JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

The Development of Target-Specific Pose Filter Ensembles To Boost Ligand Enrichment for Structure-Based Virtual Screening

Jie Xia,[†] Jui-Hua Hsieh,[‡] Huabin Hu,[†] Song Wu,[*,†] and Xiang Simon Wang[*,§]

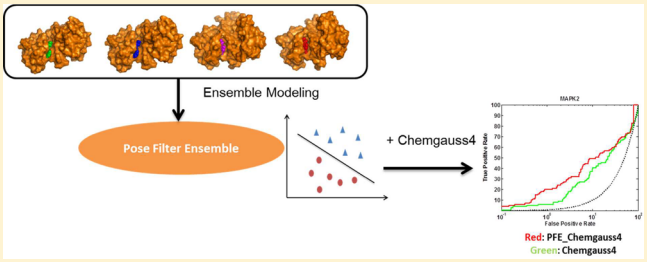
[†]State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Department of New Drug Research and Development, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China

[‡]Kelly Government Solutions, Research Triangle Park, North Carolina 27709, United States

[§]Molecular Modeling and Drug Discovery Core Laboratory for District of Columbia Center for AIDS Research (DC CFAR), Department of Pharmaceutical Sciences, College of Pharmacy, Howard University, Washington, D.C. 20059, United States

Supporting Information

ABSTRACT: Structure-based virtual screening (SBVS) has become an indispensable technique for hit identification at the early stage of drug discovery. However, the accuracy of current scoring functions is not high enough to confer success to every target and thus remains to be improved. Previously, we had developed binary pose filters (PFs) using knowledge derived from the protein−ligand interface of a single X-ray structure of a specific target. This novel approach had been validated as an effective way to improve ligand enrichment. Continuing from it, in the present work we attempted to incorporate knowledge collected from diverse protein−ligand interfaces of multiple crystal structures of the same target to build PF ensembles (PFEs). Toward this end, we first constructed a comprehensive data set to meet the requirements of ensemble modeling and validation. This set contains 10 diverse targets, 118 well-prepared X-ray structures of protein−ligand complexes, and large benchmarking actives/decoys sets. Notably, we designed a unique workflow of two-layer classifiers based on the concept of ensemble learning and applied it to the construction of PFEs for all of the targets. Through extensive benchmarking studies, we demonstrated that (1) coupling PFE with Chemgauss4 significantly improves the early enrichment of Chemgauss4 itself and (2) PFEs show greater consistency in boosting early enrichment and larger overall enrichment than our prior PFs. In addition, we analyzed the pairwise topological similarities among cognate ligands used to construct PFEs and found that it is the higher chemical diversity of the cognate ligands that leads to the improved performance of PFEs. Taken together, the results so far prove that the incorporation of knowledge from diverse protein−ligand interfaces by ensemble modeling is able to enhance the screening competence of SBVS scoring functions.



INTRODUCTION

Virtual screening (VS) is a powerful technique for early-stage drug discovery, as it is able to rapidly and inexpensively discover active compounds of novel scaffolds for a specific target from large-scale chemical libraries.[1] Structure-based virtual screening (SBVS) requires three-dimensional structure(s) of the target and employs molecular docking/scoring to identify hit compounds that fit the binding site.[2,3] As a result of the rapid growth in the number of high-quality macromolecular structures in recent years,[4] SBVS is now playing a pivotal role in modern drug discovery and has led to many cases of success in hit identification.[5,6]

In the process of VS, a docking algorithm first samples possible poses of each compound within the binding site of a target, then a scoring function predicts binding free energy of each ligand pose.[7] Currently, it appears that for most targets there is no problem for a docking algorithm to generate poses that cover native-like ones, i.e., poses close to the native pose in the X-ray structure of protein−ligand complex.[8,9] However, it remains a challenge for current scoring functions to accurately predict binding affinity.[8,10−12] The inaccurate scoring functions mistakenly score and rank pose decoys/native-like poses and binding decoys/true binders,[13] which results in their limited ligand enrichment (i.e., screening power).[14] To improve the hit rate in real-world drug screening, it is necessary to develop new methodologies to improve the accuracy of current scoring functions, in particular the power of enriching ligands of diverse scaffolds at the top of the rank-ordered result list.

Visual inspection after molecular docking and scoring has been an empirical but effective way to increase the hit rate, as it has appeared in the protocols of many successful stories reported to date.[15−21] This mainly refers to manual recognition of native-like poses for each compound according to its binding modes and the subsequent pose selection from them according to the scores. This process is time-consuming, and thus, the

Received: December 12, 2016
Published: May 16, 2017

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414−1425

Table 1. Summary of Selected Targets, Refined X-ray Structures of Protein−Ligand Complexes for Building Pose Filter Ensembles (PFE), Representative Structures in DUD-E and Their Benchmarking Parameters (i.e., Actives/Decoys) in the Current Study

target class (no.)	target name	abbr.	structures/PFE no.	structures/PFE name	structure/DUD-E	no. of actives/decoys[a]
other enzymes (33)	adenosine deaminase	ADA	19	1a4l, 1a4m, 1add, 1krm, 1ndv, 1ndy, 1ndz, 1o5r, 1qxl, 1uio, 1uip, 1uml, 1v79, 1wxy, 1wxz, 2e1w, 2z7g, 3iar, 3km8	2e1w	84/5443
	HMG-CoA reductase	HMDH	18	1dq8, 1hw8, 1hw9, 1hwi, 1hwj, 1hwl, 2q1l, 2q6b, 2q6c, 2r4f, 3bgl, 3cct, 3ccw, 3ccz, 3cd0, 3cd5, 3cda, 3cdb	3ccw	166/8735
kinases (25)	MAP kinase-activated protein kinase 2	MAPK2	10	1ny3, 3a2c, 3ka0, 3kc3, 3kga, 3m2w, 3m42, 3r2b, 3r2y, 3wi6	3m2w	99/6130
	insulin-like growth factor I receptor	IGF1R	10	2oj9, 3d94, 3f5p, 3i81, 3lvp, 3nw5, 3nw6, 3nw7, 3o23, 3qqu	2oj9	144/9272
proteases (13)	leukotriene a4 hydrolase	LKHA4	17	2r59, 3b7r, 3b7u, 3cho, 3chp, 3chq, 3chs, 3fh5, 3fh7, 3fh8, 3fuf, 3fui, 3fuk, 3ful, 3fum, 3fun, 3u9w	3chp	166/9426
nuclear receptors (11)	progesterone receptor	PRGR	13	1a28, 1e3k, 1sqn, 1sr7, 1zuc, 2w8y, 3d90, 3g8o, 3hq5, 3kba, 3zra, 3zrb, 4a2j	3kba	274/15431
cytochromes 450 (1)	cytochrome P450 3A4	CP3A4	6	3nxu, 3ua1, 4i4h, 4k9t, 4k9v, 4k9w	3nxu	170/11781
ion channels (2)	glutamate receptor ionotropic kainate 1	GRIK1	13	1vso, 2f35, 2ojt, 2pbw, 2qs1, 2qs2, 2qs3, 2qs4, 2znt, 2znu, 3fvo, 3gbb, 4dld	1vso	96/6537
GPCRs (3)	β2 adrenergic receptor	ADRB2	6	2rh1, 3d4s, 3ny8, 3ny9, 4lde, 4ldl	3ny8	229/14986
miscellaneous (5)	fatty acid binding protein adipocyte	FABP4	6	1tou, 1tow, 2nnq, 3fr2, 3fr4, 3fr5	2nnq	42/2711

[a]Ligands used for benchmarking calculations.

number of poses/compounds that human experts can realistically inspect is limited.[22] Besides, the decision made from visual inspection is normally subjective, and the outcome varies with the level of experience of human experts. The SBVS community has been working on this issue by developing automated pose filters (PFs)/classifiers to assist rational selection of potential hits. Such PFs exclude undesired poses on the basis of knowledge of protein−ligand interactions that exist in X-ray structures of a specific target.[22−24] The early PFs were constructed by directly defining essential protein−ligand interactions as constraints (e.g., hydrogen-bond constraints)[23] or applying pharmacophore-feature-like constraints derived from protein−ligand interactions.[24] The most popular class of those PFs is interaction fingerprints, and their development is normally accompanied by encoding protein−ligand interactions into a one-dimensional binary string. The pairwise similarity of bit strings is used to quantitatively measure pose similarity. The highly cited interaction-fingerprint-based PFs include SIFt[25] and w-SIFt,[26] APIF,[27] TIFP,[28] PLIF implemented in Molecular Operating Environment,[29] and SPLIF.[8] The third class of PFs is based on descriptors that define protein−ligand interfaces. Tropsha and co-workers designed geometrical chemical descriptors, i.e., Pauling electronegativity based on Delaunay tessellation (ENTess) descriptors,[30] then updated them to obtain protein−ligand pairwise atomic maximal charge transfer potential based on Delaunay tessellation (PL/MCT-tess) descriptors.[31,32] Initially, these types of descriptors were used for quantitative structure−binding affinity relationship (QSBAR) modeling.[30,31] We recently employed PL/MCT-tess descriptors to characterize protein−ligand interactions of multiple docking poses. On that basis, we applied support vector machine (SVM) algorithms to build target-specific PFs that can differentiate native-like poses from pose decoys. The add-on of PFs has consistently improved ligand enrichment of a force-field-based scoring function, i.e., MedusaScore.[32] This class of knowledge-based PFs was also evaluated during the CSAR 2013/2014 benchmark exercise and succeeded in native-like pose identification.[33,34]

Incorporating multiple protein−ligand interfaces into molecular docking is also critical to the success of SBVS. The most widely used strategy, i.e., ensemble docking against multiple receptor structures/conformations, deals with multiple protein−ligand interfaces in a computationally inexpensive way. A multitude of publications have introduced feasible methods for ensemble docking and demonstrated their potentials in improving ligand enrichment.[35−40] Also, cases of success in hit identification using ensemble docking have been reported.[19,37,41−44] In light of the promising outcome of ensemble docking, we formulated a hypothesis that the employment of multiple X-ray structures for constructing PFs could bring forth correct pose classification and eventually improve ligand enrichment. In this study, we selected 10 diverse targets from the target list of DUD-E as exemplified cases and compiled 10 ready-to-use data sets that contain multiple X-ray structures of protein−ligand complexes and benchmarking actives/decoys for each target. Using these data sets, we built a novel and thorough workflow to incorporate knowledge collected from diverse protein−ligand interfaces, i.e., to incorporate multiple protein−ligand interfaces into the construction of PFs. Then we applied the new type of PF, called a PF ensemble (PFE), to an empirical/regression-based scoring function,[45] i.e., Chemgauss4, which was different from our previously used MedusaScore, a force-field-based scoring function.[32] For performance evaluation, we performed comparative studies on ligand enrichment of Chemgauss4 prior to and after the use of PFE/PF. The current study is a continuation of our prior work on PF modeling, and it aims to answer three unexplored questions: (1) how to integrate knowledge from diverse protein−ligand interfaces into the construction of PFs in order to improve ligand enrichment; (2) whether PFs work for empirical scoring functions (e.g., Chemgauss4) apart from force-field-based scoring functions; and (3) whether PFE performs better than PF alone.

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414−1425

METHODS

Target Selection. This study includes two consecutive parts, i.e., construction of target-specific PFEs and assessment of their performance in ligand enrichment for SBVS. The former requires multiple X-ray structures of protein−ligand complexes, while the latter depends on benchmarking data sets. Since the sc-PDB database collects druggable binding sites of high-quality cocrystal structures that are ready to use (http://bioinfo-pharma.u-strasbg.fr/scPDB/, accessed June 2015),[46] it became our source of protein−ligand complexes for construction of PFEs. In addition, DUD-E includes gold-standard benchmarking data sets that are ideal for molecular docking,[47,48] and thus, it was used for performance evaluation of PFEs. To meet the above-mentioned data requirements, i.e., both protein−ligand complexes and benchmarking data sets, only the overlapping targets between DUD-E and sc-PDB were selected for further analysis. Because three targets out of 102, i.e., histone deacetylase 2 (HDAC2), trypsin I (TRY1), and tryptase β1 (TRYB1) were not included by sc-PDB (cf. Table S1), they were excluded from our list of targets.

Other criteria for target selection are as follows. (1) To build a PFE, there must be at least two available X-ray structures of protein−ligand complexes. Toward this end, all of the UniProt accession codes covered by DUD-E ligands (i.e., uniprot.txt) were obtained from DUD-E (http://dude.docking.org/targets, accessed June 2015). Then their UniProt names were retrieved from UniProt (http://www.uniprot.org/, accessed June 2015). With those Uniprot names as input, the X-ray structures of protein−ligand complexes in sc-PDB for each target were counted (cf. Table S1). Six targets for which only one X-ray structure was available, i.e., monoamine oxidase B (AOFB), cytochrome P450 2C9 (CP2C9), CXC chemokine receptor type 4 (CXCR4), dopamine D3 receptor (DRD3), β-glucocerebrosidase (GLCM), and protein kinase C β (KPCB), were excluded. (2) The remaining 93 targets cover eight classes, i.e., other enzymes, kinases, proteases, nuclear receptors, miscellaneous, GPCRs, ion channels and cytochromes P450. To construct and test PFEs on diverse targets, one or two representative targets from each class were selected as example cases. To be specific, for the classes that include over 20 targets, i.e., other enzymes and kinases, two targets were selected. For the other classes, i.e., proteases, nuclear receptors, cytochromes P450, ion channels, GPCRs, and miscellaneous, only one target was selected. Table 1 lists the 10 selected targets, including adenosine deaminase (ADA) and HMG-CoA reductase (HMDH) for other enzymes, MAP kinase-activated protein kinase 2 (MAPK2) and insulin-like growth factor I receptor (IGF1R) for kinases, leukotriene A4 hydrolase (LKHA4) for proteases, progesterone receptor (PRGR) for nuclear receptors, cytochrome P450 3A4 (CP3A4) for cytochromes P450, glutamate receptor ionotropic kainate 1 (GRIK1) for ion channels, β2 adrenergic receptor (ADRB2) for GPCRs, and fatty acid binding protein adipocyte (FABP4) for miscellaneous.

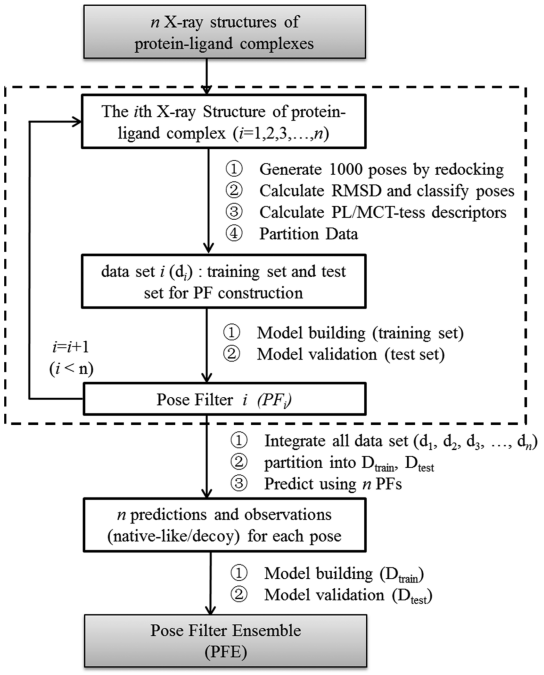
Preparation and Selection of Protein−Ligand Complexes. For each selected target, all available X-ray structures of protein−ligand complexes that correspond to the UniProt names (cf. Table S1) were retrieved from sc-PDB. Then the structures of the protein and its cognate ligands were downloaded. Although the protein structures in sc-PDB have been curated and thus are readily compatible with molecular docking,[46] additional steps of protein preparation and selection were performed prior to the construction of PFEs: (1) Chains not involved in protein−ligand interactions, e.g., T4 lysozyme fusion protein (ADRB2) were removed. (2) All water molecules were excluded from protein structures, but essential cofactors that interact with the cognate ligand in the binding site (e.g., zinc ions in ADA and LKHA4) were preserved. (3) Each protein structure and its cognate ligand were assembled into a protein−ligand complex. With the protein structure listed in the DUD-E benchmarking set as a reference, all of the X-ray structures of protein−ligand complexes were superimposed to check whether all of the cognate ligands were bound to the same site. The complexes in which the cognate ligand was bound to a site different from the reference (e.g., an allosteric site) were excluded (e.g., IGF1R, PDB entry 3lw0; cf. Table S2). (4) In order to further check protein structures, we performed redocking of each cognate ligand against its binding site using OMEGA (version 2.5.1.4; OpenEye Scientific Software)[49] and FRED (now OEDocking, version 3.0.1; OpenEye Scientific Software)[50−52] with default settings. For those structures on which FRED failed to generate docking poses because of problems with hydrogen atoms, hydrogen atoms were readded using Discovery Studio (version 2.5; Accelrys Software), and redocking was repeated afterward. In the current study, the readdition of hydrogen atoms and the repetition of the docking process were performed for the crystal structures 1ndv, 1ndy, 1ndz, 1qxl, and 2e1w (ADA) to make them feasible for molecular docking. Through the above-mentioned steps, the structures of proteins and ligands were prepared for model building.

Construction of Target-Specific PFEs. *General Workflow.* The input of the workflow is n X-ray structures of protein−ligand complexes, while the output is a constructed PFE that can classify poses. Herein, n represents the number of structures per PFE (cf. Tables 1 and S2). The resulting PFE is a scheme of a two-layer chained classifier, where the pose classes predicted by n first-layer models (i.e., n PFs) constitute the input to a second-layer model. Accordingly, the workflow consists of two essential steps: (1) building the first-layer models, i.e., n SVM-based PFs, where each model makes use of knowledge from one X-ray structure alone, and (2) building a second-layer model that uses pose classes predicted using all of the first-layer models as input (cf. Scheme 1)

Building the First-Layer Model/PF Based on Each X-ray Structure. The protocol to build PFs in this study was a modified version of our previous method for building target-specific PFs.[32] The steps of the protocol are described as follows. (1) OMEGA and FRED were applied to generate 1000 geometric poses. To be specific, OMEGA was employed to build a multiconformer database for each cognate ligand, for which its maximum number of conformations was set to the default value of 200. FRED was used to dock the cognate ligand from its multiconformer database into the binding site defined by the cognate ligand itself. As a result, the 1000 top-scoring poses were retained for further analysis. (2) "RMSD Calculator" implemented in Pipeline Pilot (version 7.5; Accelrys Software) was used to calculate heavy-atom root-mean-square deviation (RMSD) values between 1000 poses and the native pose (i.e., the pose in the X-ray structure) of the cognate ligand. According to the RMSD values, the 1000 poses were classified into two classes, i.e., native-like poses (RMSD $\leq$ 4 Å) and pose decoys (RMSD > 4 Å). The X-ray structures for which no more than 10 native-like poses were generated were not included in the structures/PFE list (e.g., 1v7a (ADA), 1dqa

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414−1425

Scheme 1. Workflow To Build a Pose Filter Ensemble (PFE); The Major Steps To Build a PF from an Individual X-ray Structure Are Shown in the Dashed Box



(HMDH), 1w0g and 2v0m (CP3A4); cf. Table S2). (3) PL/MCT-tess descriptors were calculated using our in-house program ENTESS (now available at https://github.com/moggces, accessed October 2016) to characterize the 1000 protein−ligand interfaces from generated docking poses of the cognate ligand. The program uses Delaunay tessellation to partition each protein−ligand interface into Delaunay tetrahedra.[30] Each type of Delaunay tetrahedron is then characterized by pairwise atomic potentials based on maximal charge transfer (MCT).[53] MCT determines the maximal electron flow between the donor and acceptor atoms at the protein−ligand interface based on conceptual DFT.[53,54] The value of the PL/MCT-tess descriptor for a specific tetrahedron type is calculated according to eq 1:

$$\text{PL/MCT-tess}_m = \sum_{k=1}^{n}\left(\sum_{p}^{1-3}\sum_{l}^{1-3}\text{MCT}_p \times \text{MCT}_l/d_{pl}\right)_k \quad (1)$$

In this equation, m is the index of an individual descriptor type, i.e., the tetrahedron type (m = 1, 2, 3, ..., 554), PL/MCT-tess$_m$ represents the potential of the mth tetrahedron type, n is the number of occurrences of the mth tetrahedron type in the protein−ligand complex, k is the index of tetrahedra for the mth tetrahedron type (k = 1, 2, ..., n), p is the index of a vertex represented by a protein atom in the interfacial Delaunay tetrahedron, l is the index of a vertex represented by a ligand atom, and d_{pl} is the corresponding distance between the pth protein atom and the lth ligand atom. Since the PL/MCT-tess calculation is based on each protein−ligand interfacial tetrahedron, the maximal number of protein atoms or ligand atoms is 3 (i.e., p = 1, 2, 3 and l = 1, 2, 3).

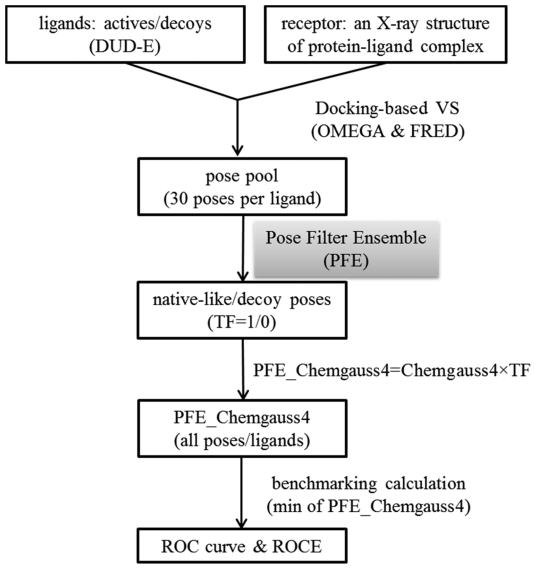
(4) According to the ratio of native-like poses to decoy poses, the data set was defined as either balanced (ratio ≤ 2) or unbalanced (ratio > 2). Each balanced data set was randomly partitioned into a training set for model building (80% of the poses) and test set for model validation (20% of the poses) using "cvpartition" (i.e., the function that creates a random partition on a set of data of a specified size) in MATLAB (version 7.6.0.324). For each unbalanced data set composed of a major class and a minor class, a downsampling strategy was applied before data partition. Let p and q be the total numbers of poses in the major and minor classes, respectively. First, each pose in the major class was characterized by its Euclidean distance to its nearest neighbor in the minor class with respect to PL/MCT-tess descriptors. Then all p poses in the major class were sorted by Euclidean distances. The approximate value of the distance for the qth pose was set as a threshold, and those poses in the major class whose distances were no more than the threshold were selected as a new counterpart of the minor class. In this way, a new balanced data set was constructed. Subsequently, random data partition was performed to generate the training and test sets. (5) The open-source program LibSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm/, accessed June 2015) was used to build and validate the base binary classifier, i.e., the PF.[55] On the basis of the training set, a grid search (grid.py) on values of the parameters γ and C for the radial basis function (RBF) was performed using fivefold cross-validation (CV). The output from the grid search, i.e., the values of the parameters that led to the highest CV accuracy, was picked and used to build the PF. For model validation, the PF was then applied to predict pose classes in test set.

Building Second-Layer Model/PFE Using Predictions from PFs as Input. A large data set (D) was built by integrating all of the data sets previously used for PF construction (d_i). It was also randomly partitioned into a training set (D_{train}, 80%) and a test set (D_{test}, 20%). On the basis of PL/MCT-tess descriptors, the potential class of each pose in D_{train}/D_{test} (i.e., native-like or decoy) was predicted by all of the PFs, then an array of predicted pose classes from multiple PFs became binary descriptors for PFE modeling. The same grid search (grid.py) using LibSVM was performed on D_{train} to search for optimal values of the parameters γ and C. With these parameters, the second-layer model was built and validated on D_{test}. This model was an ensemble that used n first-layer models (i.e., the PFs) as base classifiers and thus is called a PFE.

Application of PFEs to SBVS and Benchmarking Study. *General Workflow.* To test the potential effect of a PFE on SBVS, the PFE was integrated with a structure-based empirical scoring function implemented in the docking program FRED, i.e., Chemgauss4. Then the ligand enrichment of Chemgauss4 after the use of the PFE was assessed via a benchmarking study. Here "ligand enrichment" means the capacity to enrich actives to the top rank of a screening list by discriminating actives from decoys/random molecules.[48] The detailed workflow is shown in Scheme 2. First, the benchmarking set composed of true actives and binding decoys for the selected target was downloaded from http://dude.docking.org/ (accessed June 2015). To avoid potential bias in the benchmarking study, the true actives that overlapped with the cognate ligands used to train the PFE were excluded from the benchmarking set (cf. Table S3). With the cleaned benchmarking set as a small-scale chemical library, docking-based VS was performed against the reference structure for that target (cf. structure/DUD-E in Table 1) using OMEGA and FRED. The parameters for OMEGA were set to the default values. In docking simulations using FRED, a maximum of 30 docking poses for each ligand were generated, and all of the poses were scored by the scoring function of Chemgauss4,

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414−1425

Scheme 2. Coupling of a Pose Filter Ensemble (PFE) to Chemgauss4 for the Current Benchmarking Studies



which resulted in a large pool of poses. Second, the potential class (TF) of each pose, i.e., native-like (TF = 1) or pose decoy (TF = 0), was predicted by PFE. A new score named "PFE-coupled Chemgauss4" (PFE_chemgauss4) was generated by simply multiplying Chemgauss4 by TF. Lastly, the pose with the minimum PFE_chemgauss4 score was retained for each ligand. On the basis of a rank-ordered list of PFE_chemgauss4 scores for all of the true actives and binding decoys, the ligand enrichment of PFE_Chemgauss4 was calculated.

On the basis of specific aims in this study, the second and the last steps of the original workflow were replaced by our corresponding approaches. In order to assess the effect of the PFE, the prior scoring function, i.e., Chemgauss4 without using the PFE, was also benchmarked. In that benchmarking study, the pose with the minimum Chemgauss4 score instead of PFE_Chemgauss4 score was retained for each ligand. To explore the potential advantage of ensemble modeling, the single PF derived from the protein structure for docking-based VS was used as a pose classifier and coupled with Chemgauss4 to yield a score named "PF-coupled Chemgauss4" (PF_Chemgauss4). The ligand enrichment of PF_Chemgauss4 was then obtained for comparative analysis.

Metrics for Benchmarking Calculations. Parameters from the receiver operating characteristic (ROC) curve, i.e., the area under the ROC curve (AUC) and the ROC enrichment (ROCE) at 1%/0.5% were the main metrics for the benchmarking calculations. The AUC value is an indicator of the overall performance in ligand enrichment, while the value of $ROCE_{1\%}/ROCE_{0.5\%}$ is used to measure early enrichment. ROCE is defined as the ratio of the true positive rate to the false positive rate at a given percent of recovered known decoys (e.g., 1% for $ROCE_{1\%}$). Its value can be obtained by calculating the slope at each point from the ROC curve.[56,57] Apart from the major metrics, the percentage of actives recovered (AR%) was also calculated in order to show the performance of recognizing actives.

To explore the potential of each approach in novel scaffold identification, Murcko–Bemis frameworks[58] of the actives recovered at the early stage (i.e., 1% of binding decoys) were generated using the "Generate Fragments (Murcko Assemblies)" component in Pipeline Pilot, and only the unique ones were counted. On the basis of the numbers of actives and unique Murcko scaffolds, the overlaps of PFE_Chemgauss4 with PF_Chemgauss4 and with Chemgauss4 were also counted. For comparative analysis between PFE_Chemgauss4 and PF_Chemgauss4, two metrics were designed and applied, i.e., unidentified% and unique%. The former represents the percentage of actives/scaffolds recovered by Chemgauss4 but not identified after the use of PFE/PF. The latter indicates the percentage of additional actives/scaffolds identified after the use of PFE/PF. By definition, smaller values of unidentified% and greater values of unique% indicate desirable performance of an approach.

RESULTS AND DISCUSSION

Comprehensive Data Set for Ensemble Modeling and Benchmarking. As shown in Table 1, the data set we compiled and used for ensemble modeling and benchmarking includes 10 representative targets that cover eight classes. To be specific, ADA and HMDH represent the class of other enzymes, while MAPK2 and IGF1R are representative targets for the class of kinases. LKHA4, PRGR, CP3A4, GRIK1, ADRB2, and FABP4 are example cases for the classes of proteases, nuclear receptors, cytochromes P450, ion channels, GPCRs, and miscellaneous, respectively. The subset for each target consists of two indispensable components: (1) multiple X-ray structures of protein–ligand complexes for building the PFE (i.e., structures/PFE) and (2) actives/decoys for the benchmarking study. The numbers of protein–ligand complexes used to construct PFEs range from 6 to 19 across 10 different targets. For seven out of 10 targets, the number is no less than 10. For each target, all of the cognate ligands from the complexes bind to the same site, while the protein atom coordinates of that binding site vary along with different chemical structures of cognate ligands (cf. Figure S1). In addition, all of the included X-ray structures are feasible for molecular docking by FRED and able to generate native-like poses. The other component, i.e., actives/decoys of this data set, was retrieved from the ready-to-use benchmarking set in DUD-E and processed to ensure that there were no overlaps of actives with the cognate ligands.

Though all of the associated data were originally obtained from the publicly available databases sc-PDB and DUD-E, we recompiled them into a ready-to-use data set that includes both the structures of the protein–ligand complexes and their corresponding benchmarking sets. In this study, we used it to build the PFE and evaluate its effect on SBVS. We anticipate that the public access to such data set by the community would facilitate the methodology development associated with ensemble modeling and benchmarking.

Native-like versus Decoy Pose Classifiers: PFs and PFEs. *Characteristics of the PFs.* Using the program LibSVM, we built multiple PFs based on all of the included X-ray structures of protein–ligand complexes for each target. All of the details for model building and validation are shown in Table S4, including the training and test sets, the parameters for model building, and model performance in pose classification. The following points are worthy of note: (1) The distributions of native-like poses and pose decoys in both the training set and test set are balanced, demonstrating that our downsampling strategy was effective in converting an unbalanced distribution to a balanced distribution. (2) As indicated by the values of the parameters C and γ, only one RBF model was kept as a pose

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414–1425

Table 2. Ensemble Modeling of Pose Filters (PFs): Parameter Values, CV Accuracy, and Prediction Accuracy for Each PFE

	D_{train}		RBF model			D_{test}		
targets	no. of native-like poses	no. of pose decoys	C	γ (10^{-4})	CV accuracy (%)	no. of native-like poses	no. of pose decoys	prediction accuracy (%)
ADA	2682	2977	32	1250.0	96.3	648	766	96.1
HMDH	4158	4436	2	5000.0	88.6	977	1171	88.5
MAPK2	2991	2977	8	1250.0	84.6	775	717	84.4
IGF1R	1969	1947	512	312.5	86.1	489	490	87.3
LKHA4	3677	3145	128	1250.0	86.7	911	794	85.7
PRGR	4155	3571	2048	312.5	88.1	1060	871	88.0
CP3A4	961	1014	8	78.1	93.1	244	249	93.1
GRIK1	3275	2729	512	78.1	85.8	840	660	86.9
ADRB2	2084	2256	2048	312.5	85.1	496	589	87.2
FABP4	1525	1496	2	78.1	88.3	361	394	88.7

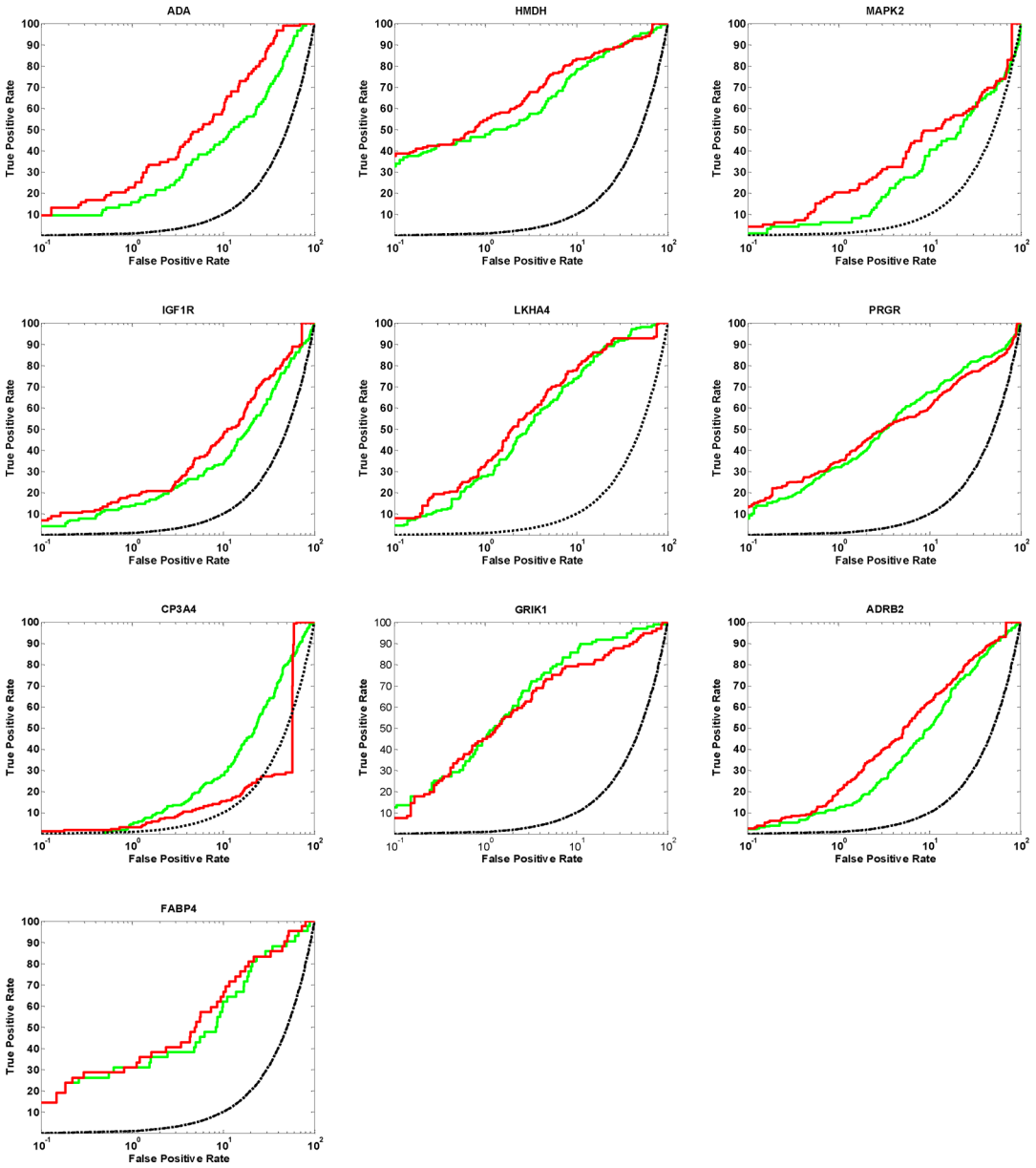


Figure 1. Ligand enrichments represented by ROC curves. Color code: red, PFE_Chemgauss4; green, Chemgauss4; dotted black, random distribution.

classifier for each X-ray structure. This is distinct from our earlier strategy that kept all models of eligible CV accuracy.[32] To save computational time, we performed a grid search in this study to find the optimal values of C and γ and then used them

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414−1425

Table 3. Ligand Enrichments for Chemgauss4, PFE_Chemgauss4, and PF_Chemgauss4 As Measured by the Metrics $ROCE_{1\%}$, $ROCE_{0.5\%}$, AUC, and Percentage of Actives Recovered (AR%)

	$ROCE_{1\%}$			$ROCE_{0.5\%}$			AUC			AR%		
target	C4[a]	PFE_C4[a]	PF_C4[a]	C4	PFE_C4	PF_C4	C4	PFE_C4	PF_C4	C4	PFE_C4	PF_C4
ADA	15.48	22.62	39.29	23.81	35.72	57.16	0.78	0.88	0.95	100.00	98.81	96.43
HMDH	47.59	54.22	42.17	89.17	90.37	78.32	0.90	0.92	0.90	100.00	93.37	96.99
MAPK2	6.06	20.20	31.31	10.10	22.23	60.62	0.65	0.71	0.67	100.00	82.83	52.53
IGF1R	13.89	18.75	20.14	23.61	26.39	29.17	0.72	0.79	0.83	100.00	88.89	87.50
LKHA4	27.71	33.13	36.75	33.74	44.58	44.58	0.91	0.90	0.89	100.00	92.77	88.55
PRGR	32.12	34.67	27.37	48.18	55.48	39.42	0.83	0.80	0.74	100.00	93.43	71.90
CP3A4	4.71	2.94	4.71	1.18	3.53	3.53	0.71	0.55	0.37	100.00	28.24	7.65
GRIK1	44.79	44.79	34.38	58.35	68.76	45.84	0.94	0.90	0.80	100.00	98.96	68.75
ADRB2	12.23	20.52	15.72	17.47	19.21	19.21	0.81	0.85	0.84	100.00	93.01	91.70
FABP4	30.96	30.96	33.34	52.45	57.22	57.22	0.84	0.87	0.71	100.00	100.00	78.57
average	23.55	28.28	28.52	35.81	42.35	43.51	0.81	0.82	0.77	100.00	87.03	74.06

[a]C4, Chemgauss4; PFE_C4, PFE_Chemgauss4; PF_C4, PF_Chemgauss4.

Table 4. Statistics of Actives Recovered at 1% of Binding Decoys by Chemgauss4, PFE_Chemgauss4, and PF_Chemgauss4

	actives			overlap		unidentified%		unique%	
target	C4[a]	PFE_C4[a]	PF_C4[a]	C4 and PFE_C4	C4 and PF_C4	PFE_C4	PF_C4	PFE_C4	PF_C4
ADA	13	19	33	12	13	5.0	0.0	35.0	60.6
HMDH	79	90	70	77	67	2.2	14.6	14.1	3.7
MAPK2	6	20	31	6	5	0.0	3.1	70.0	81.3
IGF1R	20	27	29	17	15	10.0	14.7	33.3	41.2
LKHA4	46	54	60	38	36	12.9	14.3	25.8	34.3
PRGR	88	94	74	78	59	9.6	28.2	15.4	14.6
CP3A4	8	5	8	0	0	61.5	50.0	38.5	50.0
GRIK1	43	43	31	36	21	14.0	41.5	14.0	18.9
ADRB2	28	47	36	22	24	11.3	10.0	47.2	30.0
FABP4	13	13	14	12	12	7.1	6.7	7.1	13.3
average	34	41	39	30	25	8.9	18.8	24.4	29.2

[a]C4, Chemgauss4; PFE_C4, PFE_Chemgauss4; PF_C4, PF_Chemgauss4.

to build the model that led to highest CV accuracy. (3) Out of 118 binary pose classifiers (i.e., PFs), 106 were optimal models for pose prediction, whose values of both CV accuracy and prediction accuracy (PA) were greater than 90%. Among the other 12 PFs, models for seven X-ray structures (i.e., 1ndy, 1uip, and 2e1w for ADA, 3nw5 for IGF1R, and 2f35, 2qs1, and 2qs3 for GRIK1) predict well, as either the CV accuracy or PA for these PFs was greater than 90%. Five models (i.e., 1dq8 and 3cdb for HMDH, 4k9t and 4k9v for CP3A4, and 1tou for FABP4), reached acceptable CV accuracy and PA, both of which were between 70% and 90%. All of the models for each target were used for ensemble modeling.

Characteristics of the PFEs. Table 2 shows the PFEs for the 10 selected targets, their parameter values, and their model performances in pose prediction. As shown in that table, the distributions of native-like poses and pose decoys in all of the data sets for construction of PFEs were balanced. Likewise, only the model of highest CV accuracy was kept for each target. The CV accuracy of every model was greater than 80%, with a maximum of 96.3% for ADA and a minimum of 84.6% for MAPK2. The value of PA (on D_{test}) for each model was close to that of the CV accuracy (on D_{train}). The minimum was 84.4% (for MAPK2), and the maximum was 96.1% (for ADA).

Ligand Enrichment: PFE_Chemgauss4 versus Chemgauss4. *Benchmarking Calculations.* For each target, we performed retrospective SBVS based on a benchmarking set and assessed the ligand enrichment of Chemgauss4 before and after the use of the PFE. Notably, a few ligands (i.e., actives or decoys) failed in pose generation because of their unspecified stereochemistry or could not be docked because their size was unfit for the site, and thus, they were not included in the benchmarking calculations. As shown in Tables 1 and S5, the numbers of actives/decoys used for benchmarking calculations were fewer than the initial numbers listed in DUD-E (http://dude.docking.org/, accessed June 2015). To obtain metrics, we plotted ROC curves (cf. Figure 1) and calculated their parameters (cf. Table 3 and Figure S2) to evaluate the ligand enrichment for different approaches. In practice, only a small percentage of compounds instead of all compounds from VS are submitted for bioassay, and thus, early enrichment is regarded as a significant metric to assess ligand enrichment.[56]

Early Enrichment. As shown in Table 3 and Figure S2, the $ROCE_{1\%}$ values for PFE_Chemgauss4 are greater than those for Chemgauss4 for seven targets (i.e., ADA, HMDH, MAPK2, IGF1R, LKHA4, PRGR, and ADRB2). For these targets, the $ROCE_{0.5\%}$ values for PFE_Chemgauss4 are also greater. For the targets GRIK1 and FABP4, although the $ROCE_{1\%}$ values for PFE_Chemgauss4 are equal to those for Chemgauss4, the $ROCE_{0.5\%}$ values are greater than those for Chemgauss4. For CP3A4, although the $ROCE_{1\%}$ value for PFE_Chemgauss4 is less than that for Chemgauss4, the $ROCE_{0.5\%}$ value is greater. On average across the 10 targets, the values of both $ROCE_{1\%}$ and $ROCE_{0.5\%}$ for PFE_Chemgauss4 are greater than those for Chemgauss4 ($ROCE_{1\%}$, 28.28 vs 23.55; $ROCE_{0.5\%}$, 42.35 vs 35.81). The portion of the ROC curve whose false positive rate is between 0.1% and 1% shows a similar trend (cf. Figure 1). All

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414−1425

Table 5. Statistics of Murcko Scaffolds Recovered at 1% of Binding Decoys by Chemgauss4, PFE_Chemgauss4, and PF_Chemgauss4

	scaffolds			Overlap		unidentified%		unique%	
target	C4[a]	PFE_C4[a]	PF_C4[a]	C4 and PFE_C4	C4 and PF_C4	PFE_C4	PF_C4	PFE_C4	PF_C4
ADA	11	13	19	10	11	7.1	0.0	21.4	42.1
HMDH	77	88	68	75	65	2.2	15.0	14.4	3.8
MAPK2	6	19	26	6	5	0.0	3.7	68.4	77.8
IGF1R	20	27	29	17	15	10.0	14.7	33.3	41.2
LKHA4	46	54	59	38	36	12.9	14.5	25.8	33.3
PRGR	79	80	67	69	54	11.1	27.2	12.2	14.1
CP3A4	8	5	8	0	0	61.5	50.0	38.5	50.0
GRIK1	24	23	17	19	14	17.9	37.0	14.3	11.1
ADRB2	27	46	35	21	23	11.5	10.3	48.1	30.8
FABP4	4	5	6	4	4	0.0	0.0	20.0	33.3
average	30	36	33	26	23	10.0	17.5	25.0	25.0

[a]C4, Chemgauss4; PFE_C4, PFE_Chemgauss4; PF_C4, PF_Chemgauss4.

of the above data, i.e., the ROC curves and corresponding parameters, indicate that the use of PFE improves the early enrichment of Chemgauss4.

We further analyzed the actives recovered at the early stage (i.e., 1% of binding decoys) before and after the application of PFE to Chemgauss4. Also, we generated Murcko−Bemis frameworks to characterize those actives and explored their scaffold diversity. According to Tables 4 and 5, for all of the targets except CP3A4, PFE_Chemgauss4 recovered more or an equivalent number of actives/scaffolds than Chemgauss4. This indicates that PFE_Chemgauss4 was able to identify additional actives/scaffolds that were initially not recovered by Chemgauss4. For the target of CP3A4, PFE_Chemgauss4 identified fewer actives/scaffolds than Chemgauss4. This explains well the situation for CP3A4 that the $ROCE_{1\%}$ value for PFE_Chemgauss4 is less than that for Chemgauss4. However, it is interesting to note that there is no overlap between Chemgauss4 and PFE_Chemgauss4 for this target. As for the overall performance, the average numbers of actives/scaffolds identified by PFE_Chemgauss4 and Chemgauss4 are 41/36 and 34/30, respectively, which demonstrates greater performance because of the use of PFE. On the basis of the above data, it can be concluded that the use of PFE improves early enrichment by identifying additional actives and Murcko scaffolds.

Overall Enrichment. Apart from early enrichment, the ROC AUC values that represent overall enrichment are also listed in Table 3 and shown in Figure S2. As shown by the averages of ROC AUC values across 10 targets for PFE_Chemgauss4 and Chemgauss4 (i.e., 0.82 vs 0.81), Use of the PFE does not have a significant effect on overall enrichment. For nine individual targets, the difference in the AUCs for PFE_Chemgauss4 and Chemgauss4 is trivial. The only significant change appears in CP3A4, and its AUC value drops from 0.71 (Chemgauss4) to 0.55 (PFE_Chemgauss4).

To elucidate a potential mechanism for the change in AUC, we calculated the percentages of actives recovered (AR%) in the whole screening. The average value of AR% across the 10 targets is 87.03%. The AR% value is greater than 90% for seven targets and between 80% and 90% for two targets. The AR% value is extremely low for CP3A4 (28.24%; cf. Table 3 and Figure S2). According to the scoring method of PFE_Chemgauss4, AR% is associated with the prediction of the PFE (i.e., TF = 1/0). If the PFE predicts all of the poses of an active as pose decoys, the active will not be recovered. Therefore, the failure of the PFE to recognize native-like poses of most actives for CP3A4 results in the extremely low AR% and overall enrichment of CP3A4. In summary, addition of the PFE maintains the overall enrichment power of Chemgauss4 in most situations.

Key Factors That Affect the Efficacy of PFEs. As noted above, the efficacy of the PFE varies for different targets. For instance, the PFE performs the best for MAPK2, as it improves the ligand enrichment of Chemgauss4 by 233%. However, it performs the worst for CP3A4, with an increase rate of −38% (cf. Table S6). We hypothesized that the chemical diversity of the cognate ligands used for model building may affect the early enrichment. Thus, we employed function-class fingerprints of maximum diameter 6 (FCFP_6) to characterize those cognate ligands and calculated their pairwise topological similarity in terms of the Tanimoto coefficient (Tc). The average of all of the Tc values, i.e., mean(Tc), was then correlated to $ROCE_{1\%}$ to uncover the potential relationship between chemical diversity and early enrichment. Prior to that, we used Cook's distance to identify the outlier(s) in the data.[59,60] The case order plot clearly shows that MAPK2 is an outlier, as its Cook's distance is much greater than the threshold value, i.e., 3 times the mean Cook's distance (cf. Figure S3). On the basis of the unleveraged data, i.e., nine pairs of mean(Tc)/$ROCE_{1\%}$ values, the Pearson correlation coefficient (r) between the increase rate in early enrichment ($ROCE_{1\%}$) and mean(Tc) is −0.68 (cf. Figure 2). This strong correlation indicates that in general lower pairwise similarity among the cognate ligands results in greater improvement in early enrichment for Chemgauss4. For

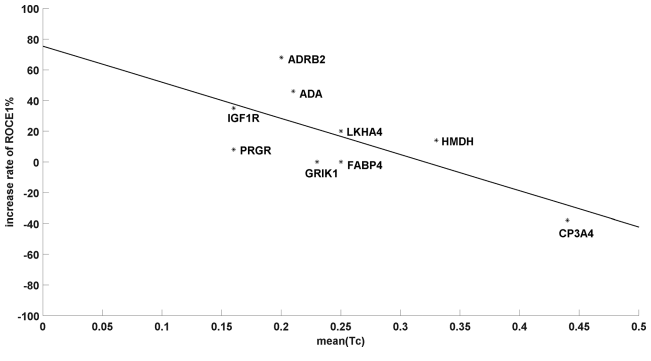


Figure 2. Regression plot between the mean(Tc) value and increase rate of $ROCE_{1\%}$, with a Pearson correlation coefficient of −0.68.

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414−1425

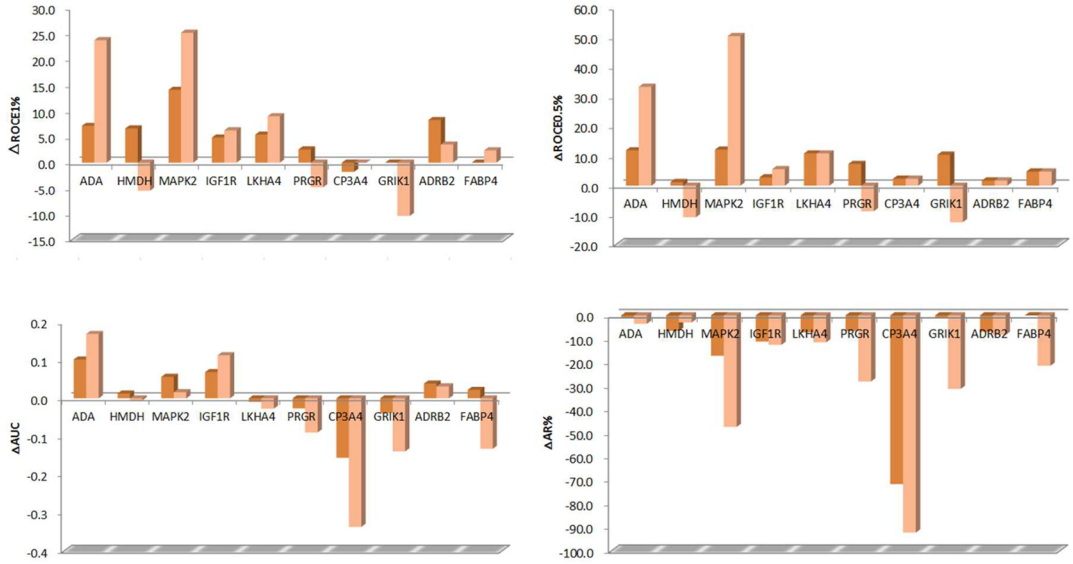


Figure 3. Comparative analysis of potential effects of PFEs and PFs in terms of $\Delta ROCE_{1\%}$, $\Delta ROCE_{0.5\%}$, ΔAUC, and ΔAR%. Color code: dark orange, PFE; light orange, PF.

instance, the mean(Tc) value for ADRB2 (0.20) is much less than that for HMDH (0.33), and thus, its improvement of $ROCE_{1\%}$ is much greater (increase rate = 68% vs 14%) correspondingly. The above data reveal that a main factor that influences the efficacy of the PFE is the chemical diversity of the cognate ligands used for ensemble modeling.

Advantages of Ensemble Modeling: PFEs versus PFs. *Contents of Benchmarking Outcomes.* In order to explore potential advantages of ensemble modeling, we benchmarked PF_Chemgauss4 and conducted extensive comparison studies with PFE_Chemgauss4. The evaluation outcomes are shown in Table 3, Table S7, and Figure 3 ($ROCE_{0.5\%}$, $ROCE_{1\%}$, AUC, and AR%), Table 4 (actives recovered), and Table 5 (Murcko scaffolds recovered).

Common Feature: Overall Improvement of Early Recognition. According to Table 3, we found that the PFE shares common features with the PF in overall effect on early enrichment. Higher average values of $ROCE_{1\%}$ and $ROCE_{0.5\%}$ for PFE_Chemgauss4 than for Chemgauss4 (i.e., 28.28 vs 23.55 and 42.35 vs 35.81, respectively) indicate that overall the addition of the PFE improves early enrichment of Chemgauss4 (cf. Table 3). As shown in Tables 4 and 5, the average numbers of actives and Murcko scaffolds at 1% of binding decoys recovered by PFE_Chemgauss4 are 41 and 36, which are greater than those recovered by Chemgauss4, i.e., 34 and 30. Therefore, the improvement of early enrichment appears to be derived from the greater numbers of actives/scaffolds identified by PFE_Chemgauss4 as well as the fact that the number of additional actives/scaffolds identified by PFE_Chemgauss4 are greater than the number of unidentified ones, as indicated by the average numbers of overlap actives/scaffolds. As shown in the tables, the pair of numbers of the actives/scaffolds identified is 41/36, which is close to that for PF_Chemgauss4 (39/33) but distant from that for Chemgauss4 (34/30). All the above data demonstrate that overall PFE_Chemgauss4 possesses the basic capacity of PF_Chemgauss4, i.e., improving early recognition by enriching more actives/scaffolds than Chemgauss4.

PFEs Gave More Consistent Improvement of Early Enrichment. For early enrichment measured by average $ROCE_{1\%}/ROCE_{0.5\%}$ values, the differences between PFE_Chemgauss4 and PF_Chemgauss4 were quite small (i.e., 28.28 vs 28.52 for $ROCE_{1\%}$ and 42.35 vs 43.51 for $ROCE_{0.5\%}$, respectively; cf. Table 3). On the basis of such general data, it appears first that the PFE does not show advantages compared with the PF. To uncover merits of the PFE, we further compared the outcomes after the use of the PFE and PF for each individual target by calculating their respective degrees of improvement in terms of $\Delta ROCE_{1\%}$, $\Delta ROCE_{0.5\%}$, ΔAUC, and ΔAR%. Notably, we observed an increase in early enrichment for all of the targets after using the PFE. Unlike the PFE, use of the PF improves early enrichment for only seven of the targets. For the other three targets, it significantly impairs the early recognition power of the original scoring function, i.e., Chemgauss4. To be specific, the $ROCE_{1\%}/ROCE_{0.5\%}$ values drop by 5.42/10.84 for HMDH, by 4.74/8.76 for PRGR, and more significantly by 10.42/12.50 for GRIK1 (cf. Table S7 and Figure 3). Correspondingly, their associated parameters are also much worse than those for the original scoring function. For instance, PF_Chemgauss4 identified considerably fewer actives/scaffolds than Chemgauss4 for these targets. This could be the major factor that causes an average of two fewer actives and three fewer scaffolds identified by PF_Chemgauss4 compared with PFE_Chemgauss4 (cf. Tables 4 and 5). At this point, the PFE is a more robust approach since it improves the early enrichment in a more consistent manner.

PFEs Showed Improved Maintenance of Active/Scaffold Recovery Power from Chemgauss4. We also compared PFE_Chemgauss4 and PF_Chemgauss4 in terms of unidentified% and unique% (cf. Tables 4 and 5). The outcomes for actives and their Murcko scaffolds recovered at the early stage are consistent among all targets except for GRIK1 and PRGR. According to the average scores across the 10 targets, the values of unidentified% for PFE_Chemgauss4 are much lower than those for PF_Chemgauss4, especially for actives (8.9 vs 18.8). Meanwhile, the average of unique% for PFE_Chemgauss4 is close to that for PF_Chemgauss4 for both actives and scaffolds. Ideally, an approach should be able to (1) recover all actives/scaffolds that were originally identified by Chemgauss4 (i.e., unidentified% = 0) and (2) identify as many more additional

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414−1425

actives/scaffolds as possible (i.e., greater unique%). The current data do show that the use of PFEs maintains the initial enrichment of Chemgauss4 more effectively and has equivalent capacity to discover additional actives/scaffolds compared with the use of PFs.

PFEs Result in Greater Overall Enrichment. According to the prior analysis, PFEs keep the overall enrichment of Chemgauss4 alone in most cases. The plots of ΔAUC and ΔAR% in Figure 3 further illustrate this point. To differentiate the effects of PFEs and PFs on overall enrichment, we first compared the numbers of targets on which the PFE and PF show positive effects (cf. Table S7 and Figure 3). On one hand, the PFE improved the AUC for six targets, while the PF did the same for only four targets. On the other hand, the PFE was able to identify over 80% of the actives for nine targets, while the PF did the same for only five targets. Next, we compared the significance of the changes that the PFEs and PFs cause (cf. Table 3). On average, the PFEs keep the AUCs at a constant level, while the PFs cause the AUCs to decline from 0.81 to 0.77. Meanwhile, the AR% value decreases from 100% to 87.03% after the use of the PFEs but to 74.06% after the use of the PFs. On the basis of the above two points, it can be concluded that PFEs are advantageous over PFs in maintaining a high level of overall enrichment.

CONCLUSIONS

In the current study, we have compiled a comprehensive data set that covers a total of 10 diverse targets. Each subset for its corresponding target contains carefully curated X-ray structures of protein−ligand complexes and benchmarking actives/decoys. This ready-to-use data set is not only designed for ensemble modeling and benchmarking (e.g., the construction and validation of PFEs as done in this study) but is also appropriate for other types of studies that require both knowledge of protein−ligand interactions and benchmarking sets. Most importantly, we have designed a novel ensemble modeling method based on multiple X-ray structures to construct pose filter ensembles (PFEs). The PFEs can be coupled with a structure-based empirical scoring function, i.e., Chemgauss4. As a means of validation, we have applied the approach to all of the diverse targets in the data set and extensively assessed the effect of each target-specific PFE on the ligand enrichment of Chemgauss4 by benchmarking the actives/decoys for that target.

The comparative analysis of ligand enrichment by Chemgauss4 before and after the use of PFEs has demonstrated that (1) as characterized by the increased $ROCE_{1\%}/ROCE_{0.5\%}$ and equivalent ROC AUCs, the use of PFEs can improve the early enrichment of Chemgauss4 while maintaining its overall enrichment, and (2) according to the numbers of Murcko scaffolds of the actives recovered at the early stage, the use of PFEs improves the chances of identifying actives of novel scaffolds by Chemgauss4. In addition, we have observed that the efficacy of the PFE varies with different targets. Further analysis of pairwise topological similarity, i.e., the *Tc* distribution within cognate ligands based on FCFP_6 fingerprints, has shown that the use of more diverse cognate ligands for PFE construction may lead to greater improvement in terms of ligand enrichment of Chemgauss4.

We have also evaluated the effect of ensemble modeling by comparing ligand enrichments of PFE_Chemgauss4 and PF_Chemgauss4. The evaluation outcome has demonstrated that the use of PFEs improves early recognition more consistently while maintaining the active/scaffold recovery power of Chemgauss4 and displaying greater overall enrichment than the use of PFs. These features of PFEs have validated our hypothesis that the introduction of additional knowledge from protein−ligand interfaces to the construction of the pose classifier/filter would bring forth much improvement in current SBVS scoring functions.

This study is a continuation of our efforts on target-specific PF construction using PL/MCT-tess descriptors to improve ligand enrichment. To the best of our knowledge, the workflow to construct PFEs based on the ensemble modeling technique and couple them with Chemgauss4 has not been reported to date. It is also interesting that the use of PFEs benefits the empirical scoring function as well. Since it works on a multitude of diverse targets, we anticipate its wide application and successes in future SBVS campaigns. Our most immediate goal is to apply this novel method to the discovery of farnesoid X receptor (FXR) agonists for the treatment of diabetes, as multiple X-ray crystal structures are currently available for this important target.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.6b00749.

Target list of DUD-E and numbers of X-ray structures of protein−ligand complexes covered by sc-PDB (Table S1); PDB IDs of excluded X-ray structures (Table S2); overlaps between true actives and cognate ligands (Table S3); parameters of each SVM-based PF (Table S4); numbers of ligands not used for benchmarking calculations (Table S5); increase rate of $ROCE_{1\%}$ and mean(*Tc*) values within cognate ligands (Table S6); ligand enrichments improved by PFE/PF (Table S7); superimposed X-ray structures of protein−ligand complexes (Figure S1); ligand enrichments for Chemgauss4/PFE_Chemgauss4 (Figure S2); case order plot of Cook's distance (Figure S3) (PDF)

AUTHOR INFORMATION

Corresponding Authors

*X.S.W.: E-mail: x.simon.wang@gmail.com. Address: Howard University College of Pharmacy, 2300 Fourth Street NW, Washington, DC 20059, United States.

*S.W.: E-mail: ws@imm.ac.cn. Address: State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Department of New Drug Research and Development, Institute of Materia Medica, Chinese Academy of Medical Sciences, and Peking Union Medical College, Beijing 100050, China.

ORCID

Jie Xia: 0000-0002-9567-3763

Xiang Simon Wang: 0000-0002-4156-3753

Notes

The authors declare no competing financial interest.

The comprehensive data sets for the ensemble modeling and benchmarking study are freely accessible at http://www.xswlab.org. The codes and manual for PFE construction are available upon request.

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414−1425

■ ACKNOWLEDGMENTS

This work was supported in part by the Fundamental Research Funds for the Central Universities (2016ZX350036) and the National Natural Science Foundation of China (NSFC 81603027). We are grateful to OpenEye Scientific Software for providing an academic license for their packages. We also thank Dr. Liangren Zhang (Peking University School of Pharmaceutical Sciences) for free access to computing facilities in his group. We are also thankful for the support from the District of Columbia Developmental Center for AIDS Research (P30AI087714), National Institutes of Health Administrative Supplements for U.S.−China Biomedical Collaborative Research (5P30A10877714-02), and the National Institute on Minority Health and Health Disparities of the National Institutes of Health under Award G12MD007597. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

■ ABBREVIATIONS

VS, virtual screening; SBVS, structure-based VS; PF, pose filter; PFE, PF ensemble; ENTess, electronegativity based on Delaunay tessellation; PL/MCT-tess, protein−ligand pairwise atomic maximal charge transfer potential based on Delaunay tessellation; QSBAR, quantitative structure−binding affinity relationship; SVM, support vector machine; HDAC2, histone deacetylase 2; TRY1, trypsin I; TRYB1, tryptase β1; AOFB, monoamine oxidase B; CP2C9, cytochrome P450 2C9; CXCR4, CXC chemokine receptor subtype 4; DRD3, dopamine D3 receptor; GLCM, β-glucocerebrosidase; KPCB, protein kinase C β; ADA, adenosine deaminase; HMDH, HMG-CoA reductase; MAPK2, MAP kinase-activated protein kinase 2; IGF1R, insulin-like growth factor I receptor; LKHA4, leukotriene A4 hydrolase; PRGR, progesterone receptor; CP3A4, cytochrome P450 3A4; GRIK1, glutamate receptor ionotropic kainate 1; ADRB2, β2 adrenergic receptor; FABP4, fatty acid binding protein adipocyte; MCT, maximal charge transfer; RBF, radial basis function; PFE_Chemgauss4, PFE-coupled Chemgauss4; PF_Chemgauss4, PF-coupled Chemgauss4; ROC, receiver operating characteristic; AUC, area under curve; ROCE, ROC enrichment; AR, actives recovered; PA, prediction accuracy; FCFP_6, function-class fingerprints of maximum diameter 6; *Tc*, Tanimoto coefficient

■ REFERENCES

(1) Schneider, G.; Bohm, H. J. Virtual Screening and Fast Automated Docking Methods. *Drug Discovery Today* **2002**, *7*, 64−70.
(2) Waszkowycz, B.; Clark, D. E.; Gancia, E. Outstanding Challenges in Protein−Ligand Docking and Structure-Based Virtual Screening. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 229−259.
(3) Cheng, T.; Li, Q.; Zhou, Z.; Wang, Y.; Bryant, S. H. Structure-Based Virtual Screening for Drug Discovery: A Problem-Centric Review. *AAPS J.* **2012**, *14*, 133−141.
(4) Zheng, H.; Handing, K. B.; Zimmerman, M. D.; Shabalin, I. G.; Almo, S. C.; Minor, W. X-Ray Crystallography over the Past Decade for Novel Drug Discovery − Where Are We Heading Next? *Expert Opin. Drug Discovery* **2015**, *10*, 975−989.
(5) Villoutreix, B. O.; Eudes, R.; Miteva, M. A. Structure-Based Virtual Ligand Screening: Recent Success Stories. *Comb. Chem. High Throughput Screening* **2009**, *12*, 1000−1016.
(6) Lionta, E.; Spyrou, G.; Vassilatis, D. K.; Cournia, Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Curr. Top. Med. Chem.* **2014**, *14*, 1923−1938.
(7) Tuccinardi, T. Docking-Based Virtual Screening: Recent Developments. *Comb. Chem. High Throughput Screening* **2009**, *12*, 303−314.
(8) Da, C.; Kireev, D. Structural Protein-Ligand Interaction Fingerprints (Splif) for Structure-Based Virtual Screening: Method and Benchmark Study. *J. Chem. Inf. Model.* **2014**, *54*, 2555−2561.
(9) Proctor, E. A.; Yin, S.; Tropsha, A.; Dokholyan, N. V. Discrete Molecular Dynamics Distinguishes Nativelike Binding Poses from Decoys in Difficult Targets. *Biophys. J.* **2012**, *102*, 144−151.
(10) Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martinez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. Recognizing Pitfalls in Virtual Screening: A Critical Review. *J. Chem. Inf. Model.* **2012**, *52*, 867−881.
(11) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455−1474.
(12) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912−5931.
(13) Graves, A. P.; Brenk, R.; Shoichet, B. K. Decoys for Docking. *J. Med. Chem.* **2005**, *48*, 3714−3728.
(14) Li, Y.; Han, L.; Liu, Z.; Wang, R. Comparative Assessment of Scoring Functions on an Updated Benchmark: 2. Evaluation Methods and General Results. *J. Chem. Inf. Model.* **2014**, *54*, 1717−1736.
(15) Li, T.; Yin, N.; Liu, H.; Pei, J.; Lai, L. Novel Inhibitors of Toxin Hipa Reduce Multidrug Tolerant Persisters. *ACS Med. Chem. Lett.* **2016**, *7*, 449−453.
(16) Lu, P.; Liu, X.; Yuan, X.; He, M.; Wang, Y.; Zhang, Q.; Ouyang, P. K. Discovery of a Novel Nedd8 Activating Enzyme Inhibitor with Piperidin-4-Amine Scaffold by Structure-Based Virtual Screening. *ACS Chem. Biol.* **2016**, *11*, 1901−1907.
(17) Song, Y.; Xue, X.; Wu, X.; Wang, R.; Xing, Y.; Yan, W.; Zhou, Y.; Qian, C. N.; Zhang, Y.; Xu, Y. Identification of N-Phenyl-2-(N-Phenylphenylsulfonamido)Acetamides as New Rorgamma Inverse Agonists: Virtual Screening, Structure-Based Optimization, and Biological Evaluation. *Eur. J. Med. Chem.* **2016**, *116*, 13−26.
(18) Smith, E. W.; Nevins, A. M.; Qiao, Z.; Liu, Y.; Getschman, A. E.; Vankayala, S. L.; Kemp, M. T.; Peterson, F. C.; Li, R.; Volkman, B. F.; Chen, Y. Structure-Based Identification of Novel Ligands Targeting Multiple Sites within a Chemokine−G-Protein-Coupled-Receptor Interface. *J. Med. Chem.* **2016**, *59*, 4342−4351.
(19) Hou, X.; Li, K.; Yu, X.; Sun, J. P.; Fang, H. Protein Flexibility in Docking-Based Virtual Screening: Discovery of Novel Lymphoid-Specific Tyrosine Phosphatase Inhibitors Using Multiple Crystal Structures. *J. Chem. Inf. Model.* **2015**, *55*, 1973−1983.
(20) Sorna, V.; Theisen, E. R.; Stephens, B.; Warner, S. L.; Bearss, D. J.; Vankayalapati, H.; Sharma, S. High-Throughput Virtual Screening Identifies Novel N′-(1-Phenylethylidene)-Benzohydrazides as Potent, Specific, and Reversible Lsd1 Inhibitors. *J. Med. Chem.* **2013**, *56*, 9496−9508.
(21) Vidler, L. R.; Filippakopoulos, P.; Fedorov, O.; Picaud, S.; Martin, S.; Tomsett, M.; Woodward, H.; Brown, N.; Knapp, S.; Hoelder, S. Discovery of Novel Small-Molecule Inhibitors of Brd4 Using Structure-Based Virtual Screening. *J. Med. Chem.* **2013**, *56*, 8073−8088.
(22) Jansen, J. M.; Martin, E. J. Target-Biased Scoring Approaches and Expert Systems in Structure-Based Virtual Screening. *Curr. Opin. Chem. Biol.* **2004**, *8*, 359−364.
(23) Perola, E. Minimizing False Positives in Kinase Virtual Screens. *Proteins: Struct., Funct., Genet.* **2006**, *64*, 422−435.
(24) Muthas, D.; Sabnis, Y. A.; Lundborg, M.; Karlen, A. Is It Possible to Increase Hit Rates in Structure-Based Virtual Screening by Pharmacophore Filtering? An Investigation of the Advantages and Pitfalls of Post-Filtering. *J. Mol. Graphics Modell.* **2008**, *26*, 1237−1251.
(25) Deng, Z.; Chuaqui, C.; Singh, J. Structural Interaction Fingerprint (Sift): A Novel Method for Analyzing Three-Dimensional

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414−1425

Protein-Ligand Binding Interactions. *J. Med. Chem.* **2004**, *47*, 337−344.
(26) Nandigam, R. K.; Kim, S.; Singh, J.; Chuaqui, C. Position Specific Interaction Dependent Scoring Technique for Virtual Screening Based on Weighted Protein–Ligand Interaction Fingerprint Profiles. *J. Chem. Inf. Model.* **2009**, *49*, 1185−1192.
(27) Perez-Nueno, V. I.; Rabal, O.; Borrell, J. I.; Teixido, J. Apif: A New Interaction Fingerprint Based on Atom Pairs and Its Application to Virtual Screening. *J. Chem. Inf. Model.* **2009**, *49*, 1245−1260.
(28) Desaphy, J.; Raimbaud, E.; Ducrot, P.; Rognan, D. Encoding Protein-Ligand Interaction Patterns in Fingerprints and Graphs. *J. Chem. Inf. Model.* **2013**, *53*, 623−637.
(29) *Molecular Operating Environment (MOE)*, version 2010.10; Chemical Computing Group: Montreal, QC, 2010.
(30) Zhang, S.; Golbraikh, A.; Tropsha, A. Development of Quantitative Structure-Binding Affinity Relationship Models Based on Novel Geometrical Chemical Descriptors of the Protein-Ligand Interfaces. *J. Med. Chem.* **2006**, *49*, 2713−2724.
(31) Hsieh, J. H.; Yin, S.; Liu, S.; Sedykh, A.; Dokholyan, N. V.; Tropsha, A. Combined Application of Cheminformatics- and Physical Force Field-Based Scoring Functions Improves Binding Affinity Prediction for Csar Data Sets. *J. Chem. Inf. Model.* **2011**, *51*, 2027−2035.
(32) Hsieh, J. H.; Yin, S.; Wang, X. S.; Liu, S.; Dokholyan, N. V.; Tropsha, A. Cheminformatics Meets Molecular Mechanics: A Combined Application of Knowledge-Based Pose Scoring and Physical Force Field-Based Hit Scoring Functions Improves the Accuracy of Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2012**, *52*, 16−28.
(33) Fourches, D.; Politi, R.; Tropsha, A. Target-Specific Native/Decoy Pose Classifier Improves the Accuracy of Ligand Ranking in the Csar 2013 Benchmark. *J. Chem. Inf. Model.* **2015**, *55*, 63−71.
(34) Politi, R.; Convertino, M.; Popov, K.; Dokholyan, N. V.; Tropsha, A. Docking and Scoring with Target-Specific Pose Classifier Succeeds in Native-Like Pose Identification but Not Binding Affinity Prediction in the Csar 2014 Benchmark Exercise. *J. Chem. Inf. Model.* **2016**, *56*, 1032−1041.
(35) Tian, S.; Sun, H.; Pan, P.; Li, D.; Zhen, X.; Li, Y.; Hou, T. Assessing an Ensemble Docking-Based Virtual Screening Strategy for Kinase Targets by Considering Protein Flexibility. *J. Chem. Inf. Model.* **2014**, *54*, 2664−2679.
(36) Bolia, A.; Gerek, Z. N.; Ozkan, S. B. Bp-Dock: A Flexible Docking Scheme for Exploring Protein-Ligand Interactions Based on Unbound Structures. *J. Chem. Inf. Model.* **2014**, *54*, 913−925.
(37) Da, C.; Mooberry, S. L.; Gupton, J. T.; Kellogg, G. E. How to Deal with Low-Resolution Target Structures: Using Sar, Ensemble Docking, Hydropathic Analysis, and 3d-Qsar to Definitively Map the Alphabeta-Tubulin Colchicine Site. *J. Med. Chem.* **2013**, *56*, 7382−7395.
(38) Cosconati, S.; Marinelli, L.; Di Leva, F. S.; La Pietra, V.; De Simone, A.; Mancini, F.; Andrisano, V.; Novellino, E.; Goodsell, D. S.; Olson, A. J. Protein Flexibility in Virtual Screening: The Bace-1 Case Study. *J. Chem. Inf. Model.* **2012**, *52*, 2697−2704.
(39) Korb, O.; Olsson, T. S.; Bowden, S. J.; Hall, R. J.; Verdonk, M. L.; Liebeschuetz, J. W.; Cole, J. C. Potential and Limitations of Ensemble Docking. *J. Chem. Inf. Model.* **2012**, *52*, 1262−1274.
(40) Rueda, M.; Totrov, M.; Abagyan, R. Alibero: Evolving a Team of Complementary Pocket Conformations Rather Than a Single Leader. *J. Chem. Inf. Model.* **2012**, *52*, 2705−2714.
(41) Rettenmaier, T. J.; Fan, H.; Karpiak, J.; Doak, A.; Sali, A.; Shoichet, B. K.; Wells, J. A. Small-Molecule Allosteric Modulators of the Protein Kinase Pdk1 from Structure-Based Docking. *J. Med. Chem.* **2015**, *58*, 8285−8291.
(42) Wang, B.; Buchman, C. D.; Li, L.; Hurley, T. D.; Meroueh, S. O. Enrichment of Chemical Libraries Docked to Protein Conformational Ensembles and Application to Aldehyde Dehydrogenase 2. *J. Chem. Inf. Model.* **2014**, *54*, 2105−2116.
(43) Li, Y.; Kim, D. J.; Ma, W.; Lubet, R. A.; Bode, A. M.; Dong, Z. Discovery of Novel Checkpoint Kinase 1 Inhibitors by Virtual Screening Based on Multiple Crystal Structures. *J. Chem. Inf. Model.* **2011**, *51*, 2904−2914.
(44) Bajusz, D.; Ferenczy, G. G.; Keseru, G. M. Discovery of Subtype Selective Janus Kinase (Jak) Inhibitors by Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2016**, *56*, 234−247.
(45) Liu, J.; Wang, R. Classification of Current Scoring Functions. *J. Chem. Inf. Model.* **2015**, *55*, 475−482.
(46) Desaphy, J.; Bret, G.; Rognan, D.; Kellenberger, E. sc-PDB: A 3D-Database of Ligandable Binding Sites—10 Years On. *Nucleic Acids Res.* **2015**, *43*, D399−D404.
(47) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (Dud-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582−6594.
(48) Xia, J.; Tilahun, E. L.; Reid, T.-E.; Zhang, L.; Wang, X. S. Benchmarking Methods and Data Sets for Ligand Enrichment Assessment in Virtual Screening. *Methods* **2015**, *71*, 146−157.
(49) Hawkins, P. C.; Nicholls, A. Conformer Generation with Omega: Learning from the Data Set and the Analysis of Failures. *J. Chem. Inf. Model.* **2012**, *52*, 2919−2936.
(50) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian Docking Functions. *Biopolymers* **2003**, *68*, 76−90.
(51) McGann, M. Fred Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2011**, *51*, 578−596.
(52) 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.
(53) Parr, R. G.; Szentpály, L. v.; Liu, S. Electrophilicity Index. *J. Am. Chem. Soc.* **1999**, *121*, 1922−1924.
(54) Liu, S.-B. Conceptual Density Functional Theory and Some Recent Developments. *Acta Phys.-Chim. Sin.* **2009**, *25*, 590−600.
(55) Chang, C. C.; Lin, C. J. Libsvm: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27.
(56) Nicholls, A. What Do We Know and When Do We Know It? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 239−255.
(57) Jahn, A.; Hinselmann, G.; Fechner, N.; Zell, A. Optimal Assignment Methods for Ligand-Based Virtual Screening. *J. Cheminf.* **2009**, *1*, 14.
(58) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.
(59) Cook, R. D. Influential Observations in Linear Regression. *J. Am. Stat. Assoc.* **1979**, *74*, 169−174.
(60) Cook, R. D. Detection of Influential Observation in Linear Regression. *Technometrics* **2000**, *42*, 65−68.

DOI: 10.1021/acs.jcim.6b00749
J. Chem. Inf. Model. 2017, 57, 1414−1425