlying molecules, whose similarity can be calculated from comparing the corresponding descriptor nodes. LigMatch[27] adopts geometrical hashing to align the molecules, and the similarities are calculated by the number of the coincident atoms with the same type in the mapping. Molecular shape or volume methods maximize the shape overlays between the two molecules, of which Gaussian functions are usually adopted to describe the molecular volumes.[28,29] Another variant of shape-based virtual screening method is Ultrafast Shape Recognition (USR) developed by Ballester, which encodes the shape information into a string vector and the similarity comparison can be performed in a nonsuperposition way.[30,31] Molecular field involves molecular electrostatic potential (MEP) and other field points representing hydrophilic and hydrophobic potential.[32−38] Fine description of the whole potential fields requires vast grid points with different "field color" which makes the comparison of the complete molecular fields quite time-consuming.[39] FieldScreen reduces the complete molecular fields to the corresponding local extrema and in turn significantly reduced the number of field points to be compared. In spite of general classification criterion mentioned above, increasing approaches integrated hybrid 3D similarity measurement by incorporating different similarity metrics into one to increase the accuracy in virtual screening.[40,41] The representatives of molecular volume based similarity methods include ROCS (Gaussian functions based molecular volume + chemical atom type)[42] and ShaEP (MEP + Gaussian functions based molecular volume),[43] which incorporate the chemistry properties or MEP into the hybrid similarity measurement and significantly increase the ranking accuracy and enrichment of active compounds in retrospective virtual screening.

Recently, we implemented a pharmacophore matching method based on feature triplet hashing and searching algorithm, which was applied as the engine of PharmMapper Server for drug targets identification via reversed pharmacophore matching.[44] In this study, based on the pharmacophore matching approach, we developed a method for rapid 3D molecular similarity calculation and evaluate its performance in virtual screening. The method, called SHAFTS (SHApe-FeaTure Similarity), adopts hybrid similarity metric of molecular shape and colored (or labeled) chemistry groups annotated by pharmacophore features for 3D similarity calculation and ranking, which is designed to integrate the strength of both pharmacophore matching and volumetric similarity approaches. The feature triplet hashing method from PharmMapper Server is used for fast molecular alignment poses enumeration, and the optimal superposition between the target and the query molecules can be prioritized by calculating corresponding "hybrid similarities", which is a weighted summation of volume overlap and feature points fit value described by Gaussian density functions. SHAFTS is suitable for large-scale virtual screening with single or multiple compounds as the query "templates" regardless of whether corresponding experimentally determined conformations are available. The problem of molecular flexibility is addressed by means of a conformational ensemble generated in either online or off-line way, and only the top ranked conformer with the highest similarity score is regarded as the final result for the current target molecule.

The performance of SHAFTS in virtual screening was evaluated on two public benchmark data sets for direct comparison with other literature-selected virtual screening methods. Moreover, to eliminate the bias introduced by the target compounds which share high 2D similarities with the query template molecule, the performance of scaffold-hopping for SHAFTS is accessed by means of chemotype clustering and arithmetic weighted receiver operating characteristics. The results revealed an improved early stage enrichment performance for both active compounds and novel chemotypes, which is comparable over the complete data sets, demonstrating its potential for virtual screening and scaffold hopping. As a proof-of-concept, SHAFTS was used to identify novel RSK2 kinase inhibitors by ligand-based virtual screening, which is described in detail elsewhere in the subsequent paper.[45]

EXPERIMENTAL METHODS

The outline of virtual screening process with SHAFTS was illustrated in Figure 1, which can be divided into three stages as follows. First SHAFTS requires an active molecule with reasonable 3D information as the query template. Since the basic assumption of 3D similarity searching is to identify candidates possessing similar binding properties to particular drug target with the template, the bioactive conformations determined by X-ray crystallization are preferred, however, not mandatory. The pharmacophore feature points are perceived from the input conformation of the template according to a chemical group based rule (A). SHAFTS uses a semirigid strategy to characterize molecular flexibility, of which a conformation ensemble has to be generated for each target molecule in the screening database (B) prior to similarity calculation, and an optional local optimization to the alignment poses can also be adopted to refine the conformation of the target molecules superposed on the template. During the searching, the same rule of feature perceiving rule is applied to each conformation of target molecule in the database. A feature triplet hashing and searching algorithm described previously was used to generate alignment poses of the target and template conformations, and each pose is scored and ranked by the hybrid similarity of volume and feature overlap (C). In the following section, further details of these stages are provided.

Conformation Generation. SHAFTS uses a semirigid strategy to characterize molecular flexibility. The conformation ensembles of corresponding target molecules in the database have to be generated in the process of conformation searching prior to similarity calculation. This process can be performed either online (generating conformational ensemble for each molecule sequentially using the conformation searching algorithm implemented in SHAFTS) or offline (preparing conformation database with external programs). In this study, we used Cyndi, a conformation generation program based on multiobjective evolution algorithms described previously,[46] to generate conformation ensembles (maximum size of 200) for each target molecule sets in the benchmark test prior to SHAFTS. The same method in Cyndi has also been integrated into SHAFTS for online conformation generation.

Pharmacophore Feature Triplets Hashing and Matching. The pharmacophore features in the query and target molecules are identified by means of a predefined rule, which maps different chemical groups onto corresponding feature types.[47] There are six types of pharmacophore features defined in SHAFTS,

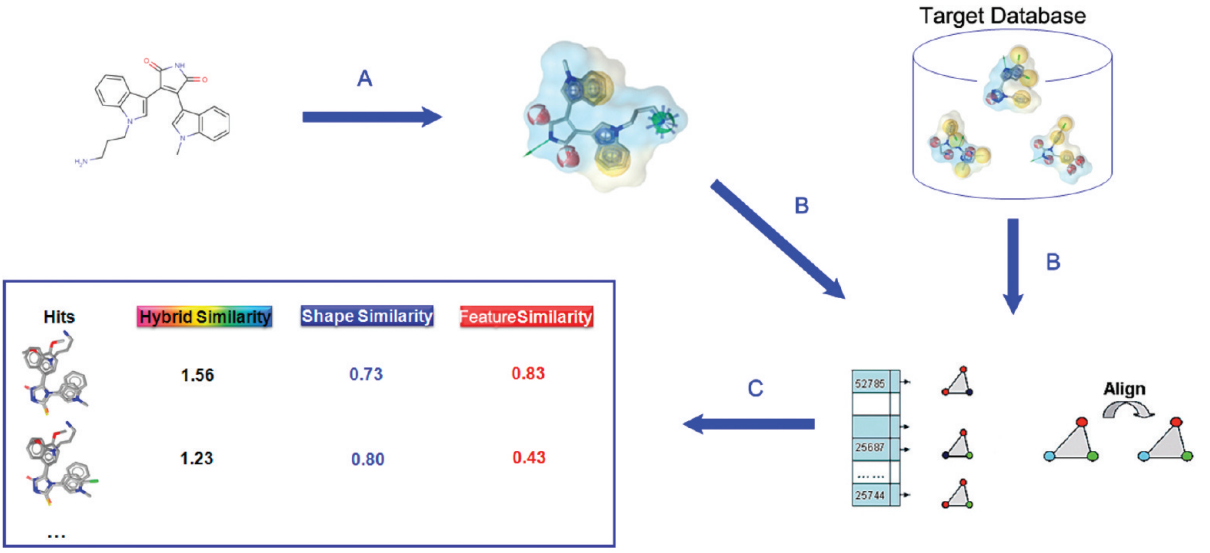


Figure 1. Schematic representation of SHAFTS' workflow for database virtual screening: (A) select an active compound, convert to a specific conformation, and add pharamacophore feature points as the query template; (B) search the database by overlaying each target structure on the query template using feature triplet hashing method; (C) rank the compounds in the hit list according to the hybrid similarities and output the corresponding alignment poses; (D) the target database is populated by conformational exploration of all molecules and feature triplets enumeration.

including Hydrophobic center (H), Positive charge center (P), Negative charge center (N), Hydrogen bond acceptor (HBA), Hydrogen bond donor (HBD), and Aromatic rings (AR). The detailed definitions of the pharmacophore features are illustrated in Figure S1. SHAFTS enumerates all possible feature triplets from both query template and target molecular conformations in the database and stored them into triplets hashing table whose vertex types and edge lengths are encoded as the searching key. All triplets from the template are used as the query one by one to search the triplet hashing table of each target molecule, and those with the same key are defined as a match. For each match, a least-squares fitting routine is performed to align the corresponding triplets and the relevant molecules (as shown in Figure S2). Then the hybrid similarity is calculated for each alignment mode, and only the one from the conformation with the highest similarity score is retained. The top-scored alignment modes can be further improved optionally through a simplex optimization.

Like most virtual screening methods, SHAFTS provided customized filter prior to feature triplet searching to reduce the overall computational cost in the screening. For example, knowledge from SAR analysis of the template molecule can be taken into consideration. If some mandatory pharmacophore feature types are required and labeled in the template, all the target molecules without such types of feature will be skipped, and only the feature triplet matches involved with the same types of feature will be aligned and scored. The feature triplet hashing method is extremely fast, taking about 5 ms for each conformation and 1−2 s for a typical query against a conformational ensemble of one target molecule.

Hybrid Similarity Calculation. The hybrid similarity used to score and rank alignment modes consists of shape-densities overlap (*ShapeScore*) and pharmacophore feature fit values (*FeatureScore*). For *ShapeScore*, the overlap of shape between molecules A and B can be defined as the sum of the overlap integrals of individual atomic shape-densities expressed in Gaussian function. Similar to that in ROCS and ShaEP,[28,29] the whole space shape-density overlap between molecules A and B can be formulated as follows

$$V_{AB} = \sum_{i \in A} \sum_{j \in B} \int d\vec{r}\, \rho_i(\vec{r})\rho_j(\vec{r}) = \sum_{i \in A} \sum_{j \in B} p_i p_j \exp\left(-\frac{\gamma_i \gamma_j}{\gamma_i + \gamma_j}\right)\left(\frac{\pi}{\gamma_i + \gamma_j}\right)^{3/2} \quad (1)$$

where i and j run over the atoms in A and B, respectively, and d_{ij} is the interatomic distance between atom i and j. γ is the width of a Gaussian which is relevant with atomic van der Waals radii. The final *ShapeScore* can be normalized to $[0, 1]$ using the cosine similarity metric

$$ShapeScore = \frac{V_{AB}}{\sqrt{V_A V_B}} \quad (2)$$

FeatureScore is essentially the fit value between the pharmacophore points extracted from template and target molecules, which is defined as the sum of the overlap between the feature points in A and B with the same type. An exponential function is used to normalize the overlap between single pair of feature points to $[0, 1]$

$$F_{AB} = \sum_{f \in F} \sum_{i \in A} \sum_{j \in B} \exp\left[-2.5\left(\frac{d_{ij}}{R_f}\right)^2\right] \quad (3)$$

where i and j run over the feature points with the same type f in A and B, respectively, d_{ij} is the distance between point i and j, and R_f is the overlap tolerance with a default value of 0.8 Å. Similar to *ShapeScore*, the final *FeatureScore* is normalized to $[0, 1]$ using the cosine similarity metric

$$FeatureScore = \frac{F_{AB}}{\sqrt{F_A F_B}} \quad (4)$$

The hybrid similarity is the weighted sum of *FeatureScore* and *ShapeScore* with the default weighting factor of 1.0 and scaled to $[0, 2]$

$$HybridScore = ShapeScore + w \cdot FeatureScore \quad (5)$$

dx.doi.org/10.1021/ci200060s |J. Chem. Inf. Model. 2011, 51, 2372–2385

Data Sets for Virtual Screening. *DUD Sets.* The performance of SAHFTS in ligand-based virtual screening was evaluated using a tailored version of the Directory of Useful Decoys (DUD).[48] DUD contains sets of active compounds and "decoys" with similar physical properties but dissimilar topology for 40 target proteins and has been used as an unbiased benchmark for virtual screening.[49] The original DUD is compiled for molecular docking and therefore not suitable for ligand-based virtual screening.[50] Recently, Good and Opera[51] carried out an analysis of DUD using a lead-like filter and clustering algorithms based on a reduced graph presentation to remove large molecules with unsuitable physiochemical properties and reduce the bias from structural analogues on actives enrichment.[52,53] We adopted the DUD LIB VS 1.0 set compiled by Jahn[54] according to the analysis results of Good as the first test set, in which each cluster of the active compounds can be regarded as a unique chemotype and used to evaluate the scaffold-hopping capability of the screening methods. To remove the bias resulted from target sets with less number of chemotypes, we also highlight the virtual screening results from a subset of DUD (13 targets) with a minimum of 17 chemotypes for each target. This subset was first used by Cheeseright et al. to compare the virtual screening results of FieldScreen with other methods[41] and has been adopted widely to evaluate the performance of active compounds enrichment and scaffold-hopping capability in other recent studies.[27,43,54]

As a 3D similarity calculation method, SHAFTS needs one or multiple bioactive query template molecules with reasonable 3D conformations. To make a fair comparison with the results from other similarity methods, the conformations of crystallized ligands in the complex crystal structures which were used to identify the binding site of each target were extracted from the corresponding PDB files and manually checked. The single bioactive conformations were set as the query templates for all the 3D virtual screening methods in this study. A comprehensive overview of the DUD set is presented in Table 1.

Data Sets of Jain. Another test set that was used to evaluate the performance of ligand-based virtual screening is the "revised" benchmarks used by Jain to model the biological targets of the known drugs with Surflex-Sim.[55] The set consists of active compounds for 23 drug target proteins and one universal decoy set containing 850 drug-like molecules compiled by Rognan.[56] Different from DUD, nearly half of the target proteins in this test set are G-protein coupling receptors (GPCRs) and ion channels, which prove to be a challenge for 3D similarity based virtual screening because very limited structural information is available due to lack of crystal structures as the bioactive query templates. Jain used Surflex-Sim to align 2−3 active compounds to build a combined query model (titled as automatically generated hypothesis) for each target, which is essentially overlaying poses of multiple active compounds. In this study, to test the independence of SHAFTS on bioactive conformations, both single active compounds and combined query model were used as the query templates, respectively, for virtual screening. A comprehensive overview of Jain's set is presented in Table 2.

Comparison with Other Methods. To make a fair evaluation of the performance of SHAFTS in virtual screening, ROCS and ShaEP were used as reference screening methods for both DUD set and Jain's set. Similar to SHAFTS, ROCS and ShaEP also take shape and chemical properties into consideration and rank the screening results according to hybrid similarity. We use ROCS and ShaEP to perform virtual screening against the two test sets with the default setting. Besides, to provide a reference for common methods in ligand-based virtual screening, we compared the results of SHAFTS with pharmacophore matching method and molecular fingerprint similarity method. The pharmacophore matching was performed by our in-house program PharmMapper based on the same feature triplet hashing algorithm in SHAFT, and the query pharmacophore models for each target in the DUD subset were

Table 1. Summary of the DUD Data Set Used in This Study,[a] Showing the Target Name, Protein Data Bank (PDB) Code for the Crystal Template Structure, Number of Actives, Decoys, and Clusters According to Chemotypes for Each Target

target	template PDB code	no. of actives	no. of decoys	no. of clusters
ace	1o86	46	1796	19
ache	1eve	99	3859	19
cdk2	1ckp	47	2070	32
cox2	1cx2	212	12606	44
egfr	1m17	365	15560	40
fxa	1f0r	64	2092	19
hivrt	1rt1	34	1494	17
inha	1p44	57	2707	23
p38	1kv2	137	6779	20
pde5	1xp0	26	1698	22
pdgfrb	1t46	124	5603	22
src	2src	98	5679	21
vegfr2	1fgi	48	2712	31
ada	1ndw	23	927	8
alr2	1ah3	26	986	14
ampc	1xgj	21	786	6
ar	1xq2	68	2848	10
comt	1h1d	11	468	2
cox1	1q4g	23	910	11
dhfr	3dfr	190	8350	14
er_an	3ert	18	1058	8
er_ag	1l2i	63	2568	10
fgfr1	1agw	71	3462	12
gart	1c2t	8	155	5
gpb	1a8i	52	2135	10
gr	1m2z	32	2585	9
hivpr	1hpx	4	2038	3
hmgr	1hw8	25	1423	4
hsp90	1uy6	23	975	4
mr	2aa2	13	636	2
na	1a4g	49	1713	7
parp	1efy	31	1350	7
pnp	1b8o	25	1036	4
ppar	1fm9	6	3127	6
pr	1sr7	22	920	4
rxr	1mvc	18	575	3
sahh	1a7a	33	1346	2
thrombin	1ba8	23	1148	13
tk	1kim	22	891	7
trypsin	1bju	9	718	7

[a] The DUD set for ligand-based virtual screening (http://dud.docking.org/jahn/) filtered and clustered as suggested by Good and Opera[51] and regenerated by Jahn, Hinselmann, Fechner, and Zell.[54]

Table 2. Summary of Jain's Set Used in This Study, Showing the Target Name, Number of Templates, Actives, and Decoys[a]

target	no. templates	no. actives	no. decoys
lanosterol demethylase	3	6	846
D-Ala-D-Ala carboxypeptidase	3	22	846
dihydropteroate synthase	2	13	846
DNA gyrase	3	6	846
HIV reverse transcriptase	2	5	846
L-type calcium channel	2	4	846
acetylcholinesterase	2	5	846
angiotensin I converting enzyme	2	5	846
β-adrenergic receptor	2	5	846
opioid receptor μ	3	10	846
voltage-gated sodium channel	3	14	846
estrogen receptor	3	10	846
progesterone receptor	3	9	846
androgen receptor	2	10	846
gluco/corticosteroid receptor	3	30	846
COX-I COX-II	3	21	846
GABAA barbiturate site	3	9	846
GABAA benzodiazepine site	2	14	846
muscarinic acetylcholine receptor	3	15	846
histamine receptor	2	5	846
Na−Cl cotransporter	3	9	846
sulfonyl urea receptor	3	7	846

[a] Two or three template molecules are included in Jain's set, which were used to build the ligand-based models described by Cleves and Jain (http://www.jainlab.org/Public/SF-Test-Data-DrugSpace-2006.zip).[55] The decoy set was originally compiled by Rognan et al. and contained 850 druglike molecules, of which 4 molecules failed the conformational generation by Cyndi due to unsuccessful force field atom types assignment.

extracted from the same crystal structures whose ligands were used as the query templates in SHAFTS. The screening results of ECFP_4 fingerprint calculated with Pipeline Pilot were also incorporated, using the same template molecules as the queries and Tanimoto coefficient (Tc) as the scoring and ranking metric.

Evaluation of Virtual Screening Performance. The evaluation of virtual screening performance is an important but error-prone process.[57−59] Enrichment factor (EF) measures the improvement of hit rate by a virtual screening method compared with random selection at different screening stages (e.g., 1% or 20%), which used to be one of the most important performance descriptors for virtual screening.[60] However, it has been criticized by its high dependency on the ratio of active and decoy molecules in the test sets, which renders a direct comparison of screening performance between the test sets with different active/decoy ratios inappropriate. Alternatively, we use ROC enrichment (ROCE), which is defined as the ratio of active rate to the decoy rate at a given stage that particular percentage of the decoys has been observed.[58] As a result, this measurement is similar to EF but does not show dependency on the active/decoy ratio. To reflect the capability of active enrichment at early stage of virtual screening, we report the ROCEs at the decoy rates of 0.5%, 1.0%, 2.0% and 5.0%, as suggested by Jain and Nicholls.[58] Additionally, we also reported an overall measurement called area under the curve (AUC) of the receiver operating characteristic (ROC) for virtual screening against the complete data sets because comparison of many ROC plots is not straightforward. As with EF, ROCE values over 1.0 indicate enrichments are better than random at different decoy stages, and an AUC value approaching 1.0 signifies an ideal discrimination of actives from decoys.

Moreover, one of the advantages of the 3D similarity methods over the 2D methods is the scaffold-hopping potential.[6,8,61] However, ROCE and AUC calculation equally treat every active compound in the data set, which in turn are not unbiased evaluation measures for the tests involving many active analogues. To characterize the performance of novel scaffolds/chemotypes retrieval, as recommended by Clark and Webster-Clark,[62] arithmetic weighting ROCE (awROCE) and AUC (awAUC) were also reported for DUD LIB VS 1.0 data sets, of which the scaffolds information can be revealed from the clustering results of the active compounds.

RESULTS AND DISCUSSION

In this study, the screening performance and the scaffold-hopping potential of SHAFTS were investigated considering the hybrid benefits from pharmacological feature matching and volume overlay. The virtual screening performances on the DUD set and the Jain's set are outlined in Figure 2, Table 5, and Figure 6, along with the equivalent analysis of the results generated by other methods. For each target, the overall screening performance was characterized by the AUC and the awAUC values of the corresponding ROCs. In addition, active enrichment at early stage is also a practical performance snapshot to evaluate the virtual screening methods, highlighting the capability of discriminating the actives from the decoys. Several popular early stage ROCEs and awROCEs are 0.5%, 1%, 2%, and 5%, which were selected in this study for screening performance characterization, and the detailed results for each target in DUD and Jain's set are tabulated in Tables S1 to S15. Again, the equivalent analysis for the other methods is also presented for comparison.

Screening Performance (DUD). The results in Figure 2A,C show the AUCs for each target in the DUD set and average ROCEs at different screening stages. The detailed results for each target are tabulated in Tables S1 to S5. Comparing with other methods, SHAFTS obtains acceptable virtual screening performance and the average AUC value of SHAFTS is 0.77, which is higher than that of ROCS, ShaEP, and PharmMapper. The average ROC enrichments are higher at early stages of screening (ROCE0.5% and ROCE1.0% are 64.6 and 37.5, respectively), indicating that SHAFTS can rank a large portion of the actives as the top virtual screening hits and discriminate the actives from the decoys efficiently. SHAFTS achieved best overall performance in terms of AUC for 18 out of 40 targets, especially for ace, ache, p38, src, alr2, and hivpr, which can also be appreciated from the heatmap shown in Figure 2A. For other cases in which SHAFTS were outperformed, it still achieved competitive results and was never the worst one. However, similar to the observation of Kirchmair et al.,[63] for 3D similarity based virtual screening methods, there is not an obvious correlation between AUC values and early stage ROC enrichments. For example, the trends between AUC and ROCE0.5% values in the cases of pdgfrb and trypsin are significantly different. The inconsistency between AUC and early stage ROCE implies the limitations of using single benchmark and highlights the complementary between the two evaluation methods particularly in the early stage of virtual screening.

Across the tested methods, ROCS gave superior or similar results for some targets. The similarity metric implemented in SHAFTS is similar to that of ROCS for that the combo score in

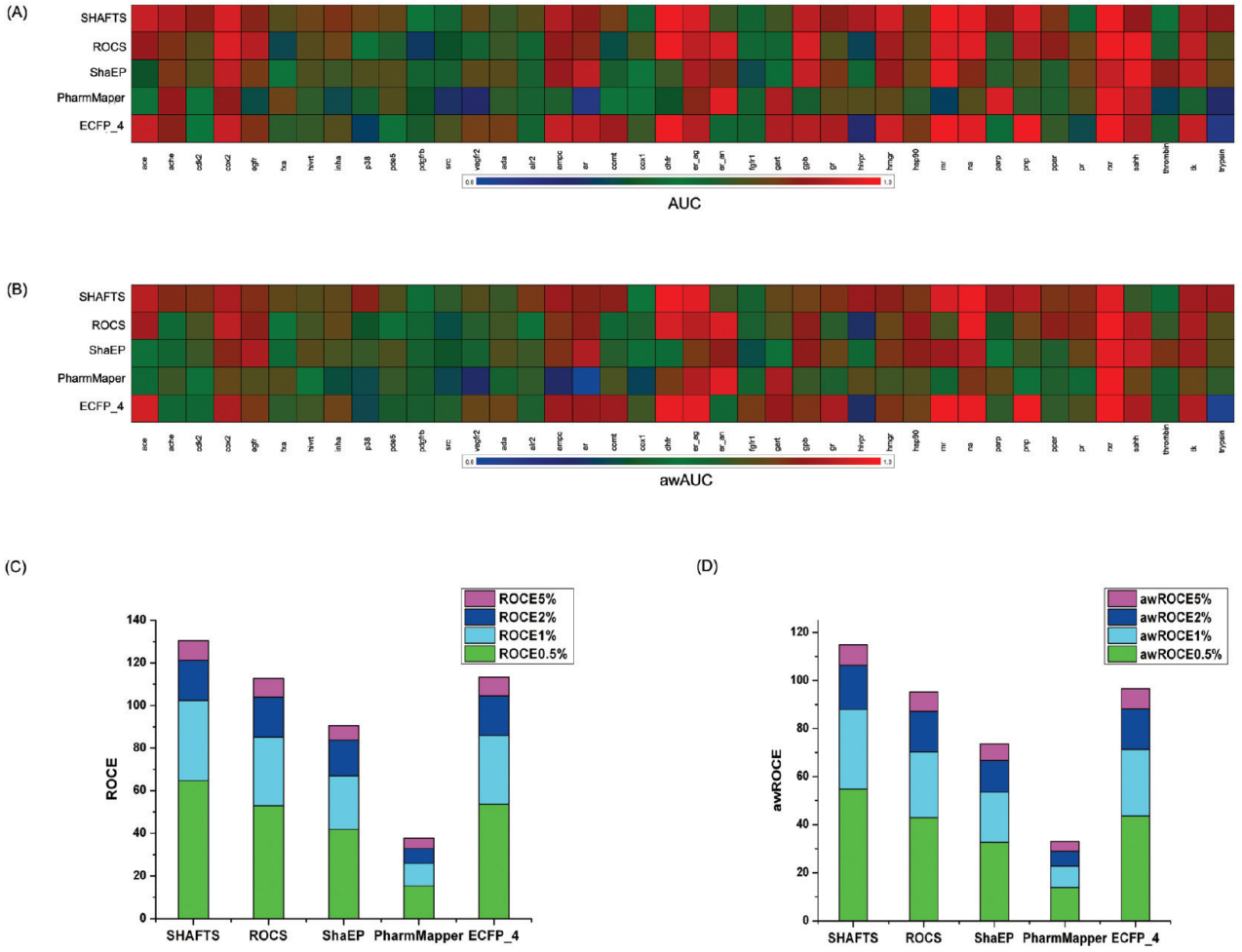


Figure 2. Comparison of the virtual screening performances of SHAFTS and other methods. Heat map plots showing the (A) AUC and (B) awAUC values for SHAFTS and other methods for each of the 40 DUD targets. Stack bar charts showing the average (C) ROCEs and (D) awROCEs at the 0.5%, 1%, 2%, and 5% screening stages, respectively.

ROCS is essentially a hybrid similarity score in combination of volume overlaying and atom type consistency. However, SHAFTS generates alignment poses by enumerating all potential pharmacophore feature triplets matches, while ROCS optimizes the initial alignments from the self-orientation into the internal frames by maximizing the volume overlay between the molecules. The pharmacophore features are defined in a fuzzy way independent of the atom types which consist of them. One of the advantages of feature triplet hashing over the local alignment optimization is that the molecules with inconsistent internal frames and even shapes can still be aligned and scored with a hybrid similarity metric. For instance in the case of src kinase, the query template compound is ANP, which appears troublesome for ROCS because the AUC value is only 0.37 and the ROCE at every early stage of virtual screening is around zero. The alignment pose generated by ROCS for one of the actives ZINC01609260 is shown in Figure 3B, which was ranked 4028 in the hit list of ROCS. Obviously, the pyrimidine mimic ring is expected to overlay on the pyrimidine ring of ANP like the one in Figure 3A generated by SHAFTS, which was ranked fifth in the hit list. Of course this is an extreme case because ANP is not appropriate to be a template for shape similarity based virtual screening and sometimes partial alignment between templates

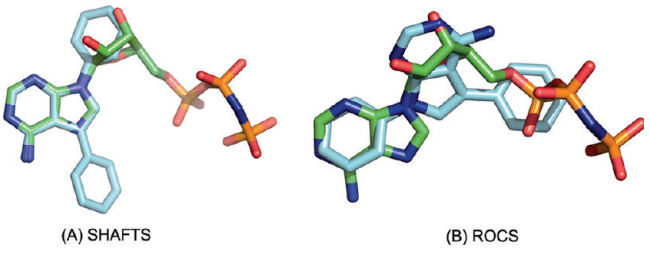


Figure 3. Superposition of an active (ZINC01609260) of src kinase onto the query template ANP using (A) SHAFTS and (B) ROCS, which are colored according to atom types (green carbon atoms for ANP and cyan carbon atoms for ZINC01609260).

and actives with high deviation of molecular weights may undermine the performance of SHAFTS. However, this case implies that although the similarity metric used in SHAFTS and ROCS are similar, the difference of initial alignment poses generating strategies can still play a pivotal role in the hit list ranking. On the other hand, PharmMapper uses the same feature triplet hashing method to generate alignment poses. Although it also takes advantage of molecular shape information by incorporating the position of the residues around the binding pockets as the

dx.doi.org/10.1021/ci200060s |J. Chem. Inf. Model. 2011, 51, 2372–2385

exclude volumes, SHAFTS and ROCS outperformed it in most cases, indicating shape similarity calculation by quantitative Gaussian volume overlay is more practical and reliable than simple position-restraints in pharmacophore matching.

Scaffold Hopping Potential (DUD). Traditional similarity based virtual screening is believed to suffer from biased enrichments caused by analog structures. To reduce the influence on overestimates of AUC and ROCE, the structures in the DUD set were preclustered according to corresponding 2D similarities and arithmetic weighted AUC and ROCE were applied as the performance metrics, in which the structures classified into one cluster have the equal weights.

The results of scaffold hopping potential for the tested methods are shown in Figure 2B,D. Also, detailed results can be referred to Tables S6 to S10. Comparing with other methods, SHAFTS still exhibits superior performances against most of the targets as to that of AUC and ROCE. SHAFTS obtains best overall results in 14 out of 40 targets in terms of awAUC value. The average awAUC value is 0.76, which is fairly better than that of ROCS. The awROCEs at early stage of virtual screening are also as high as that of ROCEs (the average awROCE0.5% and awROCE1.0% are 54.7 and 33.2, respectively), indicating the new chemotypes are efficiently enriched at early stages, especially in the cases of ache, egfr, pde5, dhfr, hmgr, and pnp. Cheersright et al.[41] used 13 of 40 target sets in DUD for the evaluation of FieldScreen, a 3D alignment based virtual screening program using a molecular field method. For each of the 13 targets, there are at least 15 clusters of actives. The same subset was also used by Jahn et al.[54] to compare the screening performance of ROCS, FieldScreen, MACCS, and their optimal assignment methods. Although the difference between data preparation may not allow for direct ranking of these methods, a qualitative analysis can be performed, and the results are reproduced in Tables 3 and 4. The results indicate that SHAFTS is still competitive in terms of both awAUC and awROCE values against this scaffold-focused subset.

The results presented above show that the scaffold-hopping performances of the tested 3D similarity based screening methods vary for different target sets. To make a detail exploration of this phenomenon, two cases are selected for further analysis at the early screening stage: hivpr, for which only SHAFTS outperformed other methods; and p38, for which all of the methods performed poorly. Figure 4 shows the alignment poses of the 2 scaffolds of the actives on the query template in the case of hivpr. The two actives show noticeable differences with the query template in 2D structure; however, the alignment poses generated by SHAFTS indicate that the common chemical groups are fairly well superposed, although the molecules are extremely flexible, and both the scaffolds and the molecular shapes exhibit remarkable deviations. For the case of p38 in Figure 5, although SHAFTS achieves an overall good performance, a large number of actives of which the scaffolds are very similar to that of query template are retrieved at the early screening stage. Mitogen-activated protein kinase p38 has been proved a challenge for most

Table 3. awAUC Values Obtained for "Scaffold-Focused" Subset of DUD Using SHAFTS and Different Methods Other than the Ones in This Study[a]

target	SHAFTS	LigMatch	FieldScreen	2SHA	DOCK	MACCS
ace	0.89	0.70	0.64	0.86	0.67	0.86
ache	0.77	0.76	0.62	0.50	0.57	0.37
cdk2	0.76	0.68	0.44	0.48	0.53	0.55
cox2	0.86	0.87	0.82	0.79	0.68	0.56
egfr	0.77	0.82	0.82	0.72	0.55	0.60
fxa	0.70	0.82	0.74	0.56	0.72	0.45
hivrt	0.68	0.84	0.63	0.53	0.73	0.54
inha	0.70	0.80	0.72	0.64	0.26	0.64
p38	0.79	0.64	0.27	0.76	0.36	0.38
pde5	0.67	0.71	0.62	0.38	0.48	0.28
pdgfrb	0.53	0.79	0.40	0.49	0.40	0.54
src	0.61	0.92	0.39	0.74	0.52	0.50
vegfr2	0.68	0.80	0.53	0.54	0.42	0.42
mean	0.73	0.78	0.59	0.61	0.53	0.51

[a] The awAUC values for LigMatch, FieldScreen, 2HSA, DOCK, and MACCS keys are reproduced from refs 27, 41, 43, and 54.

Table 4. awROCE Values at Early Screening Stages (0.5% and 1.0% of the Decoys) Obtained for "Scaffold-Focused" Subset of DUD Using SHAFTS and Different Methods Other than the Ones in This Study[a]

	SHAFTS		LigMatch		FieldScreen		2SHA		DOCK		MACCS	
target	0.5%	1%	0.5%	1%	0.5%	1%	0.5%	1%	0.5%	1%	0.5%	1%
ace	84.3	53.7	40	20.0	14.7	12.6	73.5	42.0	17.0	13.5	55.7	36.8
ache	77.7	49.3	35.4	20.8	16.7	20.4	24.8	15.2	0.0	0.0	19.1	9.8
cdk2	46.9	26.6	18.1	15.9	7.5	3.8	9.4	11.1	4.0	9.5	9.4	4.9
cox2	92.3	51.9	67.6	41.5	48.8	29.5	38.3	33.7	1.9	5.3	17.0	10.7
egfr	127.0	66.0	26.4	23.1	52.4	29.5	103.6	56.2	7.6	7.7	40.3	20.4
fxa	10.5	5.3	0.0	26.0	0.0	2.8	10.0	5.2	15.1	13.5	30.0	15.7
hivrt	35.3	17.6	57.8	31.1	40.0	11.7	20.1	10.7	4.4	2.2	22.0	10.7
inha	44.9	22.4	89.1	51.1	56.7	31.2	57.9	42.3	0.0	0.0	49.9	31.0
p38	6.3	5.2	12.4	24.2	3.7	1.8	28.2	16.3	0.0	0.0	1.0	0.5
pde5	27.3	13.6	18.2	11.4	6.8	4.5	4.3	2.3	4.4	6.7	4.3	2.0
pdgfrb	36.4	22.7	NA	18.2	27.3	13.6	43.9	22.3	0.0	0.0	42.0	23.2
src	13.9	7.7	104.2	60.0	13.7	7.0	4.3	9.7	0.0	4.7	0.0	0.0
vegfr2	8.1	4.0	65.3	39.1	12.9	8.1	12.5	9.4	6.2	3.2	18.7	9.4
mean	47.3	26.8	44.5	29.4	23.2	13.6	33.1	21.3	4.7	5.1	23.8	13.5

[a] The awROCE values for LigMatch, FieldScreen, 2HSA, DOCK, and MACCS keys are reproduced from refs 27, 41, 43, and 54.

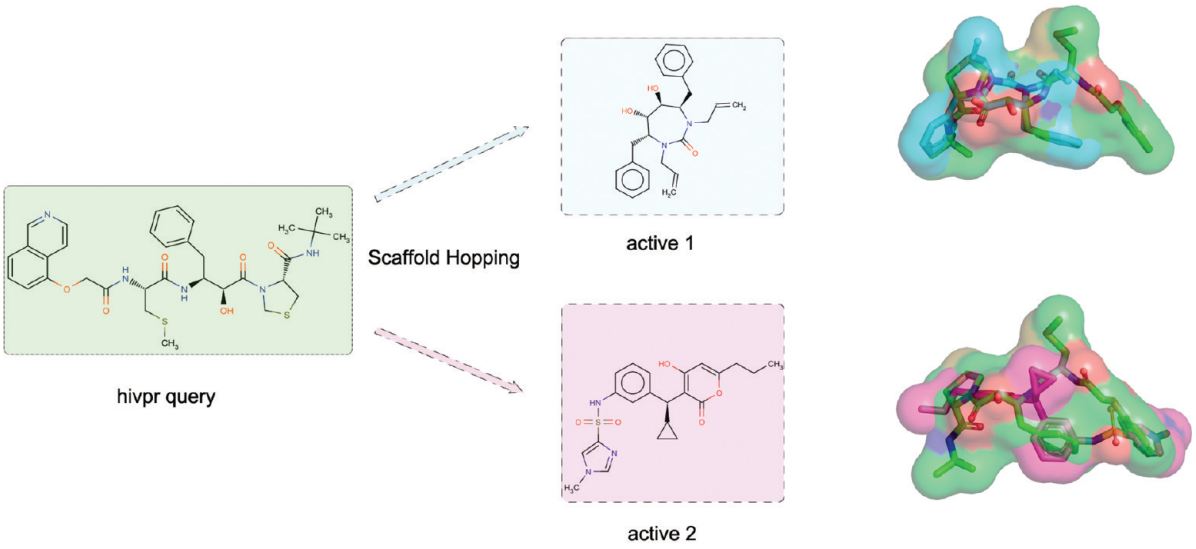


Figure 4. The chemical structures and corresponding superposition poses onto the query for two hivpr actives with different scaffolds which were only retrieved by SHAFTS ranking 2% of the database.

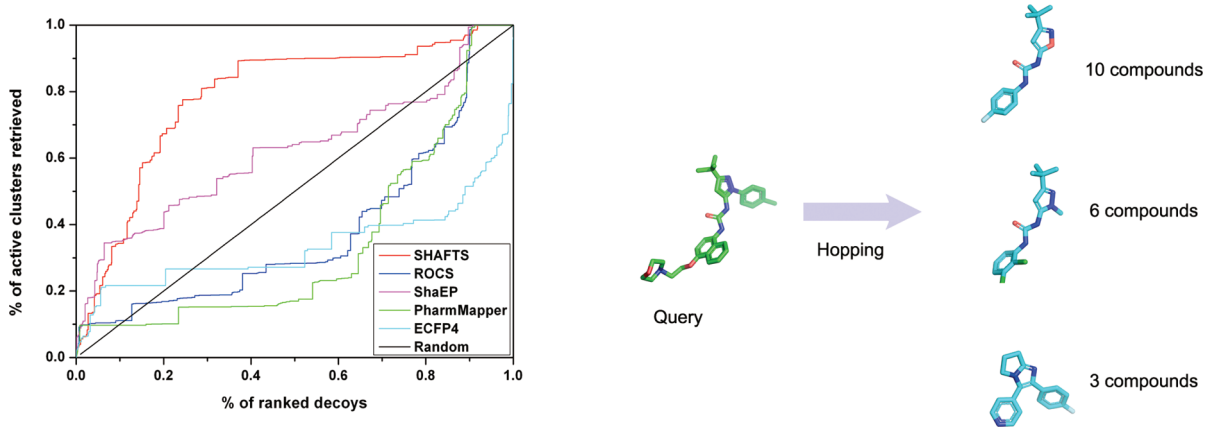


Figure 5. Scaffold enrichment of p38 kinase inhibitors represented by awROC curves of each method. The representatives of the chemotypes and corresponding numbers of analogues retrieved by SHAFTS ranking 2% of the database are shown in the right figure.

3D similarity methods, which can be attributed to the binding poses deviation resulted from the flexibility of the ATP binding site. The ATP binding site of p38 kinase has two stable conformations, namely "DFG-in" and "DFG-out", which can be bound by different types of inhibitors. The query template B96 is a typical "DFG-out" ligand, but some of the smaller actives in the p38 subset bind in the "DFG-in" conformation of p38 with remarkably different binding poses from the "DFG-out" ones. As a result, none of the 3D alignment based methods can align such type of compounds on the template and give appropriate rankings. The above results imply that single crystal structural conformation is not necessarily the best query template because it is unrealistic to expect that single query can retrieve all types of actives, let alone the actives with complete different binding poses to the target sites. Recently, it is highly recommended to use multiple active compounds as combined queries to improve both the actives and chemotypes enrichments and significant improvements were observed for some targets subsets (especially for the kinase families). SHAFTS also support multiple conformations (or compounds) as combined queries. Although we did not compare the results with multiple query templates for DUD set (primarily for increased computational costs introduced in some cases), remarkable improvements are still observed in most cases of the Jain's data set with 2 or 3 active query templates as described in the following section.

Screening Performance (Jain's Data Set). The results in the DUD data set were calculated with respect to the specific actives and decoys for each target. To make a more generic evaluation of the screening performance against the random selected drug-like decoys, a small-scale decoy set compiled by Rognan containing 825 drug-like molecules was used as the background to the 22 targets of Jain's data set. The same data set has been used to evaluate the screening performance of Surflex-Sim. The results of screening performance of SHAFTS and other methods on Jain's data set are outlined in Table 5 and Figure 6, and the detailed results are listed in Tables S11 to S15. Noting that for each target, we used two or three active compounds from the "best hypothesis" generated by Surflex-Sim as the respective query templates, and the screening performances were evaluated by the average AUC and ROCE values at the early stages across all the templates.

dx.doi.org/10.1021/ci200060s |J. Chem. Inf. Model. 2011, 51, 2372–2385

Table 5. AUC Values Obtained for the Targets in Jain'S Set by SHAFTS and Other Methods Used in This Study

target	SHAFTS[a]	SHAFTS_M[b]	ROCS[a]	ShaEP[a]	ShaEP_M[b]	ECFP_4[a]
lanosterol demethylase	0.98	0.98	0.83	0.77	0.89	0.94
D-Ala-D-Ala carboxypeptidase	0.99	0.99	0.99	0.61	0.66	0.99
dihydropteroate synthase	0.97	0.97	0.96	0.91	0.92	0.94
DNA gyrase	1.0	1.0	0.99	0.84	1.0	0.99
HIV reverse transcriptase	0.96	0.93	0.93	0.81	0.75	0.98
L-type calcium channel	0.93	0.90	0.84	0.77	0.77	0.89
acetylcholinesterase	0.61	1.0	0.75	0.88	0.99	0.68
angiotensin I converting enzyme	0.99	0.98	0.93	0.61	0.92	1.0
β-adrenergic receptor	0.81	0.98	0.82	0.96	0.97	0.90
opioid receptor μ	0.81	0.90	0.89	0.98	0.98	0.81
voltage-gated sodium channel	0.80	0.92	0.84	0.72	0.92	0.76
estrogen receptor	0.73	1.0	0.86	0.72	0.90	0.78
progesterone receptor	0.99	1.0	0.99	0.99	0.98	0.99
androgen receptor	0.72	0.97	0.81	0.88	0.86	0.76
gluco/corticosteroid receptor	0.98	0.96	0.94	0.95	0.95	0.95
COX-I COX-II	0.69	0.85	0.76	0.60	0.69	0.83
GABAA barbiturate site	1.0	1.0	1.0	1.0	1.0	0.98
GABAA benzodiazepine site	0.95	0.93	0.95	0.92	0.92	0.86
muscarinic acetylcholine receptor	0.84	0.89	0.81	0.95	0.96	0.90
histamine receptor	0.99	1.0	0.99	0.99	1.0	0.90
NaCl cotransporter	1.0	1.0	0.97	0.98	0.97	0.99
sulfonyl urea receptor	1.0	1.0	1.0	0.89	0.94	1.0
mean	0.9	0.96	0.90	0.85	0.91	0.90

[a] Average AUC values obtained from the multiple runs with different single templates as the queries. [b] AUC values obtained from the single run with all templates as the combined queries.

All the tested methods performed well across the 22 targets, and the average AUC values are over 0.85. SHAFTS obtained competitive overall performance compared with other methods (the average AUC is 0.9) and the ROCE values at early screening stage are also inspiring because the majority of the actives can be retrieved at the very earliest stages (the average ROCE%0.5 and ROCE%1 values are 106.9 and 59.3, respectively).

Unlike DUD, not all the query templates have crystal structural conformation, especially for the GPCRs and ion channels. Only hypothesis conformations from Surflex-Sim were used as the query templates. Although the superior screening performance across all the tested methods may be attributed to the small size of the decoy set used for each target, the independence of bioactive conformation determined from crystal structures for the 3D similarity based screening methods are still promising because for most GPCRs and ion channels, the crystal structures of the modulators are not available yet. Our results indicate that the bioactive conformations are preferred, but not mandatory, for 3D similarity based virtual screening in some cases.

Impact of Multiquery Molecules on the Performance. It is strongly recommended that incorporating multiple template molecules spanning diverse active scaffold families as the queries to obtain a combined ranking list can enhance the screening performance for most of 3D similarity methods. Using multiple templates as the combined query can alleviate inappropriate ranking of the actives resulted from the structural and binding modes deviation because it is not realistic to assume a global match that exists between most active compounds, just like the multiple templates for SHAFTS in the cases of Jain's data set. Significant improvement against DUD set with multiple crystal structures and multiple conformations of single query has also been reported recently for 3D alignment based similarity calculation.[27,63,64] Since SHAFTS and ShaEP support utilizing multiple template compounds as the combined query, we also compared the results from using all the hypothesis compounds in Jain's data set with the average results from single query method.

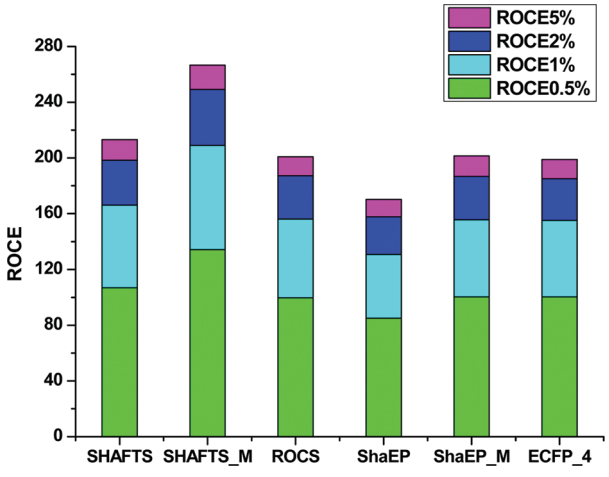


Figure 6. Comparison of the virtual screening performances of SHAFTS and other methods for the targets in Jain's data set. Stack bar charts showing the average ROCEs at the 0.5%, 1%, 2%, and 5% screening stages, respectively. SHAFTS_M and ShaEP_M are the ones of which multiple templates were used as the combined query and MAX fusion strategy was used to rank hit lists.

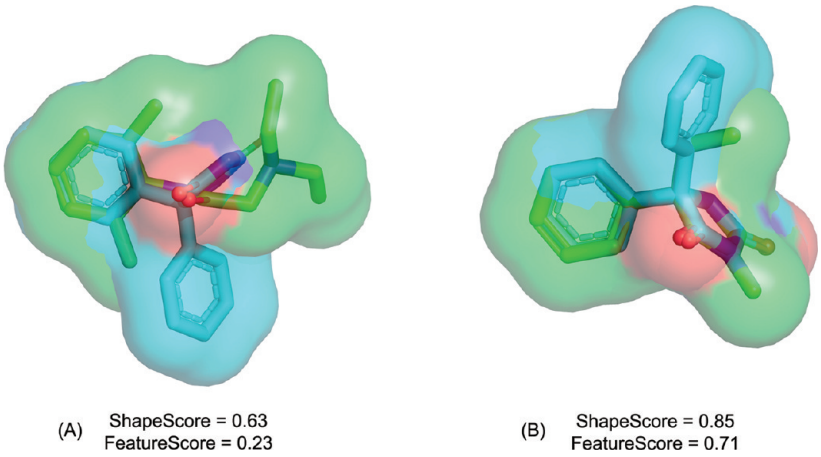


Figure 7. Superposition of a voltage-gated sodium channel antagonist (phenytoin) onto two query templates (A) lidocaine and (B) mephenytoin with corresponding *ShapeScores* and *FeatureScores* calculated by SHAFTS. The molecules are colored according to atom types (green carbon atoms for templates and cyan carbon atoms for phenytoin).

The effects on the screening performance using multiple active compounds for the 22 target sets are also shown in Table 5 and Figure 6. For SHAFTS, the multiple templates method generally outperforms the single template version in terms of overall enrichment across the data set (the average AUC value is 0.961) and early stage enrichments (the ROCE%0.5 and ROCE%1 values are 134.2 and 74.9, respectively). The early stage enrichment (ROCE%0.5) improves for 12 out of 22 targets for when using SHAFTS with multiple templates method. For 5 out of 22 targets the performances remain the same, and in the cases of the other five targets the performances declined.

The most significant improvements in early stage ROCE can be observed in the cases of HIV Reverse Transcriptase (160%), Acetylcholinesterase (100%), Voltage-Gated Sodium Channel (160%), Estrogen Receptor (170%), Androgen Receptor (140%), and Muscarinic Acetylcholine Receptor (70%), which show greatest percentage increases with the multiple templates method (160, 100, 160, 170, 140, and 70%, respectively). The improvement incorporating multiple templates can be attributed to the enrichment of the actives presented in the data sets by the queries with different scaffolds. For instance, in the case of Voltage-Gated Sodium Channel, the early stage enrichments for the three single templates are all below 50.0, indicating none of them can retrieve all types of active scaffolds. However, using the 3 templates as the combined query, the ROCE0.5% was increased to 100.0, implying that some of the actives missed by single template can be retrieved by others in a complementary way. The alignment poses and corresponding hybrid similarities for one of the actives, phenytoin, and two of the templates are shown in Figure 7. Although the shape overlay scores between the two templates, namely hypo1 and hypo2, are similar (0.63 and 0.85 for hypo1 and hypo2, respectively), the pharmacophore feature similarities deviate remarkably (0.23 and 0.71 for hypo1 and hypo2, respectively), which results in different ranking in the hit lists (173 and 2 in the cases of hypo1 and hypo2, respectively). It seems that using multiple templates including hypo1 and hypo2 removes the bias with respect to specific single template and gives a promotion in the ranking list to phenytoin (6 in the case of multiple templates), leading to an overall robust result.

Effect of Weight Variation on the Performance. To study the effect of the weight factor variation in eq 5, we described the performances of SHAFTS against the subset of DUD containing 13 targets with various weight factors (0.5, 0.8, 1.0, 1.2, and 1.5) to scale the contribution of FeatureScore. The results are shown in Figure S3 and Table S16. In most cases the weight factor of 1.0 gives the better results, and increasing the weight factor in some cases improved the screening performance. Obviously, it is a case-by-case setting parameter to decide the contribution of the feature score against the shape score depending on the shapes and chemotypes of both the query template and the target compounds. We did not do full optimization for the weight in this study because we aimed to make a relatively fair comparison with other methods, of which the contributions of shape score and "color score" in ROCS and ESP score in ShaEP are all deemed as equivalent with default values.

Comparing with 2D Similarity Method. It is often expected that the 3D similarity based methods can provide a better way to identify novel active compounds comparing with 2D methods because the 3D shape and pharmacophore feature distribution are generally believed to play a pivotal role in ligand−receptor recognition. However, the overall analysis of the screening performance and scaffold hopping potential on DUD and Jain's data set reveals that the 2D fingerprint method ECFP_4 is competitive to 3D methods and even outperforms others in some target cases (as shown in Figures 2 and 4). Similar observations have been made by many other researchers with different data sets, where the 2D methods outperformed 3D methods, including ligand-based and structure-based ones.[65−67] In the very recent study, Venkatraman et al.[64] made a comprehensive comparison between 2D and 3D similarity methods against the DUD set and observed the similar phenomenon: 2D methods generally give competitive or better performance than the 3D shape based methods for many targets. A common caveat is the 2D methods find actives that are not quite as diverse as found by 3D methods, but the difference in diversity is not necessarily large. However, ECFP_4 fingerprint has recently proved overall the highest scaffold hopping potential among the 2D methods.[68] One may ask how it is possible for 2D similarity methods to perform nearly as well as 3D methods at lead hopping. ECFP_4 fingerprint is a combinatorial molecule-specific fingerprint that encodes layered atom environments with a maximum diameter of four bonds around each atom in a molecule, whose concept is

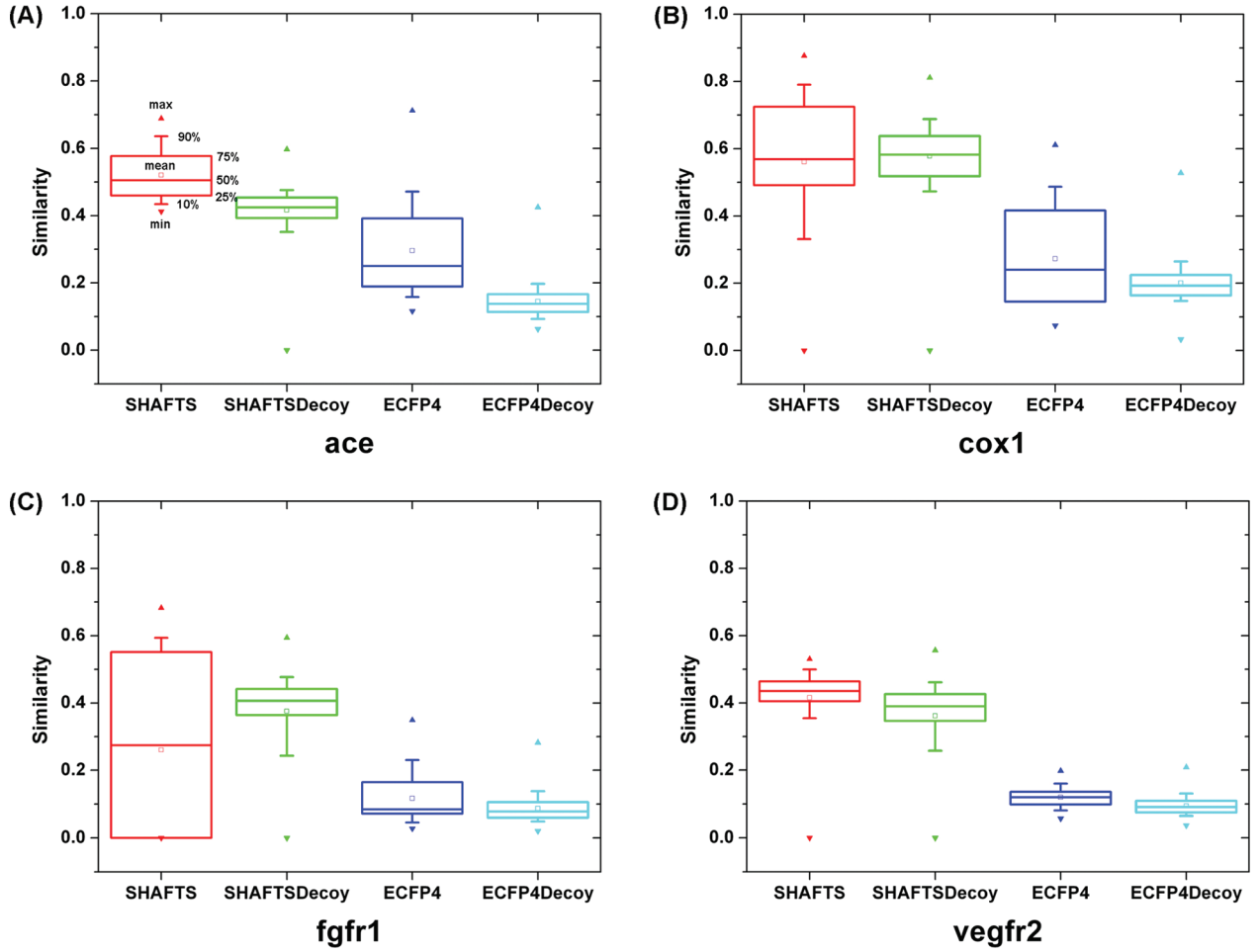


Figure 8. Box plots showing the distribution of SHAFTS (scaled to [0,1] by dividing the *HybridScore* with 2.0) and ECFP_4 similarities of the actives (SAHFTS and ECFP4) and decoys (SHAFTSDecoy and ECFP4Decoy) between the corresponding queries in 4 DUD targets in which ECFP_4 outperformed SHAFTS: (A) ace, (B) cox-1, (C) fgfr1, and (D) vegfr2. The upper and lower edges of the box correspond to the 75th and 25th percentile of the similarity, respectively. The line in the box indicates the median value and the upper and lower ends of the "whiskers" correspond to 90th and 10th percentile, respectively. The minimum, mean, and maximum values of similarities are indicated by ▼, □, and ▼, respectively.

essentially similar to the chemotype features generation in SHAFTS. Distribution of the similarities between the actives and decoys are considerably different from that of SHAFTS. As Figure 8 shows, the distributions of ECFP_4 similarity are considerably shifted to the low Tc values (0.1 to 0.4) which is in accordance with the results of Vogt et al.,[68] and the overlaps between the similarity to the actives and decoys are much smaller than that of SHAFTS in the 4 representative DUD targets where ECFP_4 outperformed SHAFTS, namely, the high enrichments of the actives obtained by using ECFP_4 method dominantly stem from the surprisingly dissimilarities between the decoys and queries rather than from the similarities between the actives and queries. In DUD set the decoys were berry picked which share low 2D similarity with the actives (Tc values less than 0.9 for annotated actives calculated with the CACTVS fingerprint).[48] Nevertheless, the active molecules may be similar enough to the query by some 2D descriptors which are not common among the decoys, though they are not so similar with the close analogs of the query. Thus it is possible for 2D similarity searches to hop to another series and have a high enrichment in a diverse database.

Despite our observations that 2D method yields higher enrichment rates than 3D methods in most cases in our test, in this paper we note that SHAFTS and ROCS are nearly on a par with ECFP_4. In our study, SHAFTS is the only 3D method which can retrieve competitive results to the ECFP_4 method: for DUD data set, SHAFTS outperforms ECFP_4 in 24 of 40 targets in terms of AUC and 29 of 40 targets in terms of both ROCE%0.5 and ROCE%1; for Jain's data set, SHAFTS still outperforms ECFP_4 in 13 and 16 of 22 targets in terms of AUC and ROCE values, respectively. Also, we have demonstrated that SHAFTS, a sophisticated 3D ligand-based algorithm, is able to select more diverse actives than 2D methods at the early stage (23 of 40 targets in terms of awROCE0.5%). This result justifies the use of computationally more expensive methodology comparing with the 2D methods for virtual screening in practice which was described elsewhere.[45]

CONCLUSION

A hybrid ligand 3D similarity based virtual screening method called SHAFTS was developed, and the corresponding performances of retrospective virtual screening and scaffold-hopping against the DUD set and Jain's data set were described. SHAFTS incorporates the shape overlay and pharmacophore feature matching into the hybrid similarity. A feature triplet hashing

method was implemented to enumerate the potential alignment poses between query and target molecules, rendering SHAFTS fast enough for large-scale compound database screening.

Compared with other ligand-based virtual screening methods (ROCS, ShaEP, PharmMapper, and ECFP_4), SHAFTS achieved a superior performance in terms of both overall and early stage enrichments of known actives and chemotypes. However, the accuracy of SHAFTS is prone to suffer from using single template compound as the query because of the potential deviations of molecular sizes and binding modes between the query and actives, which reduces its performances in some target cases. As a remedy, SHAFTS supports using multiple templates as combined query, and the MAX fusion strategy was used to rank the final retrieving list. The results against the Jain's data set indicate that multiple templates method outperformed the single template version in most target cases. Moreover, in spite of lacking enough test results for evaluation, it is not mandatory for SHAFTS to adopt bioactive conformations of the template compounds because the screening performances for the GPCRs and ion channels in Jain's data set are promising. As a benchmark of prospective virtual screening, SHAFTS was applied to identify novel RSK2 kinase inhibitors, which was described elsewhere.[45]

As a reliable 3D similarity calculation method, according to *Similar Property Principle*, SHAFTS can also be used to identify the potential drug targets or explore multiple drug targets for the given bioactive compounds and even explore potential mechanisms for drugs' side effects. SHAFTS has been integrated into ChemMapper Server, which is a Web-based open resource for bioactive space navigation and exploration by molecular similarity methods (unpublished result), and can be accessed at http://59.78.96.61:8080/chemmapper/download.html.

■ ASSOCIATED CONTENT

Supporting Information. Detailed pharmacophore feature definitions and workflow of feature triplet hashing algorithms in SHAFTS are shown in **Figures S1** and **S2**. Comprehensive comparisons in terms of AUC, awAUC, ROCE, and awROCE for SHAFTS and other methods across all the targets in DUD and Jain's set are listed in **Tables S1** to **S15**. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +86-21-64250213. Fax: +86-21-64250213. E-mail: hlli@ecust.edu.cn.

■ ACKNOWLEDGMENT

This work was supported by the Fundamental Research Funds for the Central Universities, the National Natural Science Foundation of China (grants 20803022, 21173076, and 81102375), the Special Fund for Major State Basic Research Project (grants 2009CB918501 and 2011CB910200), the Shanghai Committee of Science and Technology (grants 09dZ1975700 and 10431902600), and the National S&T Major Project of China (grant 2011ZX09307-002-03). Honglin Li is also sponsored by Shanghai Rising-Star Program (grant 10QA1401800) and Program for New Century Excellent Talents in University (grant NCET-10-0378). The authors thank OpenEye Scientific Software Inc. and Accelrys Inc. We are also grateful to Dr. Mikko J. Vainio for making the programs ShaEP and Balloon publicly available.

■ REFERENCES

(1) Maggiora, G. A., Johnson, M. A. *Concepts and applications of molecular similarity*; John Wiley: New York, 1990.

(2) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J Med. Chem.* **2002**, *45*, 4350–4358.

(3) Muchmore, S. W.; Edmunds, J. J.; Stewart, K. D.; Hajduk, P. J. Cheminformatic tools for medicinal chemists. *J. Med. Chem.* **2010**, *53*, 4830–4841.

(4) Maldonado, A. G.; Doucet, J. P.; Petitjean, M.; Fan, B. T. Molecular similarity and diversity in chemoinformatics: From theory to applications. *Mol. Diversity* **2006**, *10*, 39–79.

(5) Niedermeier, S.; Singethan, K.; Rohrer, S. G.; Matz, M.; Kossner, M.; Diederich, S.; Maisner, A.; Schmitz, J.; Hiltensperger, G.; Baumann, K.; Holzgrabe, U.; Schneider-Schaulies, J. A small-molecule inhibitor of nipah virus envelope protein-mediated membrane fusion. *J. Med. Chem.* **2009**, *52*, 4257–4265.

(6) Oyarzabal, J.; Howe, T.; Alcazar, J.; Andres, J. I.; Alvarez, R. M.; Dautzenberg, F.; Iturrino, L.; Martinez, S.; Van der Linden, I. Novel approach for chemotype hopping based on annotated databases of chemically feasible fragments and a prospective case study: new melanin concentrating hormone antagonists. *J. Med. Chem.* **2009**, *52*, 2076–2089.

(7) Oyarzabal, J.; Zarich, N.; Albarran, M. I.; Palacios, I.; Urbano-Cuadrado, M.; Mateos, G.; Reymundo, I.; Rabal, O.; Salgado, A.; Corrionero, A.; Fominaya, J.; Pastor, J.; Bischoff, J. R. Discovery of mitogen-activated protein kinase-interacting kinase 1 inhibitors by a comprehensive fragment-oriented virtual screening approach. *J. Med. Chem.* **2010**, *53*, 6618–6628.

(8) Rush, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **2005**, *48*, 1489–1495.

(9) Ballester, P. J.; Westwood, I.; Laurieri, N.; Sim, E.; Richards, W. G. Prospective virtual screening with Ultrafast Shape Recognition: the identification of novel inhibitors of arylamine N-acetyltransferases. *J. R. Soc. Interface* **2010**, *7*, 335–342.

(10) Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L. H.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175–181.

(11) Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* **2007**, *25*, 197–206.

(12) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983–996.

(13) *Pipeline Pilot*, version 7.5; Accelrys: San Diego, CA, 2009.

(14) *MACCS structural keys*; Symyx Software: San Ramon, CA, 2010.

(15) Daylight Fingerprints, *version 4.62*; Daylight Chemical Information Systems: Laguna Niguel, CA, 1999.

(16) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): Evaluation of performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1708–1718.

(17) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Molecular similarity searching using atom environments, information-based feature selection, and a naive Bayesian classifier. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 170–178.

(18) Raymond, J. W.; Gardiner, E. J.; Willett, P. RASCAL: Calculation of graph similarity using maximum common edge subgraphs. *Comput. J.* **2002**, *45*, 631–644.

(19) Raymond, J. W.; Gardiner, E. J.; Willett, P. Heuristics for similarity searching of chemical graphs using a maximum common edge subgraph algorithm. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 305–316.

(20) Rarey, M.; Dixon, J. S. Feature trees: a new molecular similarity measure based on tree matching. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 471–490.
(21) Whittle, M.; Gillet, V. J.; Willett, P. Analysis of data fusion methods in virtual screening: Theoretical model. *J. Chem. Inf. Model.* **2006**, *46*, 2193–2205.
(22) Whittle, M.; Gillet, V. J.; Willett, P.; Loesel, J. Analysis of data fusion methods in virtual screening: Similarity and group fusion. *J. Chem. Inf. Model.* **2006**, *46*, 2206–2219.
(23) Baber, J. C.; William, A. S.; Gao, Y. H.; Feher, M. The use of consensus scoring in ligand-based virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 277–288.
(24) Muchmore, S. W.; Debe, D. A.; Metz, J. T.; Brown, S. P.; Martin, Y. C.; Hajduk, P. J. Application of belief theory to similarity data fusion for use in analog searching and lead hopping. *J. Chem. Inf. Model.* **2008**, *48*, 941–948.
(25) von Korff, M.; Freyss, J.; Sander, T. Flexophore, a new versatile 3D pharmacophore descriptor that considers molecular flexibility. *J. Chem. Inf. Model.* **2008**, *48*, 797–810.
(26) Sperandio, O.; Andrieu, O.; Miteva, M. A.; Vo, M. Q.; Souaille, M.; Delfaud, F.; Villoutreix, B. O. MED-SuMoLig: A new ligand-based screening tool for efficient scaffold hopping. *J. Chem. Inf. Model.* **2007**, *47*, 1097–1110.
(27) Kinnings, S. L.; Jackson, R. M. LigMatch: a multiple structure-based ligand matching method for 3D virtual screening. *J. Chem. Inf. Model.* **2009**, *49*, 2056–2066.
(28) Bemis, G. W.; Kuntz, I. D. A fast and efficient method for 2D and 3D molecular shape description. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 607–628.
(29) Grant, J. A.; Pickup, B. T. A Gaussian description of molecular shape. *J. Phys. Chem.* **1995**, *99*, 3503–3510.
(30) Ballester, P. J.; Finn, P. W.; Richards, W. G. Ultrafast shape recognition: Evaluating a new ligand-based virtual screening technology. *J. Mol. Graphics Modell.* **2009**, *27*, 836–845.
(31) Ballester, P. J.; Richards, W. G. Ultrafast shape recognition to search compound databases for similar molecular shapes. *J. Comput. Chem.* **2007**, *28*, 1711–1723.
(32) Wild, D. J.; Willett, P. Similarity searching in files of three-dimensional chemical structures. Alignment of molecular electrostatic potential fields with a genetic algorithm. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 159–167.
(33) Marin, R. M.; Aguirre, N. F.; Daza, E. E. Graph theoretical similarity approach to compare molecular electrostatic potentials. *J. Chem. Inf. Model.* **2008**, *48*, 109–118.
(34) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: a method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41*, 4502–4520.
(35) Jilek, R. J.; Cramer, R. D. Topomers: A validated protocol for their self-consistent generation. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1221–1227.
(36) Cramer, R. D.; Jilek, R. J.; Guessregen, S.; Clark, S. J.; Wendt, B.; Clark, R. D. "Lead hopping". Validation of topomer similarity as a superior predictor of similar biological activities. *J. Med. Chem.* **2004**, *47*, 6777–6791.
(37) Ahlstrom, M. M.; Ridderstrom, M.; Luthman, K.; Zamora, I. Virtual screening and scaffold hopping based on GRID molecular interaction fields. *J. Chem. Inf. Model.* **2005**, *45*, 1313–1323.
(38) Bergmann, R.; Linusson, A.; Zamora, I. SHOP: Scaffold HOPping by GRID-based similarity searches. *J. Med. Chem.* **2007**, *50*, 2708–2717.
(39) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233–3243.
(40) Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. Molecular field extrema as descriptors of biological activity: Definition and validation. *J. Chem. Inf. Model.* **2006**, *46*, 665–676.
(41) Cheeseright, T. J.; Mackey, M. D.; Melville, J. L.; Vinter, J. G. FieldScreen: Virtual Screening using molecular fields. Application to the DUD data set. *J. Chem. Inf. Model.* **2008**, *48*, 2108–2117.
(42) Nicholls, A.; Grant, J. A. Molecular shape and electrostatics in the encoding of relevant chemical information. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 661–686.
(43) Vainio, M. J.; Puranen, J. S.; Johnson, M. S. ShaEP: Molecular overlay based on shape and electrostatic potential. *J. Chem. Inf. Model.* **2009**, *49*, 492–502.
(44) Liu, X. F.; Ouyang, S. S.; Yu, B. A.; Liu, Y. B.; Huang, K.; Gong, J. Y.; Zheng, S. Y.; Li, Z. H.; Li, H. L.; Jiang, H. L. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.* **2010**, *38*, W609–W614.
(45) Lu, W. Q.; Liu, X. F.; Cao, X. W.; Xue, M. Z.; Liu, K. D.; Zhao, Z. J.; Shen, X.; Jiang, H. L.; Xu, Y. F.; Huang, J.; Li, H. L. SHAFTS: A Hybrid Approach for 3D Molecular Similarity Calculation. 2. Prospective Case Study in the Discovery of Diverse p90 Ribosomal S6 Protein Kinase 2 Inhibitors To Suppress Cell Migration. *J. Med. Chem.* **2011**, *54*, 3564–3574.
(46) Liu, X. F.; Bai, F.; Ouyang, S. S.; Wang, X. C.; Li, H. L.; Jiang, H. L. Cyndi: A multi-objective evolution algorithm based method for bioactive molecular conformational generation. *BMC Bioinf.* **2009**, *10*, 101.
(47) Greene, J.; Savoj, H.; Sprague, P.; Teig, S. Chemical function queries for 3D database search. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1297–1308.
(48) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.
(49) von Korff, M.; Freyss, J.; Sander, T. Comparison of ligand- and structure-based virtual screening on the DUD data set. *J. Chem. Inf. Model.* **2009**, *49*, 209–231.
(50) Irwin, J. J. Community benchmarks for virtual screening. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 193–199.
(51) Good, A. C.; Oprea, T. I. Optimization of CAMD techniques 3. Virtual screening enrichment studies: a help or hindrance in tool selection? *J Comput.-Aided Mol. Des.* **2008**, *22*, 169–178.
(52) Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308–1315.
(53) Barker, E. J.; Gardiner, E. J.; Gillet, V. J.; Kitts, P.; Morris, J. Further development of reduced graphs for identifying bioactive compounds. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 346–356.
(54) Jahn, A.; Hinselmann, G.; Fechner, N.; Zell, A. Optimal assignment methods for ligand-based virtual screening. *J. Cheminf.* **2009**, *1*, 14.
(55) Cleves, A. E.; Jain, A. N. Robust ligand-based modeling of the biological targets of known drugs. *J. Med. Chem.* **2006**, *49*, 2921–2938.
(56) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759–4767.
(57) Nicholls, A. What do we know and when do we know it? *J Comput.-Aided Mol. Des.* **2008**, *22*, 239–255.
(58) Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133–139.
(59) Hawkins, P. C. D.; Warren, G. L.; Skillman, A. G.; Nicholls, A. How to do an evaluation: pitfalls and traps. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 179–190.
(60) Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection - What can we learn from earlier mistakes? *J Comput.-Aided Mol. Des.* **2008**, *22*, 213–228.
(61) Low, C. M. R.; Buck, I. M.; Cooke, T.; Cushnir, J. R.; Kalindjian, S. B.; Kotecha, A.; Pether, M. J.; Shankley, N. P.; Vinter, J. G.; Wright, L. Scaffold hopping with molecular field points: Identification of a cholecystokinin-2 (CCK2) receptor pharmacophore and its use in the design of a prototypical series of pyrrole- and imidazole-based CCK2 antagonists. *J. Med. Chem.* **2005**, *48*, 6790–6802.
(62) Clark, R. D.; Webster-Clark, D. J. Managing bias in ROC curves. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 141–146.
(63) Kirchmair, J.; Distinto, S.; Markt, P.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Wolber, G. How to optimize shape-based virtual

screening: Choosing the right query and including chemical information. *J. Chem. Inf. Model.* **2009**, *49*, 678–692.

(64) Venkatraman, V.; Perez-Nueno, V. I.; Mavridis, L.; Ritchie, D. W. Comprehensive comparison of ligand-based virtual screening tools against the DUD data set reveals limitations of current 3D methods. *J. Chem. Inf. Model.* **2010**, *50*, 2079–2093.

(65) Sheridan, R. P.; Kearsley, S. K. Why do we need so many chemical similarity search methods? *Drug Discovery Today* **2002**, *7*, 903–911

(66) Zhang, Q.; Muegge, I. Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: Ranking, voting, and consensus scoring. *J. Med. Chem.* **2006**, *49*, 1536–1548.

(67) McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J. F.; Cornell, W. D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 1504–1519.

(68) Vogt, M.; Stumpfe, D.; Geppert, H.; Bajorath, J. Scaffold Hopping Using two-dimensional fingerprints: True potential, black magic, or a hopeless endeavor? Guidelines for virtual screening. *J. Med. Chem.* **2010**, *53*, 5707–5715.