JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

An Integrated Virtual Screening Approach for VEGFR-2 Inhibitors

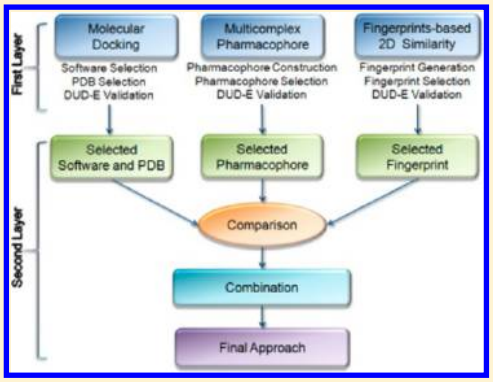
Yanmin Zhang,[†] Shangyan Yang,[‡] Yu Jiao,[‡] Haichun Liu,[†] Haoliang Yuan,[†] Shuai Lu,[†] Ting Ran,[†] Sihui Yao,[†] Zhipeng Ke,[†] Jinxing Xu,[†] Xiao Xiong,[†] Yadong Chen,*[,†] and Tao Lu*[,†,‡]

[†]Laboratory of Molecular Design and Drug Discovery, School of Science, China Pharmaceutical University, 24 Tongjiaxiang, Nanjing 210009, China

[‡]State Key Laboratory of Natural Medcines, School of Science, China Pharmaceutical University, 24 Tongjiaxiang, Nanjing, 210009, China

Supporting Information

ABSTRACT: In recent years, various virtual screening (VS) tools have been developed, and many successful screening campaigns have been showcased. However, whether by conventional molecular docking or pharmacophore screening, the selection of virtual hits is based on the ranking of compounds by scoring functions or fit values, which remains the bottleneck of VS due to insufficient accuracy. As the limitations of individual methods persist, a comprehensive comparison and integration of different methods may provide insights into selecting suitable methods for VS. Here, we evaluated the performance of molecular docking, fingerprint-based 2D similarity and multicomplex pharmacophore in an individual and a combined manner, through a retrospective VS study on VEGFR-2 inhibitors. An integrated two-layer workflow was developed and validated through VS of VEGFR-2 inhibitors against the DUD-E database, which demonstrated improved VS performance through a ligand-based method ECFP_4, followed by molecular docking, and then a strict multicomplex pharmacophore. Through a retrospective comparison with six published papers, this integrated approach outperformed 43 out of 45 methods, indicating a great effectiveness. This kind of integrated VS approach can be extended to other targets for the screening and discovery of inhibitors.



1. INTRODUCTION

Abnormal angiogenesis, which consists of blood vessel formation sprouting or splitting from pre-existing vessels,[1] is involved in many disease conditions, including cancer, inflammation, psoriasis, rheumatoid arthritis, degenerative eye conditions and some other diseases.[1,2] Vascular endothelial growth factors (VEGFs) and their respective family of tyrosine kinases receptors (VEGFRs), including VEGFR-1 (Flt-1), VEGFR-2 (Flk-1/kinase domain receptor [KDR]) and VEGFR-3 (Flt-4), are key proteins modulating angiogenesis. Particularly, VEGFR-2 exists mostly in vascular endothelial cells and hematopoietic stem cells, and mediates almost all known cellular responses to VEGF by stimulating multiplication of vascular endothelial cells and growth of blood vessels.[2,3] Therefore, VEGFR-2 emerges as a central target for discovering inhibitors against tumor-associated angiogenesis.[4] Most small-molecule inhibitors of VEGFR-2 are competitive for ATP by binding to the highly conserved ATP pocket.[5] Small-molecule VEGFR-2 kinase inhibitors including sorafenib, sunitinib, pazopanib, vatalanib, axitinib and regorafenib have been approved by the Food and Drug Administration (FDA) for clinical use.[6−8] However, since the drug resistance and low selectivity remain two concerns, discovering novel inhibitors with satisfactory outcomes is still in great demand.

As a complementary approach to experimental high-throughput screening (HTS),[9−11] virtual screening (VS) has gained lots of attention for saving time and expenses in recent years.[12,13] Molecular docking, similarity search and pharmacophore screening are the most frequently used methodologies in the discovery of small-molecule inhibitors. However, individual VS methods are facing some thorny problems; for example, insufficient accuracy of scoring functions remains the bottleneck of VS.[11,14] Molecular docking is a commonly used VS method, which generates optimal binding modes of ligands in a given binding pocket and ranks these poses on the basis of docking scores. However, the accuracy of scoring functions remains a major challenge for the success of docking-based VS. Accordingly, it is generally recommended to apply multiple scoring functions and conduct comparative evaluation/validation of the docking and scoring procedures based on known structures and activities.[15]

Similarity-based VS is considered as an integral part of modern *in silico* drug discovery, since structurally similar compounds are assumed to have similar physicochemical and biological properties.[16] This approach allows a fast identification of analogues to biologically active molecules.[17]

Received: July 23, 2013
Published: November 22, 2013

ACS Publications © 2013 American Chemical Society

Similarity search requires suitable representations of molecules, such as one-dimensional descriptors (e.g., log P), two-dimensional (2D) descriptors (e.g., MACCS structural keys), three-dimensional (3D) descriptors or spatial pharmacophores (e.g., CoMFA fields).[17] The 2D fingerprint and 3D shape similarity methods are frequently used in VS, such as extended-connectivity fingerprints (ECFP), MACCS, rapid overlay of chemical structures (ROCS) and phase Shape.[18,19] Nevertheless, compound structural classes strongly affect the performance of similarity-based methods and different approaches generate different results, which are often with little overlap in similarity relationships.[20] Independent studies have shown that using multiple reference compounds and novel algorithms can enhance the performance of similarity approaches.[20]

Pharmacophore screening has been extensively and successfully applied in VS, de novo design and lead optimization.[21] In ligand-based pharmacophore modeling, conformational flexibility of ligands and molecular alignment are the main difficulties; in structure-based pharmacophore modeling, quantitative structure−activity relationship (QSAR) cannot be effectively reflected by models derived from a single macromolecule−ligand complex or macromolecule alone.[21] To address these issues, a multicomplex-based pharmacophore model was suggested,[22] which may facilitate locating the ligand-binding site and determining the key protein−ligand interaction.

Since individual VS methods have their own advantages and disadvantages, we cannot expect the availability and applicability for any of them under all circumstances. The combined strategies for VS, either in a sequential or parallel manner, have been proposed to take the advantages of different VS methods.[14,23,24] However, even the order of the employed approaches can have a huge impact on the results of sequential VS strategy.[24] A parallel VS preselection before the sequential investigation may be a better choice as it assumes that different VS methods are supplementary to each other.[24] Therefore, a reasonable integration approach, including both parallel preselection and combination of various ligand- or structure-based methods may improve the possibility of finding more diverse active compounds for a target of interest.[11,23,25]

The purpose of this study was to evaluate the performance of molecular docking, fingerprint-based 2D similarity and multicomplex pharmacophore in the VS of VEGFR-2 inhibitors through parallel comparison and combination. We developed an integrated two-layer workflow (Figure 1) to improve the VS performance against the directory of useful decoys enhanced (DUD-E)[26] database of VEGFR-2 inhibitors. The first layer was a parallel investigation of three methods individually, and the second layer was a comprehensive comparison and combination of the optimal selection resulting from the first layer evaluation. Methods that performed well were retrospectively employed for a prospective integration which may promise good performance on the VS of VEGFR-2 inhibitors. Moreover, a retrospective comparison with methods of formerly published papers was conducted to further validate the effectiveness of this integrated approach.

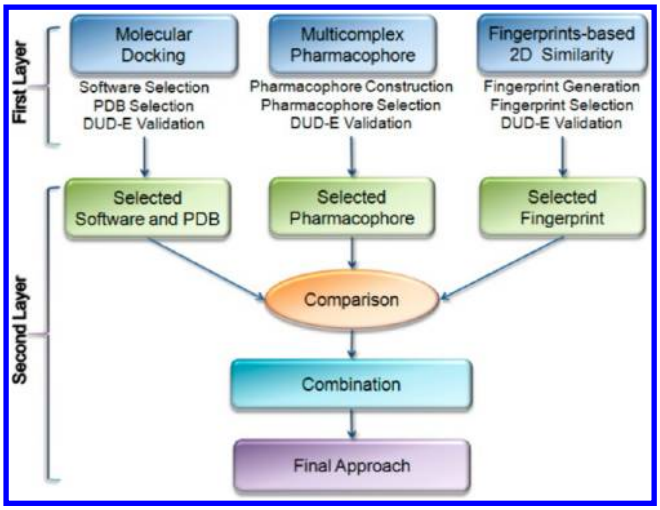


Figure 1. Workflow of the integrated two-layer VS evaluation.

Table 1. Information of All Thirty-one of the VEGFR-2 PDB Cognate Ligands

no.	PDB_ID	resolution(Å)	PDB_Ligand	ligand IC_{50} (nM)	release date
1	1Y6A	2.1	AAZ	22	2005-6-7
2	1Y6B	2.1	AAX	38	2005-6-7
3	1YWN	1.71	LIF	3	2005-8-23
4	2OH4	2.05	GIG	3.5	2007-9-18
5	2P2H	1.95	994	68	2007-3-20
6	2P2I	2.4	608	38	2007-3-20
7	2QU5	2.95	276	9	2007-9-25
8	2QU6	2.1	857	3	2007-9-25
9	2RL5	2.65	2RL	0.5	2008-4-8
10	2XIR	1.5	00J	*NT*[a]	2011-4-13
11	3B8Q	2.75	900	0.5	2008-4-1
12	3B8R	2.7	887	0.6	2008-4-1
13	3BE2	1.75	RAJ	2	2008-4-8
14	3C7Q	2.1	XIN	21 ± 13	2008-12-23
15	3CJF	2.15	SAV	6.3	2008-10-7
16	3CJG	2.25	KIM	*NT*[a]	2008-10-7
17	3CP9	2.5	C19	48	2008-6-17
18	3CPB	2.7	C92	>25000	2008-6-17
19	3CPC	2.4	C52	>5000	2008-6-17
20	3DTW	2.9	A96	353	2008-9-9
21	3EFL	2.2	706	3	2009-9-15
22	3EWH	1.6	K11	69 ± 10	2009-8-25
23	3U6J	2.15	03X	6.2	2012-2-22
24	3VHE	1.55	42Q	6.2 (4.7−8.3)	2011-11-2
25	3VHK	2.49	BPK	50000	2012-9-5
26	3VID	2.3	4TT	620	2012-8-15
27	3VNT	1.64	0JA	2.2 (2.0−2.4)	2012-4-11
28	4AGC	2.0	AXI	0.013	2012-9-26
29	4AGD	2.81	B49	2.7	2012-9-26
30	4ASD	2.03	BAX	2.3	2012-9-26
31	4ASE	1.83	AV9	0.04	2012-9-26

[a]*NT* - no accurate IC_{50} value is given.

2. MATERIALS AND METHODS

2.1. Crystal Structures Preparation. So far, more than 30 crystal structures of protein kinase VEGFR-2 (Table 1 and Supporting Information [SI] Figure S1) have been reported in the RCSB Protein Data Bank (PDB).[27] All 31 VEGFR-2 structures recently published were used to select the most appropriate docking program and crystal structure for docking compounds into VEGFR-2 pocket. These crystal complexes were downloaded and prepared using the Protein Preparation Wizard Workflow in Maestro of the Schrodinger 2009 package.

Only one chain and its cognate ligand as well as essential water molecules were kept for each PDB structure. Proteins were protonated and then the added hydrogens were minimized using the OPLS2005 force field.[28]

2.2. Ligand Query. Thirty-one crystallized ligands were extracted from their corresponding complexes after being aligned with each other. Their basic information is displayed in Table 1, including names in PDB database and activities (IC_{50} values) quoted from corresponding primary citations. Table 1 shows that except for two ligands (2XIR_L and 3CJG_L) with unavailable IC_{50} values, the IC_{50} values of the other 29 cognate ligands range from 0.013 nM (4AGC_L) to 50000 nM (3VHK_L), including 25 ligands with IC_{50} values below 10 nM and 4 ligands (3CPB_L, 3CPC_L, 3DTW_L, and 3VHK_L) over 100 nM. In addition, the 2D structures are displayed in Figure S1 in the SI where the red-marked atoms represent key structures that form one or more hydrogen bonds with amino acids (commonly Cys919 or Cys917) in the hinge region. In order to keep their original binding conformation, only missing hydrogens were added and minimized. We also calculated their molecular properties using the Calculate Molecular Property protocol in Discovery Studio 2.5 so as to compare the physical properties with that of the DUD-E validation compounds. These 31 crystallized ligands were used as validation materials for the selection of docking softwares and crystal structures in self- and cross-docking analysis, and they were also employed as the queries for 2D fingerprint-based similarity search. To avoid bias, all query ligands were excluded from the DUD-E database to prevent an early high false retrieval rate.

2.3. DUD-E Validation Database Preparation. DUD-E of VEGFR-2 inhibitors, an enhanced and rebuilt version of DUD 2.0 (78 actives and 2479 decoys), is a good benchmark library to evaluate the VS methods Duplicates were removed from the DUD-E, which contains 409 actives and 24680 decoys.[29] All active ligands and decoys were added with hydrogens and optimized using LigPrep module.

2.4. Docking Software Selection. Four commonly used docking softwares including Dock in Molecular Operating Environment (MOE), CDOCKER, Glide, and GOLD were used in a self-docking study. All 31 VEGFR-2 cognate ligands were self-docked into their corresponding crystal structures, and the docking results were evaluated by calculating the root-mean-square deviation (RMSD)[30] between the docked poses and the experimentally cocrystallized ones. Results were assessed according to the average RMSD value and the number of best poses with RMSD values below or equal to 2.0 Å. Then the selected softwares were further validated by a cross-docking analysis. Several statistical RMSD values, such as average and median RMSD values etc., were calculated on the basis of the docking results.

2.4.1. MOE. All 31 VEGFR-2 crystal structures were prepared using Protonate3D[31] tool within MOE to add hydrogen atoms and assign ionization states. Ligand atoms in the receptor were used to define a docking site. The DOCK module in MOE contains several placement methods such as Alpha PMI and Alpha Triangle and scoring functions including ASE, Affinity dG, Alpha HB, and London dG to perform protein−ligand docking. Planesas et al.[19] have assessed the reliability of these docking scoring functions by both self- and cross-docking analysis, and suggested that the combination of Alpha Triangle and Alpha HB generated the best results. Thus, Alpha Triangle was selected as the placement method, and Alpha HB was selected as the scoring function in this study. Ten poses of each ligand were retained to calculate the RMSD for further statistical calculation.

2.4.2. CDOCKER. The CDOCKER protocol in Discovery Studio 2.5,[32] an implementation of the CDOCKER algorithm,[33] allows us to run a refinement docking of any number of ligands with a single protein receptor. After preparing protein complexes by adding hydrogens and monitoring the bad valence, the CHARMm force field was applied to the structures. For each crystal structure, binding site spheres based on its cognate ligand were defined with an automatically generated radius. The top 10 hits were kept for further analysis in the self-docking process and the top 1 hit for the cross-docking study.

2.4.3. Glide. Glide module in the Schrodinger 2009 software was validated in the present study.[34] First, the 31 crystallized complex structures of VEGFR-2 were prepared with the Protein Preparation Wizard. Docking grid files were then generated using the cocrystallized ligand within the "Receptor Grid Generation" module at their default settings and an enclosing box that was similar in size to the cocrystallized ligand was generated. Only hydrogens of each cognate ligand were minimized to ensure that the input ligand conformations were their binding conformations. Ten poses per ligand were included to perform postdocking minimization in standard precision (SP), and the "Compute RMSD to input ligand geometries" option was selected. All the remaining parameters were kept as default settings.[30]

2.4.4. GOLD. GOLD program employs a genetic algorithm to explore ligand conformational flexibility in the protein of partial flexibility.[35] The docking results were indicated mainly by a GOLD fitness score calculated from contributions of hydrogen bonds and van der Waals interactions between the protein and ligand, and intramolecular hydrogen bonds and strains of the ligand.[36,37] The X-ray cognate ligands of VEGFR-2 complexes were used to define the active site region with a radius of 15 Å. The annealing parameters of van der Waals and hydrogen-bond interactions were considered within 4.0 Å and 2.5 Å, respectively. GOLDScore was selected as the fitness function and 10 best solutions for each ligand were saved at the end of the calculation and other parameters were remained as default.

2.5. 2D Fingerprints Similarity Search. Thirteen fingerprints were generated using the Find Similar Molecules by Fingerprints protocol in Discovery Studio 2.5 which finds ligands in an input library that are similar to the reference ligands. The similarity is measured by comparing a selected fingerprint property of the investigated compounds with that of the queries. The fingerprints include the MDLPublicKeys and 12 extended-connectivity fingerprints xCFz_n, where x represents the atom type (E for atom-types and F for functional-class); z indicates whether the presence or absence of a feature is only considered or the number of features is also considered (C refers to counts); and n is then the diameter of the substructure which represents a two-, four-, or six-bonds feature.[18] Hence, the 13 extended connectivity fingerprints were investigated, consisting of three ECFP fingerprints (ECFP_2, ECFP_4, and ECFP_6), three FCFP fingerprints (FCFP_2, FCFP_4, and FCFP_6), three ECFC fingerprints (ECFC_2, ECFC_4, and ECFC_6), and three FCFC fingerprints (FCFC_2, FCFC_4, and FCFC_6) and MDLPublicKeys. VS validation was performed on VEGFR-2 inhibitors using the Tanimoto index to assess the similarity subsequently. The Tanimoto coefficient T_{AB}, mostly used for binary fingerprints, is shown as eq 1:

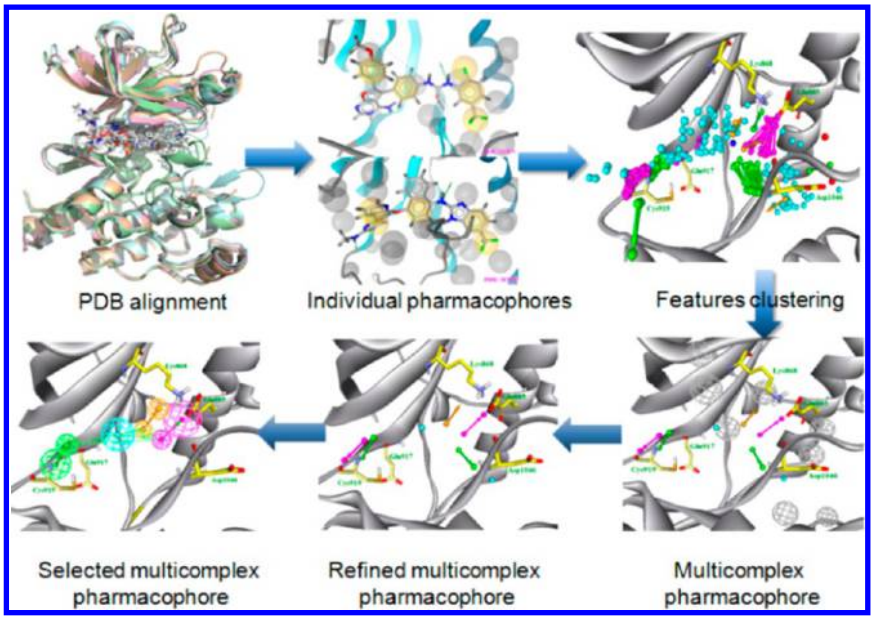


Figure 2. Workflow of generating and refining multicomplex pharmacophores.

$$T_{AB} = \frac{c}{a + b - c} \tag{1}$$

where individual fingerprint bits set in molecules A and B are represented by a and b; and then c represents the number of common bits. T_{AB} value ranges from 0 to 1, where 0 represents no detection of the same bits; however, 1 does not mean that the two molecules are totally identical. Additionally, atom pair similarity was calculated by the Strike similarity suite given in eq 2:

$$\mathrm{sim}_{AB} = \left[\sum_{k}^{m_{AB}} \omega_K \min(\mathrm{freq}_k^A, \mathrm{freq}_k^B)\right] / 0.5\left[\sum_{k}^{m_{AB}} \omega_K \min(\mathrm{freq}_k^A + \mathrm{freq}_k^B)\right] \tag{2}$$

where m_{AB} represents the total number of unique atom pair types in molecule A and molecule B, freq_k^A and freq_k^B represent the number of times atom pair type k is found in molecule A and molecule B, respectively. The atom pair similarities will be determined by the number of atom pair types shared by the two molecules, where 0 indicates no similarity and 1 indicates identity.[18]

2.6. Multicomplex Pharmacophore. The whole process of constructing a multicomplex-based pharmacophore model is depicted in Figure 2. LIGANDSCOUT 2.02,[38] an adopting algorithm that allows automatic construction of a 3D pharmacophore from the protein−ligand complex, was utilized to generate 31 individual CATALYST pharmacophore models based on previously aligned structures. To cover the most common features that may account for VEGFR-2 inhibitory potency, all pharmacophore features identified by LIGANDSCOUT 2.02 were clustered according to their interaction pattern with the receptor using the Edit and Cluster Pharmacophores tool in Discovery Studio 2.5. The obtained model was further optimized by the modification of the constraint tolerance of the spheres attending to the default values of CATALYST software. To obtain the most favorable multicomplex pharmacophore, we first set more clustering centers to make sure that the raw pharmacophore contained the most frequently existing pharmacophoric features. Then by a pruning method, each feature was deleted for one time, and the remaining combinations were adopted for the VS until the best EF was achieved within a given precondition.

2.7. Criteria of Performance Evaluation. Enrichment factor (EF) at a given percentage of the screened database is one of the most popular and simplest metric methods to evaluate the effectiveness of VS against a given protein target.[39−41] The EF is described as[42]

$$\mathrm{EF}^{x\%} = \frac{\mathrm{Hits}_{\mathrm{selected}}^{x\%}/N_{\mathrm{selected}}^{x\%}}{\mathrm{Hits}_{\mathrm{total}}/N_{\mathrm{total}}} \tag{3}$$

where $\mathrm{Hits}_{\mathrm{selected}}^{x\%}$ and $N_{\mathrm{selected}}^{x\%}$ represent the number of hits (actives) and the number of compounds filtered at the top x% of the database screened, respectively, $\mathrm{Hits}_{\mathrm{total}}$ and N_{total} refer to the number of actives and the number of compounds of the entire database correspondingly.[42] As the DUD-E validation database is relatively large, not all of the resulting compounds were used for analysis; only the top 5% (1254 compounds), 2% (502 compounds), 1% (251 compounds), 0.5% (125 compounds), 0.2% (50 compounds), 0.1% (25 compounds) hits were saved for further evaluation of the adopted methods. The docking scores after molecular docking, fitvalue after pharmacophore screening, and similarity value after similarity search, were used to plot the percentage of known actives versus the percentage of ranked database screened (enrichment plot) and calculate the EF at each stage of selected compounds.[43] Moreover, the receiving operating characteristic (ROC) curve, a graphical plot of the sensitivity (true positive rate, TP) vs 1 − specificity (false positive rate, FP), was calculated to avoid the sensitivity for small changes in ranking. The sensitivity (Se) and specificity (Sp) are given as eq 4 and eq 5:

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4}$$

$$\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{5}$$

dx.doi.org/10.1021/ci400429g | J. Chem. Inf. Model. 2013, 53, 3163−3177

where TP (true positives) represents the number of correctly identified actives, TN (true negatives) represents the number of correctly identified decoys, FP (false positives) represents the number of decoys incorrectly predicted as actives, and FN (false negatives) represents the number of actives incorrectly predicted to be inactives.[42] The area under the ROC curve (AUC plot) gives the probability of ranking a randomly selected active higher than a randomly chosen decoy. It ranges from 0 to 1, where 1 indicates a perfect ranking wherein all actives ranked above the decoys, and 0.5 corresponds to a completely random ranking.[44] A traditional academic point system for classifying the accuracy of a virtual screening test is known as follows: AUC < 0.5 is fail; 0.5 ≤ AUC < 0.70 is poor; 0.7 ≤ AUC < 0.8 is fair; 0.8 ≤ AUC < 0.9 is good; and 0.9 ≤ AUC ≤ 1 is excellent. Herein, we lifted the threshold of evaluation as we defined that AUC < 0.6 is a failure. All the numerical results are calculated within a 95% confidence interval. Additionally, if the obtained lists only contain active molecules, it makes it impossible for the calculation of AUC; however, this is an ideal situation for the VS investigation, so we regard this kind of result as an excellent level, and the AUC values are marked as 1* . Essentially independent of the actual number of positive and negative instances, the AUC of an ROC plot gives an objective measure of query performance.[42]

3. RESULTS AND DISCUSSION

3.1. Evaluation of Queries and DUD-E Database. Several 2D properties of the 31 queries (i.e., VEGFR-2 PDB cognate ligands) and compounds in the DUD-E validation database were calculated to avoid bias in various methods' results. The physical properties including charges, minimized energy, AlogP, molecular weight, number of rotatable bonds, rings, aromatic rings, hydrogen acceptors, hydrogen donors, and molecular fractional polar surface area. As shown in Figure 3 and SI Table S1, all compounds had Gasteiger charges. The

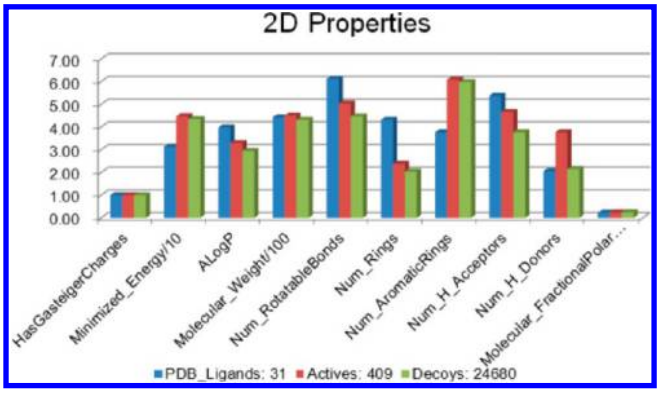


Figure 3. Statistical values comparison of 2D properties of VEGFR-2 PDB cognate ligands and DUD-E validation compounds.

molecular weight demonstrated an acceptable molecular weight error range with a mean weight of about 440. Other physical properties were also in a minor difference (error ranged from 0 to 2). For example, the average numbers of rings were 4, 2, 2; aromatic rings were 4, 6, 6; acceptors were 5, 5, 4; donors were 2, 4, 2 for the three-part compounds. All of these 2D descriptors demonstrated that the 31 queries and compounds in DUD-E database were of similar properties space.

3.2. Docking Software and Crystal Structure Selection. As the four docking softwares including MOE, CDOCKER, Glide_SP and GOLD have been investigated, a total of 2046 docking runs have been tested, including 124 self-docking and 1922 cross-docking runs. According to the self-docking validation (SI Table S2), CDOCKER and Glide_SP obtained better results. For each docking run, the RMSD values between the 10 poses generated and their original cognate ligand were calculated (SI Table S3). In addition, the statistics for self-docking containing the best-scored, minimum, maximum, average, and median RMSD values as well as the number of poses with RMSD values below 2.0 Å are available in the SI (Table S4). For further analysis, we computed the average statistical values of the above-mentioned six initial statistics. Seen from Table S2 in SI and Figure 4, CDOCKER

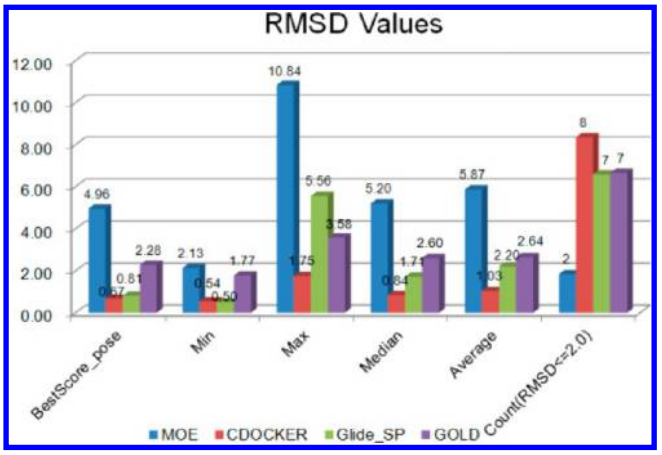


Figure 4. Average statistical RMSD values of self-docking results.

outperformed the other three software programs with all five average statistical RMSD values lower than 2.0 Å and eight PDB ligands with average RMSD values lower than 2.0 Å. Glide_SP was also favorable with an average RMSD value for the best scored pose of 0.81 Å and the lowest average minimum RMSD value of 0.50 Å. However, the average maximum RMSD value reached 5.56 Å, which may be due to a large range of rotation of the ligands during the docking pose generation process. Both Glide_SP and GOLD have resulted in seven PDB ligands with an average RMSD value below 2.0 Å. MOE performed the least best with the highest average RMSD values according to first five statistics. Thus, we selected CDOCKER and Glide_SP for further cross-docking validation.

During the cross-docking validation, each cocrystallized ligand was docked into the rest of protein structures. Hence, for each cross-docking combination, the top ranked pose with the best docking score was kept to compute the RMSD value with its original PDB cognate ligand. The original cross-docking information is available in the SI (Table S5 for Glide_SP and Table S6 for CDOCKER). In the cross-docking, 1YWN_L was not docked into 3VNT, 3C7Q_L not docked into 2QU5, 3CPC, 4AGC, 4ASD, or 4ASE and 3CPB_L not docked into PDB 3VID. Due to complicated structures, the cognate ligands in 1YWN_L and 3C7Q_L were not docked into some cavities. 3C7Q_L had a long (4-methylpiperazin-1-yl)acetamido)-phenylamino group that would extend to the solvent accessible area, but cavities were relatively small which limited the entry of molecules. Regarding 3CPB_L, the *N*-(2-cyclopentylethyl)-acetamide part being very flexible may contribute to its inability to enter the cavity.

dx.doi.org/10.1021/ci400429g | J. Chem. Inf. Model. 2013, 53, 3163−3177

Table S7 in the SI summarizes the statistical RMSD values of cross-docking results including the average, median, minimum, quartile (i.e., Quartile 1/4), three-quarters of the median (i.e, Quartile 3/4), maximum RMSD values as well as the number of cross-docking runs with RMSD values below 2.0 Å. Additionally, residues of the average and median RMSD values between CDOCKER and Glide_SP docking are given in Table S7 in the SI too. To better determine which software was more favorable, frequency statistics (SI Table S8) were calculated for the number of average and median cross-docking RMSD values. As shown Table S8 in the SI, both CDOCKER and Glide_SP did not obtain any cross-docking results with average RMSD values lower than 2 Å. Four cross-dockings of Glide_SP achieved average RMSD values between 2 Å and 4 Å. In the range of 4−6 Å, both CDOCKER and Glide_SP obtained 12 groups of cross-docking, which covered the most concentrated area. The second concentrated region was the interval of 6−8 Å, where CDOCKER returned 10 groups and Glide_SP returned eight groups of cross-docking. Still, there were nine groups of average RMSD values for CDOCKER cross-docking and eight groups for Glide_SP over 8 Å. In terms of median RMSD values, 13 groups of cross-docking were lower than 2 Å for Glide_SP, whereas CDOCKER only got six groups. For the rest intervals, CDOCKER was mainly distributed in the range of 8−12 Å, while Glide_SP was more evenly distributed with each interval two to four groups in the area <12 Å.

We also compared residues of average and median RMSD values between the CDOCKER and Glide_SP cross-docking (Table S7 in the SI and Figure 5). Figure 5a depicts the average RMSD values comparison for cross-docking results, and a histogram based on the difference between CDOCKER and Glide_SP average RMSD values was produced. Values >0 indicated that CDOCKER performed worse than Glide_SP, and the corresponding columns in the histogram were above the *X* axis. On the contrary, if the values were <0, the corresponding columns were below the *X* axis, indicating a better performance of CDOCKER. From Figure 5a, only five columns (PDBs: 3CPC, 3EFL, 3VNT, 4AGC and 4AGD) were below the *X* axis, and the rest of the 26 PDBs were above the *X* axis. Furthermore, three PDBs (1YWN, 2QU6, 2RL5) for Glide_SP were 2 Å lower than those in CDOCKER. Handled in a similar way, the median RMSD values are depicted in Figure 5b. There were also five columns (PDBs: 1Y6B, 3CPB, 3CPC, 4AGC, 4ASD) below the *X* axis and the difference of two PDBs (2P2I: 7.18 Å, 2QU6: 9.62 Å) were much lower for the Glide_SP compared with CDOCKER. Besides, the number of RMSD values lower than 2 Å are listed in Table S7 in the SI and are shown in Figure 5c. Only two PDBs (4AGC, 4ASD) for CDOCKER obtained more RMSD values below 2 Å than Glide_SP and seven PDBs (2OH4, 2P2H, 3C7Q, 3CJG, 3CPC, 3VID, 4AGD) obtained the same number of RMSD values below 2 Å with Glide_SP. All these statistics indicated that Glide_SP performed much better, and therefore it was chosen as the docking software for the selection of the most suitable crystal structures and subsequent docking-based VS.

Five PDBs including 1YWN, 2RL5, 3B8Q, 3EWH and 4ASE were tested (Figure 6a). Table S9 in the SI shows the average, minimum, median, quartile (i.e., Quartile 1/4), three-quarters of the median (i.e, Quartile 3/4), maximum RMSD values as well as the cumulative frequency for cross-docking runs with RMSD values below 2.0, 3.0, 4.0, 5.0, 6.0 Å, respectively. Moreover, a histogram (Figure 6b) was created for the first six parametric statistical values and a line chart (Figure 6c) for the

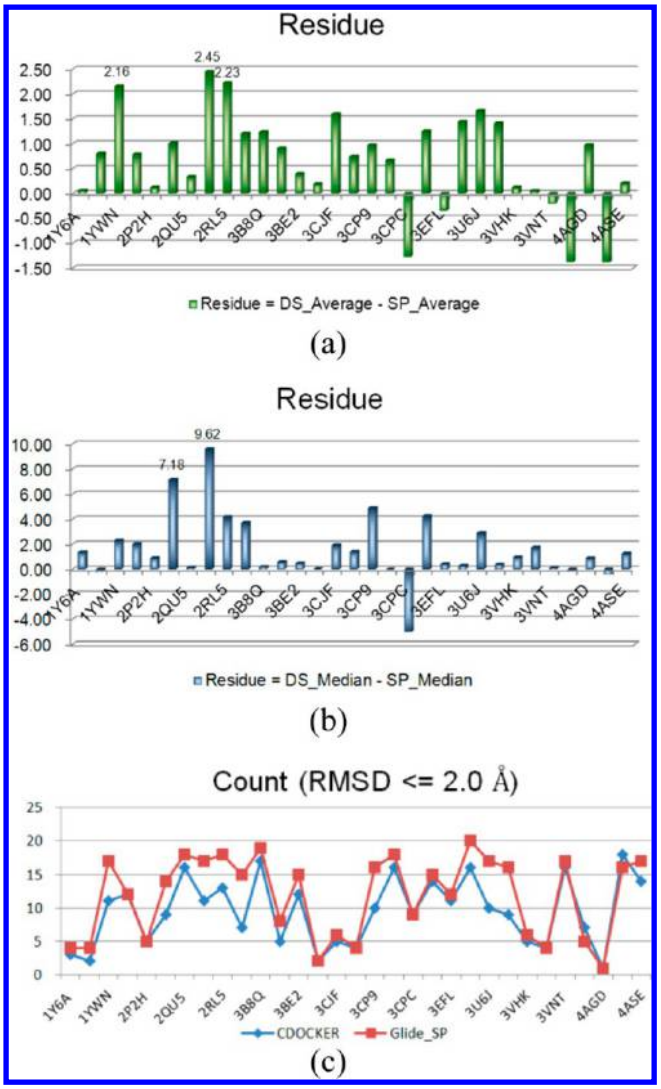


Figure 5. Average statistical RMSD values comparison of cross-docking results. (a) Residues of the average RMSD values between CDOCKER (DS_Average) and Glide_SP (SP_Average). (b) Residues of the median RMSD values between CDOCKER (DS_Median) and Glide_SP (SP_Median). (c) The number of RMSD values lower than 2 Å for CDOCKER and Glide_SP.

cumulative frequency. From Table S9 in the SI and Figure 6, we could observe that PDB 3B8Q retrieved the optimal average RMSD value of 2.91 Å. Nevertheless, its crystallographic resolution was the poorest for PDB 3B8Q (2.75 Å) among the other four PDBs (1YWN: 1.71 Å, 2RL5: 2.65 Å, 3EWH: 1.60 Å, 4ASE: 1.83 Å). Both PDB 1YWN and 3EWH retrieved average favorable RMSD values of 3.47 Å and 3.42 Å, especially PDB 3EWH performed best. It obtained a minimum RMSD value of 0.30 Å, three-quarters of the median (i.e, Quartile 3/4) RMSD value of 2.57 Å and 20 groups were retrieved in this interval. Enyedy et al.[45] and Planesas et al.[19] had also used PDB 1YWN in the docking experiments applied to an in-house database. Therefore, the top three PDBs (1YWN, 3B8Q, 3EWH) were chosen for the DUD-E validation database to achieve a favorable PDB. Among the three selected PDBs, the IC_{50} values were 3 nM,[46] 0.5 nM,[47] and 69 nM[48] for the cognate ligands in PDB 1YWN, 3B8Q, and 3EWH, respectively.

dx.doi.org/10.1021/ci400429g | J. Chem. Inf. Model. 2013, 53, 3163−3177

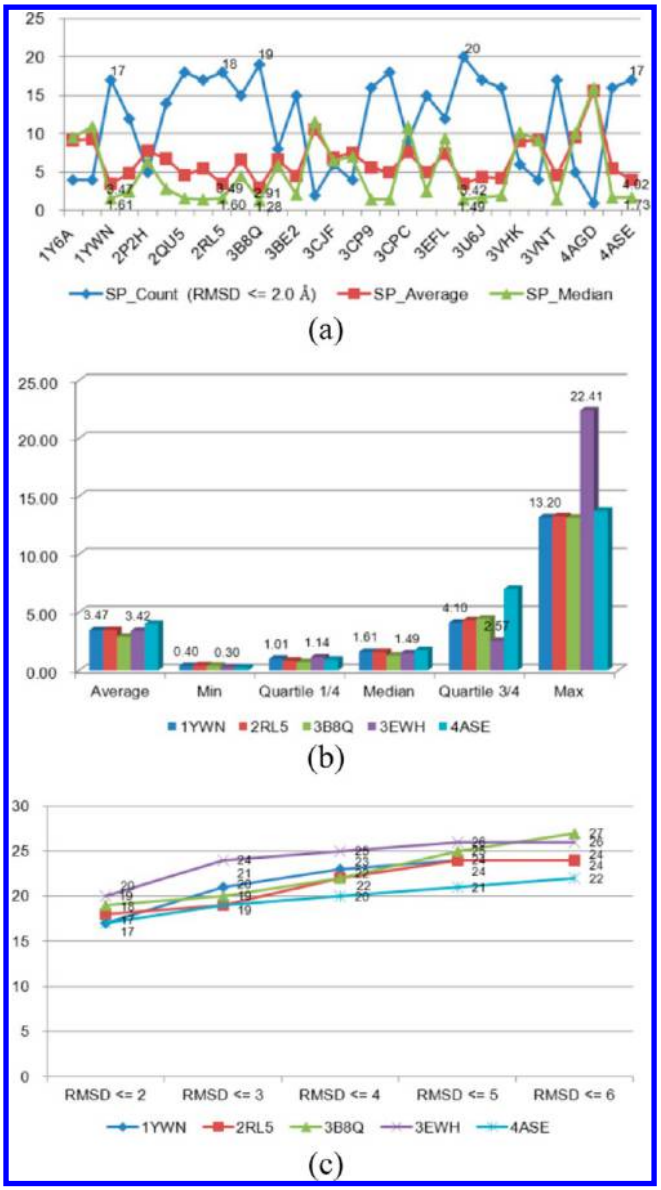


Figure 6. Selection of the most favorable PDBs. (a) Average, median, and the number of RMSD values lower than 2 Å for Glide_SP. (b) Statistical RMSD values for the selected five PDBs. (c) The number of RMSD values lower than 2 Å for the selected five PDBs.

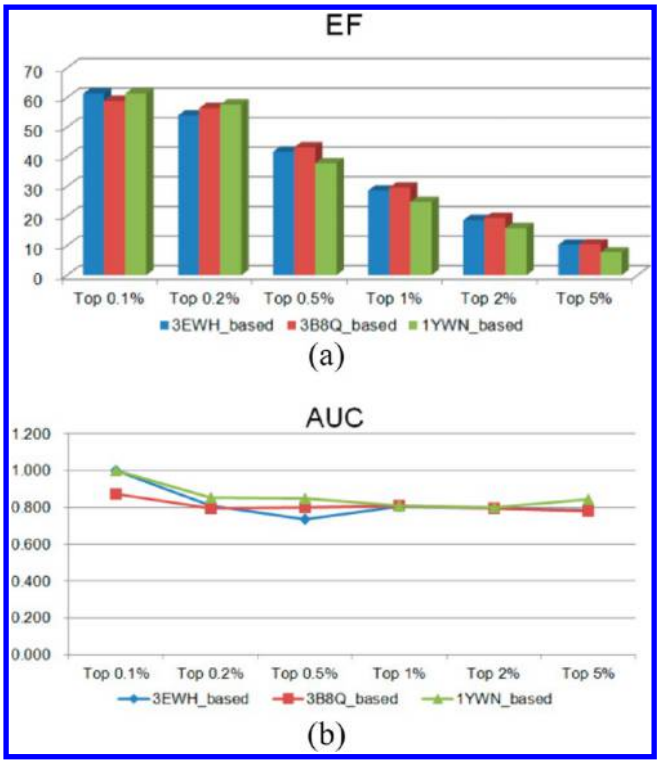


Figure 7. Selection of the most robust docking software and PDBs. (a) EF values for the three selected PDBs of the saved lists at six stages using Glide_SP docking. (b) AUC values for the three selected PDBs of the saved lists at six stages using Glide_SP docking.

3.3. Glide_SP-based VS. Thus, these three PDBs were further validated by VS of the DUD-E database. The ranked hit lists of the top 5%, 2%, 1%, 0.5%, 0.2%, and 0.1% were reserved for further analysis. Statistical values including the number of actives and decoys, the EF and AUC values, as well as the AUC levels in these six ranked lists are displayed in Table S10 in the SI and Figure 7. In addition, the selectivity and specificity values were also calculated (Table S10 in the SI). Overall, PDB 1YWN performed worst compared with the other two PDBs, since it captured the least number of actives in the six stages except the top 0.2% stage. There was only small difference between PDB 3EWH and 3B8Q in the number of obtained actives. PDB 3B8Q performed better except at the top 0.1% stage where 24 actives were achieved compared with the 25 actives obtained by PDB 3EWH.

As selectivity and specificity are closely related to the quantity of captured actives and decoys, they were in a similar trend with the number of actives and decoys. As shown in Table S10 in the SI, at the stage of the top 5%, PDBs 3EWH and 3B8Q achieved a selectivity of 0.51 and 0.52 correspondingly, whereas PDB 1YWN only obtained a selectivity of 0.38. However, when we considered the stage of the top 0.1% and the top 0.2%, the selectivities were greatly reduced to 0.06 and 0.11 for all three PDBs. For the specificity, it was worth noting that the specificity values for all three PDBs in all stages were quite satisfactory. At the stage of the top 0.1%, 0.2%, and 0.5%, the specificity values were equal to 1, and even at the stage of the top 5%, it reached 0.96. However, this was by no means the result of a perfect reference because the ranked lists were stored in accordance with the previous sets for the convenience of analysis rather than an automatic storage. Nevertheless, since all the results were done in the same way, it could certainly be referred to here for detailed analysis. Balancing the selectivity and specificity is quite challenging, and one solution is to set a threshold through a number of experiments. Herein, we set the threshold of selectivity to 0.5 which means at least half of the actives should be obtained.

All three PDBs showed improvement in EF about 6 to 7.5 times at the expense of a 50-fold decrease in the number of molecules selected. At this point, PDB 3EWH performed best. In terms of the AUC values (Table S10 in the SI) calculated by the ROC methods, from the stage of the top 5% to top 0.1%, PDB 3EWH exerted an improvement of 0.783 (Fair level) to 1* (Excellent level), PDB 3B8Q exerted an improvement of 0.781 (Fair level) to 0.875 (Good level), and PDB 1YWN demonstrated an increase of 0.845 (Good level) to 1* (Excellent level).From the aforementioned five aspects (i.e., the number of actives, selectivity, specificity, EF and AUC values), it could be summarized that PDB 1YWN could be

dx.doi.org/10.1021/ci400429g | J. Chem. Inf. Model. 2013, 53, 3163−3177

excluded from the crystal structure candidates as a result of low selectivity while PDB 3EWH performed best among these three crystal structures. Therefore, PDB 3EWH were chosen as the most appropriate VEGFR-2 PDB for the docking-based VS. This conclusion conformed well to our previous self- and cross-docking analysis.

3.4. 2D Fingerprints Similarity-Based VS. A total of 13 fingerprints including MDLPublicKeys and 12 extended-connectivity fingerprints xCFz_n were investigated in the VS of DUD-E database to identify the most favorable 2D fingerprint property. Here we employed multiple inhibitor queries for VEGFR-2 kinase inhibitors screening and the similarity based on Tanimoto coefficient T_{AB} was adopted as the evaluation criterion when ranking the hit lists. The same evaluation in Glide_SP-based VS was used to analyze the ranked hit lists of the top six levels. The five evaluation characteristics are calculated in the SI (Table S11), and the corresponding statistical charts are depicted (Figure 8). All the 13 fingerprints were divided into three categories: xCFP_n (x = E/F, n = 2/4/6); xCFC_n (x = E/F, n = 2/4/6) and the MDLPublicKeys.

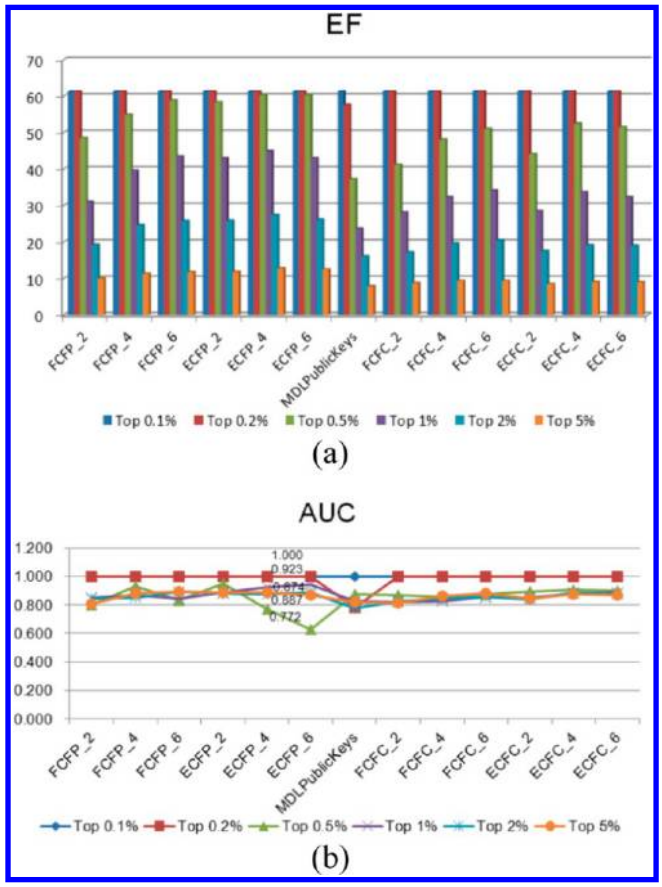


Figure 8. Selection of the most robust 2D fingerprint. (a) EF values for the 13 2D fingerprints of the saved lists at six stages. (b) AUC values for the 13 2D fingerprints of the saved lists at six stages.

As shown in Table S11 in the SI, at the stage of the top 0.1%, all 25 compounds were actives, and 50 compounds at the stage of the top 0.2% were also actives except MDLPublicKeys which only captured 47 actives. We observed that more than 200 actives were obtained in the xCFP_n (x = E/F, n = 2/4/6) category, whereas no fingerprint in the xCFC_n (x = E/F, n = 2/4/6) category achieved over 200 actives. The MDLPublicKeys performed worst with only 162 actives retrieved, relatively low in comparison with the other two categories. In xCFP_n (x = E/F, n = 2/4/6) category, ECFP_4 outperformed the other five fingerprints at all six stages of the hit lists, especially at the stage of the top 5% where 263 actives were retrieved. At the stage of the top 0.5%, the performance of ECFP_6 was the same as that of ECFP_4, and in other stages there was only a slight difference between ECFP_4 and ECFP_6. This may be due to the fact that they are of the same type only with different diameters of the substructure. The same phenomenon also existed when the selectivity was greatly reduced with a decrease of selected compounds. For example, the selectivity values of the best two fingerprints ECFP_4 and ECFP_6 decreased from 0.64 at the top 5% stage to 0.06 at the top 0.1% stage. However, it was necessary to improve the accuracy of the positive rate at the expense of decreasing selectivity. The specificity values were as high as 0.96 at the top 5% stage and even reached 1 at the stages of the top 0.1%, top 0.2%, and top 0.5% as well as another nine fingerprints in the top 1% lists, which demonstrated that almost all true negatives have been distinguished.

As for EF (Figure 8a), all 13 fingerprints gave a relatively high EF value of 61 at stages of the top 0.1% and 0.2%. At the stage of the top 5%, the EF values were all >10 for the fingerprints in xCFP_n (x = E/F, n = 2/4/6) category where both ECFP_4 and ECFP_6 reached 13, while the xCFC_n (x = E/F, n = 2/4/6) category fingerprints were all less than 10 but above 9. The MDLPublickeys performed worst with an EF value of 8. Moreover, as illustrated in Table S11 in the SI and Figure 8b, all the AUC levels of 13 fingerprints except MDLPublickeys reached an excellent level with AUC of 1* at stages of the top 0.1% and 0.2%, indicating that all the saved compounds were actives. Even at the stage of the top 5% which contained the largest number of compounds the AUC values were above 0.8 (Good level), demonstrating a very good ability to identify active compounds from the DUD-E validation database.

The detailed analysis was implemented for three categories at the stage of the top 5%. For the xCFP_n (x = E/F, n = 2/4/6) category, AUC values of four fingerprints exceeded 0.88 except FCFP_2 and ECFP_6 reaching 0.809 and 0.869, respectively. For the xCFC_n (x = E/F, n = 2/4/6) category, FCFC_6 also achieved a satisfactory result of 0.88 despite AUC values of other five fingerprints lower than 0.8. At this point, the MDLPublickeys ranked the third but last position, only higher than FCFP_2 (AUC: 0.809) and FCFC_2 (AUC: 0.811). However, it was worth noting that ECFP_4 acquired a fair level AUC value of 0.772, although it was still higher than ECFP_6 with a fail level AUC value of 0.626. This deviation was inconsistent with former results when ECFP_4 outperformed other 12 fingerprints. It was found that at the top 0.5% stage ECFP_4 captured 123 actives and 2 decoys, and these two decoys ranked in relatively top positions (ranked at the 96th and 97th positions), leading to a decrease of their AUC values. No matter how high the EF or AUC values are, they all rely on cutoffs made at various points through the ranking of actives, so they can be sensitive to small changes in ranking. This required that we could not choose the most appropriate parameter (e.g., fingerprints) for VS merely on one of the five factors aforementioned individually, but had to consider them in a comprehensive way. On the basis of this principle, ECFP_4 was selected as the most favorable fingerprint of the 2D fingerprints-based VS. Additionally, we could conclude that

dx.doi.org/10.1021/ci400429g | J. Chem. Inf. Model. 2013, 53, 3163−3177

for the 2D fingerprints-based VS to discover VEGFR-2 kinase related inhibitors, the xCFP_n (x = E/F, n = 2/4/6) category may outperform xCFC_n (x = E/F, n = 2/4/6) and the MDLPublickeys.

3.5. Multicomplex Pharmacophore-based VS. Thirty-one individual pharmacophores were automatically generated by LIGANDSCOUT 2.02 software based on the X-ray crystal structures of VEGFR-2. All the detected features are shown in the SI Figure S2a, we could categorize the binding pocket into four significant regions: the hinge region where key hydrogen interactions are formed between ligands and Cys919 of VEGFR-2 (PDB ID: 3EWH); a small hydrophobic region formed by hydrophobic amino acids such as Leu840, Gly841, Val848, Ala866, Val867, Val914, Ile915, Thr916, Phe1047, and Gly1048; a relatively conserved hydrogen interaction region where hydrogen acceptor or donor substituents are preferred, regarding the existence of two polar amino acids (Glu885 and Asp1046) and a bulky, hydrophobic substituents favored region surrounded by several hydrophobic amino acids including Ile888, Leu889, Ile892, Val898, Val899, Leu1019, Cys1024, Ile1044, and Cys1045. Except for the smaller hydrophobic region, the other three regions were in perfect accordance with our previously published structure−activity relationship of the VEGFR-2 binding pocket on the basis of a 3D-QSAR analysis and molecular docking investigation.[6] Then by the Edit and Cluster Pharmacophore tool in Discovery Studio 2.5, a multicomplex-based pharmacophore was constructed (Figure S2b in the SI), which contains two hydrogen acceptors (A1, A2) in the hinge region, two hydrogen donors (D1, D2) in the hydrogen-bond conserved region, two hydrophobic features (H1, H2) and an aromatic ring (Ar1) in the two hydrophobic regions. Eight excluded volume features (Vol1−Vol8) were also included which reflected potential steric restrictions of amino acids Thr916, Val848, Val914, Gly841, Gly1048, Ile888, Leu1019, and His1026, accordingly. The original multicomplex pharmacophore (Raw_ph_and_vol) was first employed to screen the DUD-E database, and the results are displayed in Table S12 in the SI. Initially, the top 5% of resultant lists were saved for analysis, but only seven actives have been captured, and the selectivity (0.017) was fairly low. When taking into account the influence of excluded volumes, all of the eight excluded volume features were deleted, leading to a pruned pharmacophore (Raw_ph0). However, it did not render considerable improvement in terms of actives (8) and selectivity (0.02). It has been suggested that[21] pharmacophores with 3−7 features are more effective for VS. The poor performance of the pharmacophore may be due to the excessive number of features which was too restrictive for the VS; thus, we adopted a permutation and combination methodology to solve this thorny issue. The seven-feature multicomplex pharmacophore (Raw_ph0, Figure S2c in the SI) was pruned to refined pharmacophores with six, five, four, and three features by deleting pharmacophoric features one by one, and 99 pharmacophores were generated by the following formula:

$$C_7^0 + C_7^1 + C_7^2 + C_7^3 + C_7^4 = 99$$

All 99 pharmacophores are displayed in Table S12 in the SI using a method similar to binary form where 1 represents the corresponding feature existing in the pharmacophore and 0 indicates that the pharmacophore does not contain the feature. The 99 pharmacophores were utilized to screen the DUD-E database for selecting the most favorable multicomplex pharmacophore, and their screening results are also shown in Table S12 in the SI. The top 5% lists were kept for the initial selection. For tests where the total hits were less than 1254 compounds, they were all handled as lists of 1254 compounds for the consistency of comparison. To guarantee the selectivity, pharmacophores capturing at least 100 actives (Se ≈ 0.24) were selected. Finally, we achieved eight pharmacophores, of which five contained 4 features (Raw_ph29, Raw_ph39, Raw_ph44, Raw_ph46, Raw_ph57) and three contained 3 features (Raw_ph65, Raw_ph66, Raw_ph73). The more detailed screening results are calculated in the SI (Table S13) and depicted (Figure 9). As shown in Figure 9, at the stage of the top 5%, all eight pharmacophores acquired more than 100 actives with selectivity values ranging from 0.2 to 0.4, among which Raw_ph46 exerted the best performance with 141 actives and a selectivity of 0.34. They all demonstrated a relatively high specificity of more than 0.95 on average. In terms of EF, the values ranged from 5 to 7 at stages of the top 5% and top 0.1%, and Raw_ph46 reached 29, whereas the EF values of Raw_ph57 and Raw_ph65 were 0, indicating that no actives were captured for these two pharmacophores. All the AUC values exceeded 0.5 (a random level) but lower than 0.7, belonging to the defined fail level. In contrast, Raw_ph46 with an AUC value of 0.683 was thus regarded as the best. In summary, Raw_ph46 (Figure S2d in the SI) with four features (A1, D2, H1, and Ar1) was recognized as the best multicomplex pharmacophore for screening.

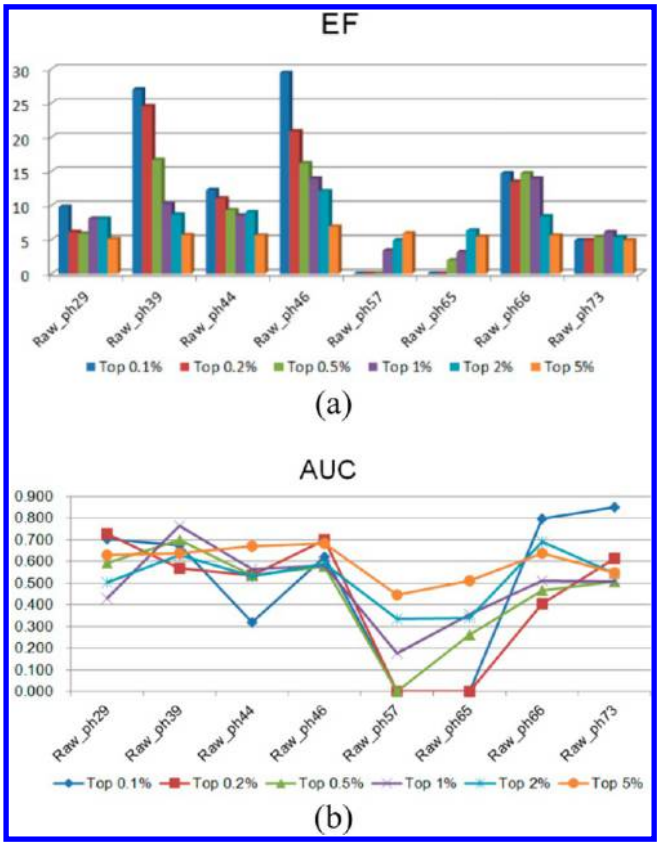


Figure 9. Selection of the most robust multicomplex pharmacophores. (a) EF values for the six selected multicomplex pharmacophores of the saved lists at six stages. (b) AUC values for the six selected multicomplex pharmacophores of the saved lists at six stages.

3.6. Comprehensive Comparison of Various Methods. After evaluating three methods individually through VS runs of

dx.doi.org/10.1021/ci400429g | J. Chem. Inf. Model. 2013, 53, 3163−3177

Table 2. Comparison of Virtual Screening Results for the Three Selected Optimal Solutions and Two Combinations[a]

	top 0.1% (25 compounds)					top 0.2% (50 compounds)					top 0.5% (125 compounds)				
solutions	actives	decoys	AUC	AUC_level	EF	actives	decoys	AUC	AUC_level	EF	actives	decoys	AUC	AUC_level	EF
ECFP_4	25	0	1*	excellent	61	50	0	1*	excellent	61	123	2	0.772	fair	60
Raw_ph46	12	13	0.622	poor	29	17	33	0.697	poor	21	33	92	0.578	fail	16
SP_3EWH	25	0	1*	excellent	61	44	6	0.807	good	54	85	40	0.737	fair	42
ECFP_4(SP_3EWH)	25	0	1*	excellent	61	50	0	1*	excellent	61	123	2	0.764	fair	60
SP_3EWH(ECFP_4)	**25**	**0**	**1***	**excellent**	**61**	**50**	**0**	**1***	**excellent**	**61**	**123**	**2**	**0.996**	**excellent**	**60**
	top 1% (251 compounds)					top 2% (502 compounds)					top 5% (1254 compounds)				
solutions	actives	decoys	AUC	AUC_level	EF	actives	decoys	AUC	AUC_level	EF	actives	decoys	AUC	AUC_level	EF
ECFP_4	184	67	0.923	excellent	45	225	277	0.874	good	27	263	991	0.887	good	13
Raw_ph46	57	194	0.581	fail	14	99	403	0.590	fail	12	141	1113	0.683	poor	7
SP_3EWH	117	134	0.806	good	29	152	350	0.794	fair	19	210	1044	0.783	fair	10
ECFP_4(SP_3EWH)	186	65	0.919	excellent	45	224	278	0.887	good	27	261	993	0.891	good	13
SP_3EWH(ECFP_4)	**186**	**65**	**0.927**	**excellent**	**45**	**224**	**278**	**0.895**	**good**	**27**	**261**	**993**	**0.870**	**good**	**13**
	true positives (TP)						selectivity (Se)						specificity (Sp)				
solutions	top 0.1%	top 0.2%	top 0.5%	top 1%	top 2%	top 5%	top 0.1%	top 0.2%	top 0.5%	top 1%	top 2%	top 5%	top 0.1%	top 0.2%	top 0.5%	top 1%	top 2%	top 5%
ECFP_4	25	50	123	184	225	263	0.06	0.12	0.30	0.45	0.55	0.64	1.00	1.00	1.00	0.99	0.99	0.96
Raw_ph46	12	17	33	57	99	141	0.03	0.04	0.08	0.14	0.24	0.34	1.00	1.00	1.00	1.00	0.98	0.95
SP_3EWH	25	44	85	117	152	210	0.06	0.11	0.21	0.29	0.37	0.51	1.00	1.00	1.00	1.00	0.99	0.96
ECFP_4(SP_3EWH)	25	50	123	186	224	261	0.06	0.12	0.30	0.45	0.55	0.64	1.00	1.00	1.00	0.99	0.99	0.96
SP_3EWH(ECFP_4)	**25**	**50**	**123**	**186**	**224**	**261**	**0.06**	**0.12**	**0.30**	**0.45**	**0.55**	**0.64**	**1.00**	**1.00**	**1.00**	**0.99**	**0.99**	**0.96**

[a]1* represents that no decoys were screened at the given stage.

dx.doi.org/10.1021/ci400429g | J. Chem. Inf. Model. 2013, 53, 3163−3177

the VEGFR-2 DUD-E validation database, we also explored a combination of them. Glide_SP (PDB 3EWH) selected from docking investigation, ECFP_4 selected from 2D fingerprints-based similarity search, and Raw_ph46 chosen from multi-complex-based pharmacophore exploration were further evaluated by a detailed comparison. Similarly, the five aspects of the six stages were gathered for analysis (Table 2). Together with Table 2, a set of statistical charts (Figure 10) was automatically generated for a better explanation.

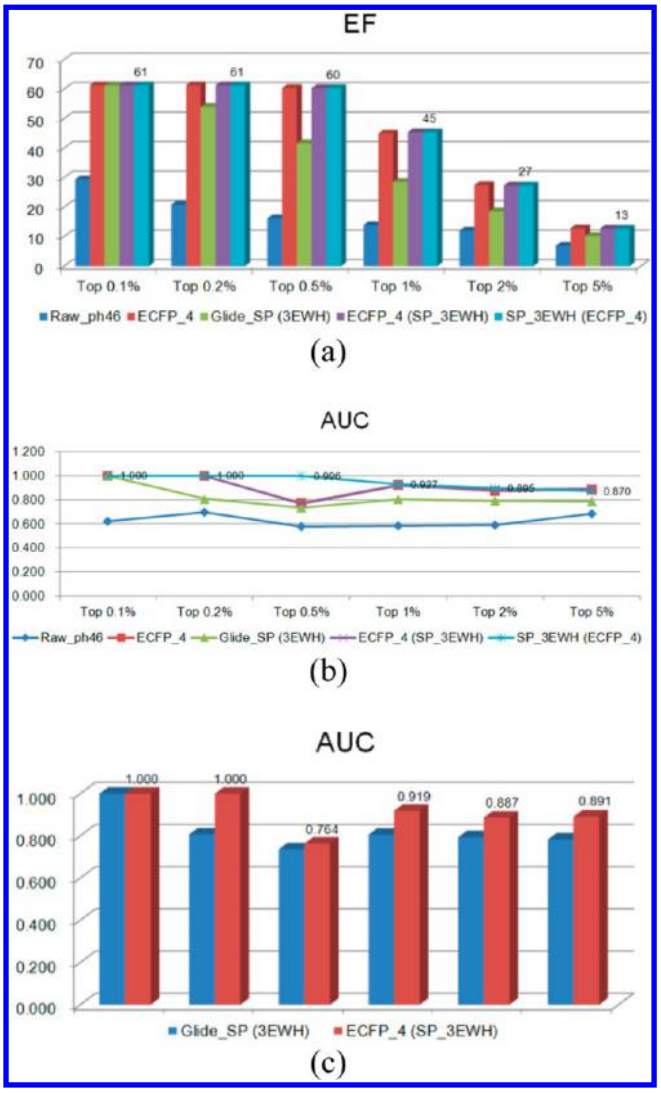


Figure 10. Comparison of the selected three optimal solutions and two combinations (a) EF values for the three optimal solutions and two combinations of the saved lists at six stages. (b) AUC values for the three optimal solutions and two combinations of the saved lists at six stages. (c) AUC comparison of the two combinations.

As shown in Table 2 and Figure 10, ECFP_4 exhibited the best performance in terms of the number of actives (263) and the selectivity (0.64). Followed by Glide_SP (3EWH), 210 actives with a selectivity of 0.51 were obtained. Raw_ph46 performed worst with 141 actives and a very low selectivity of 0.34. From the perspective of specificity, there was little difference for these three methods. At the stage of the top 5%, they reached 0.95, 0.96, and 0.96 for ECFP_4, Glide_SP (3EWH), and Raw_ph46, correspondingly. Moreover, the specificity values all reached 1* at the respective stages of the top 0.1%, top 0.2%, and top 0.5% for the three methods, indicating that all the captured compounds were actives. It was noticeable that even at the stage of the top 1%, ECFP_4 still exhibited excellent specificity of 1*. EF values were respectively 13, 10, and 7 for ECFP_4, Glide_SP (3EWH), and Raw_ph46 at the top 5% stage. ECFP_4 had achieved high EF values of 61 at stages of the top 0.1% and top 0.2%. Glide_SP (3EWH) also reached an equivalent level at the top 0.1% stage, but at the top 0.2% stage it obtained an EF value of 54. However, Raw_ph46 only achieved EF values of 29 and 21 at stages of the top 0.1% and top 0.2% respectively, which was rather low compared to the other two methods. Finally, ECFP_4 and Glide_SP (3EWH) performed much better than Raw_ph46 in terms of AUC values. At the stage of the top 5%, ECFP_4 reached a very satisfactory AUC value of 0.887 (good level) while Glide_SP (3EWH) obtained a Fail level AUC of 0.783. On the contrary, Raw_ph46 only reached 0.683, a predefined Fail level. Moreover, both ECFP_4 and Glide_SP (3EWH) acquired an excellent AUC level of 1* at the top 0.1% stage, demonstrating that all the screened compounds were actives, whereas Raw_ph46 still did not obtain a satisfactory result with its AUC level of just 0.622 (Fail level). In summary, ECFP_4 performed best followed by Glide_SP (3EWH). Raw_ph46 performed worst, indicating that the pruned four-feature multicomplex pharmacophore may not be appropriate for the VS of large VEGFR-2 databases. However, successful cases of this kind of pharmacophore-based VS were not very common, and general utilization of this method was performed on the basis of molecular alignment: for example, the molecular alignment before 3D-QSAR model (CoMFA and CoMSIA) and 3D-QSAR pharmacophore generation.[49,50]

As ECFP_4 and Glide_SP (3EWH) exerted satisfactory performance in the individual evaluation, ECFP_4 and Glide_SP (3EWH) were then integrated to improve the performance. As shown in Table 2, ECFP_4 (SP_3EWH) shows that the screening was first performed through Glide_SP (3EWH) docking followed by an ECFP_4 similarity search, while SP_3EWH (ECFP_4) adopted the opposite order. Almost all captured actives of the two combinations at all stages were more than that of Glide_SP (3EWH) individually. Both combinations displayed minor difference at stages of the top 5% and top 2% ECFP_4 individually, and the actives acquired at other four stages were comparable with that of ECFP_4. Furthermore, at the stage of the top 1%, the two combinations performed better than Glide_SP (3EWH) and ECFP_4 individually with two more actives obtained. The selectivity, specificity, and EF also showed similar trends, indicating the performance was as good as ECFP_4 and better than Glide_SP (3EWH).

Both combinations achieved good levels of AUC with values of 0.891 and 0.870 at the stage of the top 5%, which was equivalent with that of ECFP_4 (AUC: 0.887) and better than that of Glide_SP (3EWH) (AUC: 0.783). It is worth noting that at the top 0.5% stage where the AUC values were only on the fair level (0.772 for ECFP_4, 0.737 for Glide_SP (3EWH) and 0.764 for the combination ECFP_4 (SP_3EWH), the combination of SP_3EWH (ECFP_4) demonstrated an AUC value of 0.996 in the excellent level. Another major differentiation was the computational time for ECFP_4 as a 2D similarity search method which costs very little time, while Glide_SP as an accurate 3D database screening tool needs more time. Therefore, the combination of SP_3EWH (ECFP_4) was absolutely a better choice than the combination

of ECFP_4 (SP_3EWH). Besides, the ROC curves of the two combinations as well as Glide_SP (3EWH), ECFP_4 and Raw_ph46 are depicted in Figure 11, which conforms well to

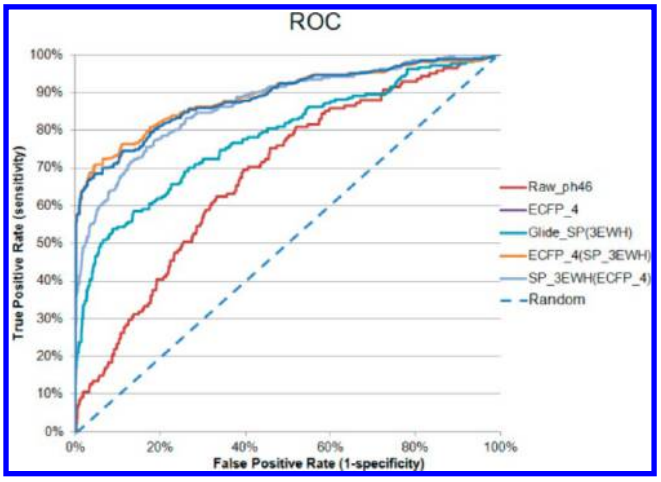


Figure 11. ROC curve of the selected three optimal solutions and two combinations.

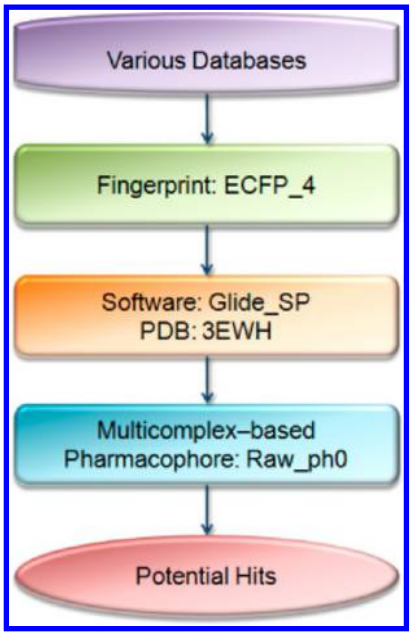


Figure 12. Final VS approach of the VEGFR-2 inhibitors.

the above analysis. The selected multicomplex pharmacophore Raw_ph46 did not perform as well as Glide_SP (3EWH) and ECFP_4; however, due to the fact that the eight captured molecules were all actives, we still investigated the multi-pharmacophore Raw_ph0 with seven features as the final step after the integration of ECFP_4 and Glide_SP (3EWH). Surprisingly, only 25 compounds were in the hit list, 23 actives, leading to a rather high EF of 56, 4 times higher than the combination of ECFP_4 and Glide_SP (3EWH). It suggested that a strict multicomplex pharmacophore serving as an efficient constraint would greatly reduce the difficulty in the final manual selection of potential hits through visual inspection. In summary, we could conclude that the combination of SP_3EWH (ECFP_4) was an efficient approach, i.e, 2D fingerprints by ECFP_4 followed by Glide_SP (3EWH), and if necessary by a strict multicomplex pharmacophore (Figure 12).

3.7. Comparison with Various Methods Published. As many published papers utilized the DUD 2.0 edition of VEGFR-2 inhibitors, for a parallel comparison, we applied the integrated protocol to the DUD 2.0 database. As the multicomplex pharmacophore filtration in the final step was an option when large hit lists were obtained, it was not considered in the comparison. Therefore, only the ECFP_4 and Glide_SP (3EWH) have been applied to the DUD 2.0 both in an individual and combined manner. The screening results are shown in Table S14 in the SI. We have reviewed several recently published papers[18,43,51−53] and found that an ROC Enrichment (ROCE) was used instead of EF. ROCE is described as the ratio of the true positive rate to the false positive rate when a given proportion of known decoys have been observed.[18,54] Superior to EF metric, it does not show a dependence on the ratio of actives to decoys. ROCE values at 0.5%, 1%, 2%, 5%, and 10% of true decoys screened and AUC values were calculated for our protocol and the methods collected from papers of another six groups[18,19,51−53,55,56] (SI Table S15) containing 45 methods in total. A histogram of the AUC (Figure 13a) and ROCE (Figure 13b) values is displayed for the comparison of our protocol and their methods. From Figure 13a and SI Table S15, we observed that the SP_3EWH (ECFP_4) has achieved the optimal results where its ROCE value at 0.5% decoys rate reached 67.99, the highest of all 49 methods. The LigMatch with single conformation method recommended by Kinnings et al.[51] also achieved satisfactory results with a ROCE value of 65.32 at stage 0.5%, and was slightly superior to the protocol at stages of 1%, 2%, and 5% decoys rate, respectively. A Shape_ele method validated by Hu et al.[18] obtained a ROCE of 8 at 5% decoys rate which was comparable with 7.47 of our protocol. In addition, a workflow containing a MACCS 2D similarity and pharmacophore postfilter after MOE dock developed by Planesas et al.[19] achieved a ROCE value of 4.13, which was also comparable to 4.11 of the protocol at 10% decoys rate. ROCE performance of other methods such as DOCK, FieldScreen, ROCS, and MM-GBSA etc. were much lower than the integrated approach. From the perspective of AUC (Figure 13b), we discovered that the AUC values of LigMatch with single conformation and multiple conformations of 0.8 and 0.76, respectively, and Shape_ele of 0.74 as well as ROCS with five conformations of 0.75 validated by Kirchmair et al.[56] were slightly higher than 0.71 of our protocol. Moreover, AUC values of 30 methods were below 0.6 and 16 methods below 0.5. In conclusion, only LigMatch and Shape_ele were comparable with the validated comprehensive protocol, and the other 43 methods mentioned by other groups performed inferiorly to our methods. In summary, all those data analyses strongly support the effectiveness of our integrated approach.

4. CONCLUSION

With a wealth of crystal structures of the VEGFR-2 in complex with various inhibitors, we have developed an integrated two-layer workflow to improve the VS performance. Molecular docking, multicomplex pharmacophore and fingerprint-based 2D similarity were explicitly investigated through the DUD-E validation database of VEGFR-2 inhibitors in an individual and a combined manner. The self- and cross-docking analyses indicated that Glide_SP combined with PDB 3EWH was the most favorable choice for molecular docking-based screening. Although the pruned multicomplex pharmacophores did not

dx.doi.org/10.1021/ci400429g | J. Chem. Inf. Model. 2013, 53, 3163−3177

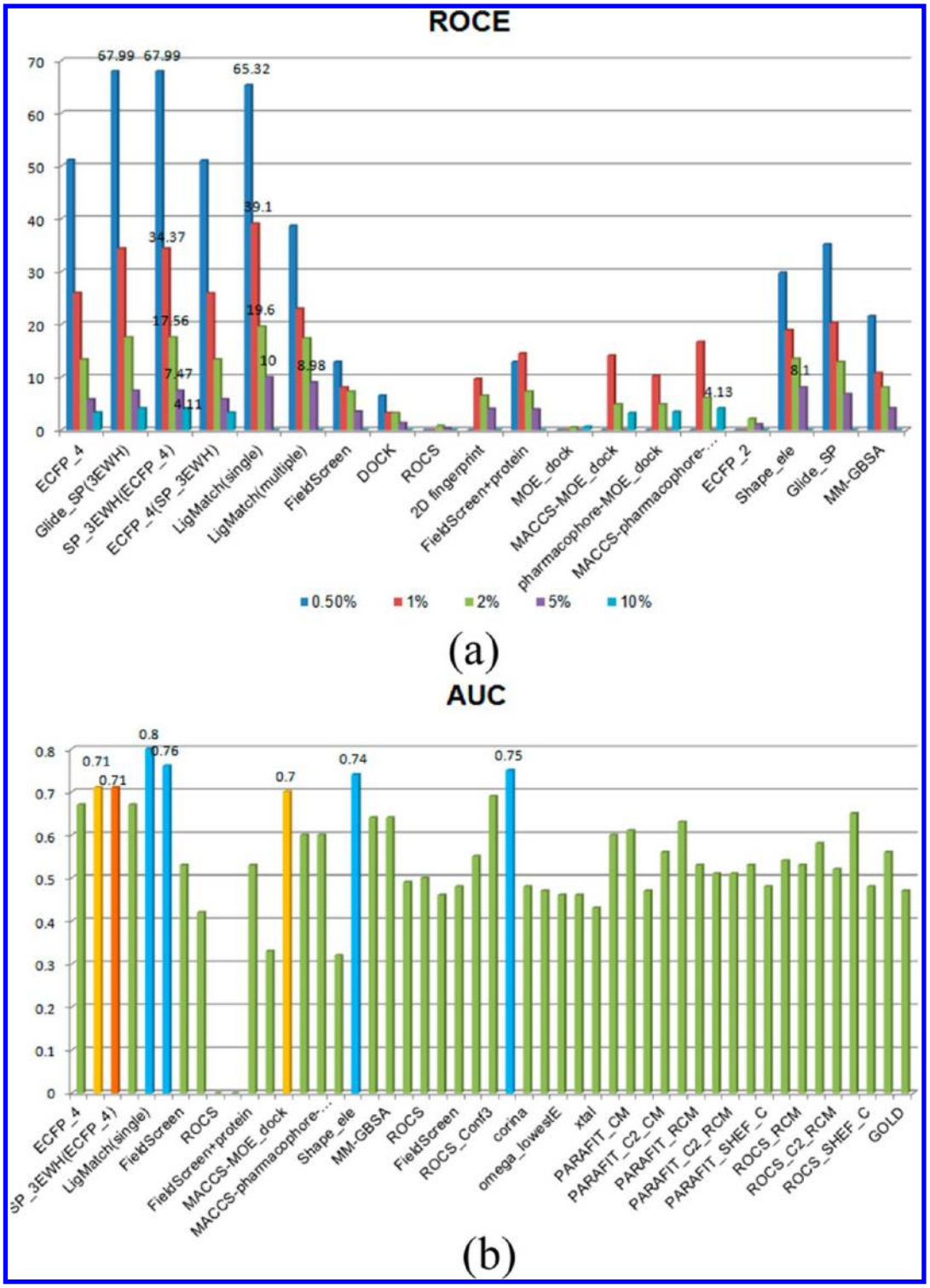


Figure 13. Comparison with other methods of six published papers.

perform as well as expected, the 2D fingerprint ECFP_4 yielded better performance among 13 fingerprints, and fingerprints (xCFPs) rather than counts (xCFCs) gave a significant improvement in the retrieval rates of active compounds. The comprehensive comparison of the best solution selected from three representative methods demonstrated that 2D fingerprint ECFP_4 outperformed Glide_SP (PDB: 3EWH) or multicomplex pharmacophore (Raw_ph46). Notably, the comparative study of ECFP_4 and Glide_SP (PDB: 3EWH) with recently published papers strongly support the effectiveness of the integrated approach as it outperformed 43 out of 45 methods. In summary, the developed two-layer workflow is applicable to identify potential inhibitors of VEGFR-2 and other drug targets by using a ligand-based method ECFP_4 followed by a structure-based method Glide_SP (PDB: 3EWH) docking, and optionally a strict multicomplex pharmacophore as the final filtration.

dx.doi.org/10.1021/ci400429g | J. Chem. Inf. Model. 2013, 53, 3163−3177

■ ASSOCIATED CONTENT

Supporting Information

Detailed information about the 31 VEGFR-2 PDB ligands; representation of multicomplex pharmacophores; 2D properties of VEGFR-2 PDB cognate ligands and DUD-E validation compounds; the original, statistical, and average statistical RMSD values of self-docking and cross-docking results; frequency distribution of statistical RMSD values of cross-docking results; VS results of Glide_SP docking, multicomplex pharmacophore screening, and 2D fingerprint-based similarity search, respectively; VS results of the integrated approach applied to DUD 2.0 database; and comparison of ROCE and AUC values with the former published papers. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*Telephone: +86-25-86185163. Fax: +86-25-86185182. E-mail: ydchen@cpu.edu.cn.

*Telephone: +86-25-86185180. Fax: +86-25-86185179. E-mail: lutao@cpu.edu.cn.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We sincerely thank Dr. Huifang Li for her professional advices in drafting the manuscript. This work was financially supported by the Fundamental Research Foundation for the Central Universities (No. JKQ2011004 and JKPZ2013002); the Natural Science Foundation of Jiangsu Province (No. 81302634 and No. 21302225), and the Postgraduate Innovative Foundation supported by Huahai Pharmaceutical Enterprise.

■ REFERENCES

(1) Kiselyov, A.; Balakin, K. V.; Tkachenko, S. E. VEGF/VEGFR signalling as a target for inhibiting angiogenesis. *Expert Opin. Investig. Drugs.* **2007**, *16*, 83−107.

(2) Musumeci, F.; Radi, M.; Brullo, C.; Schenone, S. Vascular Endothelial Growth Factor (VEGF) Receptors: Drugs and New Inhibitors. *J. Med. Chem.* **2012**, *55*, 10797−10822.

(3) Ivy, S. P.; Wick, J. Y.; Kaufman, B. M. An overview of small-molecule inhibitors of VEGFR signaling. *Nat. Rev. Clin. Oncol.* **2009**, *6*, 569−579.

(4) Huang, L.; Huang, Z.; Bai, Z.; Xie, R.; Sun, L.; Lin, K. Development and strategies of VEGFR-2/KDR inhibitors. *Future Med. Chem.* **2012**, *4*, 1839−1852.

(5) Boyer, S. J. Small molecule inhibitors of KDR (VEGFR-2) kinase: an overview of structure activity relationships. *Curr. Top. Med. Chem.* **2002**, *2*, 973−1000.

(6) Zhang, Y.; Liu, H.; Jiao, Y.; Yuan, H.; Wang, F.; Lu, S.; Yao, S.; Ke, Z.; Tai, W.; Jiang, Y. De novo design of N-(pyridin-4-ylmethyl) aniline derivatives as KDR inhibitors: 3D-QSAR, molecular fragment replacement, protein-ligand interaction fingerprint, and ADMET prediction. *Mol. Divers.* **2012**, *16*, 787−802.

(7) Wenzell, C. M. *HOPA News* 2013, *1*.

(8) DiGiulio, S. FDA Approves Stivarga for Advanced GIST. *Oncology Times* **2013**, *35* (6), 12.

(9) Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discovery.* **2002**, *1*, 882−894.

(10) Schneider, G. Virtual screening: an endless staircase? *Nat Rev. Drug Discovery* **2010**, *9*, 273−276.

(11) Tian, S.; Sun, H.; Li, Y.; Pan, P.; Li, D.; Hou, T. Development and evaluation of an integrated virtual screening strategy by combining molecular docking and pharmacophore searching based on multiple protein structures. *J. Chem. Inf. Model.* **2013**, *53*, 2743−2756.

(12) Hou, T.; Xu, X. Recent development and application of virtual screening in drug discovery: an overview. *Curr. Pharm. Des* **2004**, *10*, 1011−1033.

(13) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening–an overview. *Drug Discov. Today* **1998**, *3*, 160−178.

(14) Drwal, M. N.; Griffith, R. Combination of ligand- and structure-based methods in virtual screening. *Drug Discovery Today: Technol.* **2013**, *10*, 395.

(15) Hein, M.; Zilian, D.; Sotriffer, C. A. Docking compared to 3D-pharmacophores: the scoring function challenge. *Drug Discovery Today: Technol.* **2011**, *7*, e229−e236.

(16) Johnson, M. A.; Maggiora, G. M., *Concepts and applications of molecular similarity*; Wiley: New York, 1990; Vol. *8*.

(17) Baringhaus, K.-H.; Hessler, G. Fast similarity searching and screening hit analysis. *Drug Discovery Today: Technol* **2004**, *1*, 197−202.

(18) Hu, G.; Kuang, G.; Xiao, W.; Li, W.; Liu, G.; Tang, Y. Performance evaluation of 2D fingerprint and 3D shape similarity methods in virtual screening. *J. Chem. Inf. Model.* **2012**, *52*, 1103−1113.

(19) Planesas, J. M.; Claramunt, R. M.; Teixidó, J.; Borrell, J. I.; Pérez-Nueno, V. I. Improving VEGFR-2 docking-based screening by pharmacophore postfiltering and similarity search postprocessing. *J. Chem. Inf. Model.* **2011**, *51*, 777−787.

(20) Eckert, H.; Bajorath, J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discovery Today* **2007**, *12*, 225−233.

(21) Yang, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today* **2010**, *15*, 444−450.

(22) Zou, J.; Xie, H.-Z.; Yang, S.-Y.; Chen, J.-J.; Ren, J.-X.; Wei, Y.-Q. Towards more accurate pharmacophore modeling: Multicomplex-based comprehensive pharmacophore map and most-frequent-feature pharmacophore model of CDK2. *J. Mol. Graph. Model.* **2008**, *27*, 430.

(23) Wilson, G. L.; Lill, M. A. Integrating structure-based and ligand-based approaches for computational drug design. *Future Med. Chem.* **2011**, *3*, 735−750.

(24) Tanrikulu, Y.; Krüger, B.; Proschak, E. The holistic integration of virtual screening in drug discovery. *Drug Discovery Today* **2013**, *18*, 315−416.

(25) Krüger, D. M.; Evers, A. Comparison of Structure-and Ligand-Based Virtual Screening Protocols Considering Hit List Complementarity and Enrichment Factors. *ChemMedChem* **2010**, *5*, 148−158.

(26) 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.

(27) Berman, H. M.; Battistuz, T.; Bhat, T.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S. The protein data bank. *Acta Crystallogr., Sect D* **2002**, *58*, 899−907.

(28) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(29) 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.

(30) 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.

(31) Labute, P. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **2009**, *75*, 187−205.

(32) Konagurthu, A. S.; Whisstock, J. C.; Stuckey, P. J.; Lesk, A. M. MUSTANG: a multiple structural alignment algorithm. *Proteins* **2006**, *64*, 559−574.

(33) Wu, G.; Robertson, D. H.; Brooks, C. L.; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of

CDOCKER—A CHARMm-based MD docking algorithm. *J. Comput. Chem.* **2003**, *24*, 1549−1562.
(34) Mackey, M. D.; Melville, J. L. Better than random? The chemotype enrichment problem. *J. Chem. Inf. Model.* **2009**, *49*, 1154−1162.
(35) Bharatham, N.; Bharatham, K.; Lee, K. W. Pharmacophore identification and virtual screening for methionyl-tRNA synthetase inhibitors. *J. Mol. Graph. Model.* **2007**, *25*, 813−823.
(36) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins* **2003**, *52*, 609−623.
(37) Cheng, F.; Wang, Q.; Chen, M.; Quiocho, F. A.; Ma, J. Molecular docking study of the interactions between the thioesterase domain of human fatty acid synthase and its ligands. *Proteins* **2008**, *70*, 1228−1234.
(38) Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2005**, *45*, 160−169.
(39) Jacobsson, M.; Lidén, P.; Stjernschantz, E.; Boström, H.; Norinder, U. Improving structure-based virtual screening by multivariate analysis of scoring data. *J. Med. Chem.* **2003**, *46*, 5781−5789.
(40) Hecker, E. A.; Duraiswami, C.; Andrea, T. A.; Diller, D. J. Use of catalyst pharmacophore models for screening of large combinatorial libraries. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1204−1211.
(41) Diller, D. J.; Li, R. Kinases, homology models, and high throughput docking. *J. Med. Chem.* **2003**, *46*, 4638−4647.
(42) Hamza, A.; Wei, N.-N.; Zhan, C.-G. Ligand-based virtual screening approach using a new scoring function. *J. Chem. Inf. Model.* **2012**, *52*, 963−974.
(43) Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput. Aided Mol. Des.* **2008**, *22*, 133−139.
(44) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J.-P.; Bertrand, H.-O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* **2005**, *48*, 2534−2547.
(45) Enyedy, I. J.; Egan, W. J. Can we use docking and scoring for hit-to-lead optimization? *J. Comput. Aided Mol. Des.* **2008**, *22*, 161−168.
(46) Miyazaki, Y.; Matsunaga, S.; Tang, J.; Maeda, Y.; Nakano, M.; Philippe, R. J.; Shibahara, M.; Liu, W.; Sato, H.; Wang, L. Novel 4-amino-furo [2, 3-*d*] pyrimidines as Tie-2 and VEGFR2 dual inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2203−2207.
(47) Harmange, J.-C.; Weiss, M. M.; Germain, J.; Polverino, A. J.; Borg, G.; Bready, J.; Chen, D.; Choquette, D.; Coxon, A.; DeMelfi, T. Naphthamides as Novel and Potent Vascular Endothelial Growth Factor Receptor Tyrosine Kinase Inhibitors: Design, Synthesis, and Evaluation. *J. Med. Chem.* **2008**, *51*, 1649−1667.
(48) Cee, V. J.; Cheng, A. C.; Romero, K.; Bellon, S.; Mohr, C.; Whittington, D. A.; Bak, A.; Bready, J.; Caenepeel, S.; Coxon, A. Pyridyl-pyrimidine benzimidazole derivatives as potent, selective, and orally bioavailable inhibitors of Tie-2 kinase. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 424−427.
(49) He, G.; Qiu, M.; Li, R.; Ouyang, L.; Wu, F.; Song, X.; Cheng, L.; Xiang, M.; Yu, L. Multicomplex-Based Pharmacophore-Guided 3D-QSAR Studies of N-Substituted 2′-(Aminoaryl) Benzothiazoles as Aurora-A Inhibitors. *Chem. Biol. Drug. Des.* **2012**, *79*, 960−971.
(50) Urniaz, R. D.; Jozwiak, K. X-ray Crystallographic Structures as a Source of Ligand Alignment in 3D-QSAR. *J. Chem. Inf. Model.* **2013**, *53*, 1406−1414.
(51) Kinnings, S. L.; Jackson, R. M. LigMatch: a multiple structure-based ligand matching method for 3D virtual screening. *J. Chem. Inf. Model.* **2009**, *49*, 2056−66.
(52) 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.
(53) Pérez-Nueno, V. I.; Ritchie, D. W. Using consensus-shape clustering to identify promiscuous ligands and protein targets and to choose the right query for shape-based virtual screening. *J. Chem. Inf. Model.* **2011**, *51*, 1233−1248.
(54) Jahn, A.; Hinselmann, G.; Fechner, N.; Zell, A. Optimal assignment methods for ligand-based virtual screening. *J. Cheminf.* **2009**, *1*, 14.
(55) 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.
(56) 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.