Comparative Evaluation of 3D Virtual Ligand Screening Methods: Impact of the Molecular Alignment on Enrichment

David Giganti,[†,‡] Hélène Guillemain,[†,§] Jean-Louis Spadoni,[§] Michael Nilges,[‡] Jean-François Zagury,[§] and Matthieu Montes*[,§]

Unité de Bioinformatique Structurale, Institut Pasteur, 26 rue du Dr Roux, 75015 Paris, France, and Chaire de Bioinformatique, Conservatoire National des Arts et Métiers, 292 rue Saint Martin, 75003 Paris, France

Received December 26, 2009

In the early stage of drug discovery programs, when the structure of a complex involving a target and a small molecule is available, structure-based virtual ligand screening methods are generally preferred. However, ligand-based strategies like shape-similarity search methods can also be applied. Shape-similarity search methods consist in exploring a pseudo-binding-site derived from the known small molecule used as a reference. Several of these methods use conformational sampling algorithms which are also shared by corresponding docking methods: for example Surflex-dock/Surflex-sim, FlexX/FlexS, ICM, and OMEGA-FRED/OMEGA-ROCS. Using 11 systems issued from the challenging "own" subsets of the Directory of Useful Decoys (DUD-own), we evaluated and compared the performance of the above-cited programs in terms of molecular alignment accuracy, enrichment in active compounds, and enrichment in different chemotypes (scaffold-hopping). Since molecular alignment is a crucial aspect of performance for the different methods, we have assessed its impact on enrichment. We have also illustrated the paradox of retrieving active compounds with good scores even if they are inaccurately positioned. Finally, we have highlighted possible positive aspects of using shape-based approaches in drug-discovery protocols when the structure of the target in complex with a small molecule is known.

INTRODUCTION

In the early stage of research of drug discovery programs, high-throughput screening (HTS) procedures can be applied for hit identification in large small molecule databases. In the past decade, in silico screening has been extensively used to reduce the number of compounds going into HTS, reducing time and costs for hit finding.[1]

The classical straightforward concept aiming at identifying analogues by comparing the physicochemical, structural, or pharmacophoric properties of a known active molecule with that of compounds in a collection has been massively applied during the last decades.[2,3] Initially, these ligand-based virtual ligand screening (LBVLS) methods were based on simple 2D descriptors or fingerprints[4] derived from the structure of the reference active compound and compared to the corresponding descriptors of database compounds using a similarity metric, such as the Tanimoto coefficient (Tc). These methods were generally efficient, very fast, and provided as a result hits sharing a common chemotype with the active molecule used as the reference.[5]

To increase the structural diversity of the hits provided by LBVLS methods and thus to perform "scaffold-hopping" (i.e., change the chemotype, keep the activity[6]), different methods using more sophisticated 3D descriptors have later been developed, such as pharmacophore screening[7] or shape similarity searching.[8,9]

In pharmacophore screening, the knowledge of a set of aligned known active compounds is required, in contrast to shape similarity search methods that only require the structure of a single active compound. Shape similarity search methods thus appear as the LBVLS methods of choice when the structure of only a single active compound is available.

When the structure of the target in complex with a ligand is available, structure-based virtual ligand screening (SBVLS) methods like docking/scoring[10] or structure-based pharmacophore screening[11,12] are generally preferred. However, the coordinates of a bound ligand can also be used as a reference for shape similarity search methods. Using such methods, virtual screening success stories have recently been reported.[13−16]

It appears that the performance of virtual ligand screening methods depends on many factors, such as the selection of the query structure and its conformation,[17] the conformational search of the compounds,[18] and the quality of the alignment produced.[19]

Recently published shape/volume similarity searches, that is, 3D-ligand-based virtual ligand screening (3D-LBVLS) methods have been developed by the same research groups that had previously developed docking-based virtual ligand screening (docking-BVLS) methods such as Surflex-Sim/Surflex-dock, the ICM package, FlexX-FlexS, and OMEGA-FRED/OMEGA-ROCS. All of these 3D-LBVLS methods share their conformation generator with their respective docking-BVLS method "counterpart".

We thus decided in the present work to evaluate these four different 3D-LBVLS methods and their docking-BVLS methods counterparts. First, we wished to assess the ability of these four 3D-LBVLS methods and their docking-BVLS

* To whom correspondence should be addressed: Matthieu Montes, PhD, e-mail: matthieu.montes@cnam.fr.
† These authors contributed equally to this work.
‡ Unité de Bioinformatique Structurale, Institut Pasteur.
§ Chaire de Bioinformatique, Conservatoire National des Arts et Métiers.

10.1021/ci900507g
Published on Web 06/08/2010

methods counterparts to perform accurate molecular alignments using a large number of active compounds on 11 different targets issued from the DUD data set. Then, we wished to evaluate their performance in virtual screening and scaffold-hopping on the same challenging data sets. Recently, Kirchmair et al.[17] highlighted the importance of the query and its conformation for the quality of enrichment after virtual screening. We decided to proceed further in this direction by exploring the impact of molecular alignment on enrichment. This consisted in evaluating the impact of (1) the quality of the alignment of the query with the structure used as a reference on the enrichment obtained with 3D-LBVLS methods and (2) the docking accuracy on enrichment with docking-BVLS methods. From the results of this study, we expect to outline the most favorable approach to be used when the structure of the target in a complex with a small molecule is available and to illustrate a currently unexplained paradox between the quality of molecular alignment and ranking, that is, enrichment.

MATERIAL AND METHODS

All the 3D-LBVLS methods used in this study have a docking-BVLS counterpart with which they share an important part of the molecular alignment algorithm. For a given docking engine, the algorithm used to position, in a flexible manner, the query molecule into the binding site, according to putative ligand−receptor interaction points, is similar to the algorithm used to superimpose, in a flexible manner, the query molecule on the reference active compound, according to its shape and putative common interaction points. The biggest differences stem from the scoring functions used and from the volume to be explored as a protein binding site is generally wider than the pseudosite defined around a given known ligand. Hence, for each system, we carefully adjusted the binding site definition in all the programs to perform a realistic comparison of the methods in terms of volume to be explored. Keeping this idea of consistence between the different packages, we deliberately chose not to tune the other parameters of the different 3D-VLS programs for the difficult systems, in contrast to other studies.[20,21]

Computational Methods. *FlexX/FlexS.* FlexX[22] and FlexS[23] are a docking/scoring method and a shape-similarity search method, respectively, based on a fragmentation/reconstruction algorithm, to dock compounds flexibly into the binding site (FlexX) or to flexibly superimpose compounds on the reference active compound (FlexS). The query molecule is decomposed into fragments, whose physicochemical properties are represented by a set of spheric (FlexX) or Gaussian (FlexS) functions. The fragment defined as the base fragment is aligned with the reference active compound using the RigFit algorithm[23] (FlexS) or positioned into the binding site via a hashing-technique (FlexX). In FlexX the scoring is made using a modified Bohm empirical scoring function.[24,25] In FlexS, the scoring is made using a 3D similarity metric.[23] FlexX version 2.0.3 and FlexS 1.20.3 were used for all calculations.

Surflex-dock/Surflex-sim. Surflex-dock[26] and Surflex-sim[27] use a modified Hammerhead fragmentation/reconstruction algorithm[28] to dock compounds flexibly into the binding site (Surflex-dock) and to superimpose flexibly onto the reference active compound (Surflex-sim), respectively. The query molecule is decomposed into rigid fragments that are superimposed to the Surflex-protomol, that is, molecular fragments covering the entire binding site (Surflex-dock) or to the molecular surface points of the reference active compounds[29] (Surflex-sim). In Surflex-dock, the molecules are evaluated by an empirical scoring function. In Surflex-sim, the similarity of each fragment with the reference active compound is evaluated according to the alignment of their respective surface points. In this study, Surflex-dock and Surflex-sim, version 2.2, were used for all calculations.

ICM. ICM is based on Monte Carlo simulations in internal coordinates to optimize the position of molecules using a stochastic global optimization procedure combined with pseudo-Brownian positional/torsional steps and fast local gradient minimization.[30] The docking poses were evaluated using ICM-VLS empirical scoring function.[31] Chemical superposition with ICM-sim were performed using the atomic property fields method (APF[32]). ICM, version 3.4, was used for all calculations.

OMEGA/FRED/ROCS. FRED[33] and ROCS use atom-centered Gaussian functions parametrized to provide close approximations to hard sphere volumes. In FRED, version 2.2.3, used in this study, orientations of a single conformer are compared with a bump-grid defined using the active site during the rigid body docking procedure. Poses clashing with the bump-grid are eliminated and the remaining poses are evaluated using ShapeGauss.[34] In ROCS,[33] shape-similarity is evaluated by maximizing the volume overlap between the reference active compound and a single conformation of a query molecule using the Tanimoto coefficient. In version 2.3.1, used in this study, a "color force field" represents physicochemical properties in addition to the shape component. The conformational search of the different query compounds (up to 100 conformers per compound) has been carried out prior to all the calculations using OMEGA, version 2.1.[33]

A summary of the characteristics of the different programs evaluated in this study is available in Table 1.

DUD Database. To evaluate the ability of the methods to discriminate between active molecules and decoys and thus their enrichment performance, we decided to use the directory of useful decoys database (DUD).[21] DUD is a benchmarking data set designed for docking method evaluation containing annotated active compounds (from 30 to 120) for 40 targets, including 36 decoys for each active molecule issued from the ZINC database.[35] In the DUD, the decoys have been chosen for their similar physicochemical properties (molecular weight, logP, number of hydrogen bond donors/acceptors, etc.) with known active molecules, which renders them harder to discriminate compared to random decoys. To be as challenging as possible, the DUD-own set was selected for each target. Each DUD-own set contains only active and decoy compounds designed for the corresponding target making it more challenging than the DUD-all set.[36,37] Because of its careful construction and its high rate of decoys (36 decoys for one active compound), it appears today as the most challenging available benchmarking database.[17] Moreover, as stated recently by Hawkins et al,[37] the recent modifications of DUD 2.0 (including new active compounds and structure clustering by Good[38]) make it more suitable for ligand-based method evaluation.

Table 1. Main Characteristics of the Programs Evaluated in This Study: Flexibility Algorithm, Energy Terms Used in the Scoring Function, and Average Computational Cost in Seconds Per Molecule[a]

program	conformational search algorithm	scoring terms	average time per compound (s)
Surflex-dock	fragmentation/incremental construction	steric, hydrogen bonds, hydrophobic, pseudoentropic (rotatable bonds, size)	14.6
Surflex-sim	fragmentation/incremental construction	steric, polar (ionic + hydrogen bonds), hydrophobic	6.7
FRED	via OMEGA (fragmentation/ incremental construction)	steric, shape complementarity	1.0
ROCS	via OMEGA (fragmentation/ incremental construction)	shape, hydrogen bonds, hydrophobic, aromatic	0.5
FlexX	fragmentation/incremental construction	hydrogen bonds, ionic, hydrophobic, aromatic, pseudoentropic (rotatable bonds)	15.6
FlexS	fragmentation/incremental construction	steric, hydrogen bonds, ionic, hydrophobic	6.9
ICM	metropolis Monte Carlo	van der Waals, electrostatic, hydrophobic, hydrogen bonds desolvation, pseudoentropic (rotatable bond), size correction	17.7
ICMsim	metropolis Monte Carlo	hydrogen bonds, ionic, hydrophobic, aromatic, pseudoentropic (rotatable bonds), atomic fields (electrostatic, size)	2.4

[a] All calculations have been performed on Intel Xeon 2.4 GHz processors with 2Go of RAM.

Table 2. Description of the 11 Systems[a]

target	PDB	accessible surface (Å^2)	alternate PDB	problematic side chains	no. rot. actives (min-max)	MW actives (min-max) (g/mol)	no. rot. ref	MW ref (g/mol)
ADA	1NDW	477.05	2Z7G	L62, F65	0−8	152.1−316.1	6	256.14
CDK2	1CKP	838.29			1−10	235.1−483.2	2	259.1
DHFR	3DFR	1223.14			1−11	206.1−523.2	10	458.2
ER	3ERT	787.23			5−11	347.2−508.2	8	388.2
FXA	1F0R	872.11	1LPG	Y100, R147	4−12	293.1−574	5	450.1
HIVRT	1RT1	1774.43			1−12	244.1−582	6	302.2
NA	1A4G	1191.96			3−13	194.1−409.2	8	332.2
P38	1KV2	1217.09	1OUK, 1BL6, 1BL7, 2EWA	F160	2−10	239.1−509.2	10	523.3
THR	1BA8	971.39			2−12	250.1−647.2	19	469.2
TK	1KIM	313.44	1QHI, 2KI5	Q125	2−5	184.1−404	2	239.1
TRP	1BJU	756.41			1−12	124.1−620.3	5	283.1

[a] The accessible surface of the binding sites for each system has been computed using StrucTools. Minimum and maximum molecular weight (MW), number of rotatable bonds (no. rot) for the set of DUD-own active compounds and DUD-own reference compound are also included.

We downloaded DUD release 2 from the Web site http://dud.docking.org and carefully selected 11 out of the 40 targets available (ADA, CDK2, DHFR, ER, FXA, HIVRT, NA, P38, THR, TK, TRP) according to their presence in the literature for benchmarking studies and their diversity in binding site properties and in active compounds according to ICM. Hydrogen atoms were added in the DUD protein structures using ICM.

Performance Metrics. All enrichment graphics were produced with the statistical and graphical tool R (http://www.r-project.org/). To complete the information of enrichment graphics, receiver operating characteristics curves were plotted with the ROCR package. The area under the ROC curve (AUC) was calculated on the base of the Wilcoxon−Mann−Whitney algorithm.[39]

Ligand Cluster Definition. We classified consequently the DUD-own active compounds by ligand similarity using chemical descriptor fingerprints and Tanimoto similarity distance (Td) as implemented in ICM. We examined the resulting trees and visually inspected all the DUD-own active compounds to select a harmonized Td cutoff that resulted into at least 2 equilibrated clusters in each DUD-own data set. After multiple tests, Td was defined at 0.55.

RESULTS

Presentation of the 11 Systems. The diversity of the binding site properties of the targets selected for this study is presented in Table 2. Buried or partially buried binding sites are the most frequent (i.e., ADA, CDK2, DHFR, ER, HIVRT, NA and TK), but 2 proteins display a binding site more accessible to solvent (i.e., FXA and TRP). Some binding sites are essentially hydrophobic (i.e., CDK2, ER, HIVRT and P38), whereas others should display mainly polar interactions with a potential ligand (i.e., ADA, DHFR, NA, TK, and TRP). We also provide the number of active compounds and decoys for each system, which range respectively from 22 for TK to 454 for P38 and from 891 for TK to 9141 for P38.

Positioning: Molecular Alignment and Docking Accuracy. For all 3D-VLS methods, the conformational search is a crucial point for (1) producing a correct alignment of the known active compounds on the reference for 3D-LBVLS methods[19] and (2) producing accurate poses for docking-BVLS methods.[40]

Molecular alignment evaluation is in general performed through cognate ligand docking with one active compound per system, using a database of known protein/small molecule complexes derived from the Protein Data Bank, such as the Fisher set.[41] Since one of the goals of the present work was to evaluate the impact of molecular alignment on enrichment, we decided to use a large number of active compounds issued from the DUD data set. Since very limited structural information was known for all these compounds, we assumed their binding mode to be somewhat similar to the one provided as a reference for each system (considering its scaffold). We thus defined an expert knowledge-based visual

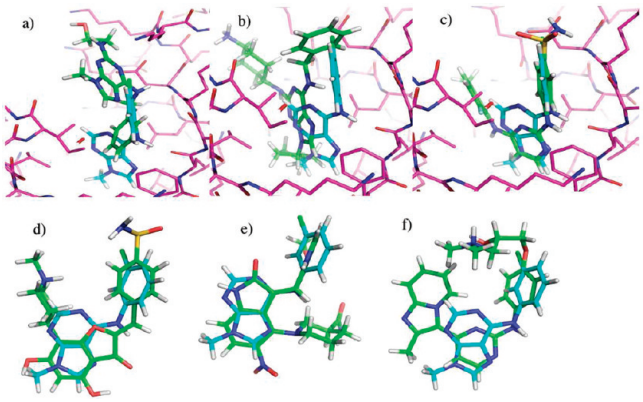

Figure 1. Examples of "0", "1", and "2" poses. The quality of molecular alignment with representative CDK2 active compounds is illustrated for docking-BVLS methods in a, b, and c and for 3D-LBVLS methods in d, e, and f. The illustration a represents a "0" pose for the ZINC01641925 compound; b represents a "1" pose for ZINC03591113, and c represents a "2" pose for ZINC03814453. Similarly, d represents a "0" pose for ZINC03814433; e represents a "1" pose for ZINC04617748, and f represents a "2" pose for ZINC03814454.

score for assessing the superimposition accuracy and the docking accuracy of the 3D-VLS methods evaluated in this study. For 3D-LBVLS methods, we used "2" for an accurate superimposition of most of the pharmacophoric interaction points, "1" for an orientation of the compound similar to the reference and at least 50% of well-superimposed pharmacophoric points, and "0" for the others. Similarly, we assessed the docking accuracy using a visual score evaluating the binding mode of the reference active compound provided by the DUD. We used "2" for an accurate pose displaying most of the expected interactions with the residues of the binding site, "1" for an orientation of the compound similar to the reference displaying 50% of the expected interactions with the residues of the binding site, and "0" for the others. We defined acceptable poses as "1 + 2" poses and unacceptable poses as "0" poses. Examples of "0", "1", and "2" poses for both 3D-LBVLS and docking-BVLS methods are presented in Figure 1. All compound positioning results are presented in Table 3.

Overall, all the 3D-LBVLS methods are able to superimpose most of the active compounds on the corresponding reference provided by the DUD as the rate of acceptable poses (visual score >0) is above 68% for all the programs. ROCS and FlexS display a very similar performance with 68.46% and 69.25%, respectively, Surflex and ICM-sim being the most efficient with an overall acceptable pose rate of 82.22% and 90.97%, respectively. As expected, the performance of the methods depends on the structure of the active compounds for each target, ADA-compounds displaying a high global rate of unacceptable poses for all the methods (mean rate of 53.21%). Some other systems seem to cause problems to 3D-LBVLS methods such as HIVRT-actives (more than 34% of unacceptable poses for all the methods but Surflex-sim) and FXA-actives (more than 50% of unacceptable poses for all the methods but ICM-sim).

Concerning docking-BVLS methods, similar trends were observed. In general, the methods provided acceptable binding modes, the most efficient being ICM-dock with an overall acceptable pose rate of 82.01%. FRED displays 79.64% of acceptable pose rate (very close to ICM-dock) and is even the best method on DHFR, ER, and P38. Surflex-dock exhibited the best performance for TK, NA, and TRP despite displaying the least overall docking accuracy with 59.93% of acceptable poses.

As expected, the global performance of all the methods were correct and depended on the structure of the target. Overall, the methods performed accurately on FXA, NA, TK, TRP (except for FRED), and DHFR (except for FlexX). Some targets such as THR (except for ICM-dock), P38, HIVRT, and ER (except for FRED) seemed to cause recurrent problems.

Enrichment in Active Compounds on the DUD-own Databases. Because the 3D-VLS methods evaluated in this study were able to provide accurate molecular alignments, we assessed their performance on enrichment of the DUD-own databases corresponding to the 11 targets we selected for this study using as a reference the compound provided by the DUD. Through the analysis of the corresponding enrichment graphs, we both evaluated their performance in early enrichment (within the top 1% of the database) and in late enrichment (within the top 10% of the database). Results are presented in Figure 2 and Table 4.

Overall, three out of the four 3D-LBVLS methods evaluated in this study (ROCS, FlexS, ICM-sim) display an early enrichment rate of 13%, ROCS displaying the best average performance with 14.12% of active compounds retrieved within the first percent of the corresponding DUD-own database and the best early enrichment rate with 34.69% observed on NA. ICM-sim seems to perform better on late enrichment with 57.25% of actives retrieved in average within the first 10% of the corresponding DUD-own database.

In similarity with the quality of superimposition, we observed a dependence of the enrichment gained by the different methods on the system tested. All the shape similarity search programs evaluated in this study displayed a good performance on ADA, ER, NA, and TK and performed poorly on P38, the best average performance being observed on ER. For THR, FXA, DHFR, and CDK2, only ICM-sim and ROCS for the two latter systems provided acceptable enrichment rates. Concerning the individual performance on each target, ICM-sim appears to be the best method on early and late enrichment on 4 out of the 11 targets (DHFR, FXA, THR, TRP) and displays the best late enrichment rate with 95.92% on TRP. ROCS and FlexS both perform best on 2 systems out of the 11: HIVRT/P38 for FlexS and CDK2/NA for ROCS.

Concerning the docking-BVLS methods, we compared the results with DOCK enrichments published on the same DUD-own databases (the only ones published with these databases to our knowledge). Overall, Surflex-dock displayed the best early and late enrichment rates with 12.36% and 45.36%, respectively, of the active compounds retrieved in the 1% and 10% of the DUD-own database. It appears as the best method for early and late enrichment rates for ADA (7.69%−48.72%), FXA (17.12%−59.59%) and TRP (18.37%−81.6%), followed by ICM-dock, which is the best for DHFR (19.02%−66.83%), ER (17.95%−53.85%), and NA (34.69%−85.71%). FRED and FlexX follow, being the best method respectively for TK (18.18%−22.73%) and THR (6.94%−52.78%). Compared to the methods evaluated herein, DOCK was the best on early enrichment for ADA and DHFR.

Table 3. Molecular Alignment of the DUD-own Active Compounds[a]

		ADA (39)	CDK2 (72)	DHFR (410)	ER (39)	FXA (146)	HIVRT (43)	NA (49)	P38 (454)	THR (72)	TK (22)	TRP (49)	mean (1100)
Surflex-sim	2	12.82	12.5	44.39	56.41	4.11	18.6	18.37	9.69	4.17	68.18	40.82	23.15
	1	35.9	51.39	53.66	43.59	33.56	72.09	75.51	69.6	93.06	31.82	59.18	59.07
	0	51.28	36.11	1.95	0	62.33	9.3	6.12	20.7	2.78	0	0	17.78
ROCS	2	12.82	43.06	74.15	41.03	14.38	30.23	79.59	9.47	2.78	86.36	8.16	35.63
	1	20.51	36.11	14.39	56.41	28.77	34.88	14.29	41.19	69.44	9.09	81.63	32.83
	0	66.67	20.83	11.46	2.56	56.85	34.88	6.12	49.34	27.78	4.55	10.2	31.54
FlexS	2	15.38	25	56.1	48.72	35.62	16.28	36.73	14.98	30.56	81.82	18.37	33.48
	1	20.51	19.44	11.71	43.59	13.7	46.51	57.14	74.01	5.56	13.64	2.04	35.77
	0	64.1	55.56	32.2	7.69	50.68	37.21	6.12	11.01	63.89	4.55	79.59	30.75
ICMsim	2	46.15	12.5	86.83	51.28	70.55	18.6	75.51	20.04	88.89	90.91	69.39	54.48
	1	23.08	68.06	11.95	46.15	16.44	46.51	14.29	68.28	9.72	9.09	28.57	36.49
	0	30.77	19.44	1.22	2.56	13.01	34.88	10.2	11.67	1.39	0	2.04	9.03
		ADA (39)	CDK2 (72)	DHFR (410)	ER (39)	FXA (146)	HIVRT (43)	NA (49)	P38 (454)	THR (72)	TK (22)	TRP (49)	mean (1100)
Surflex-dock	2	10.26	23.61	70.98	46.15	25.34	30.23	36.73	6.83	15.28	22.73	61.22	34.05
	1	56.41	45.83	9.02	15.38	57.53	30.23	61.22	18.06	26.39	77.27	36.73	25.88
	0	33.33	30.56	20	38.46	17.12	39.53	2.04	75.11	58.33	0	2.04	40.07
FRED	2	2.56	51.39	79.51	33.33	17.12	41.86	46.94	11.89	15.28	22.73	4.08	36.92
	1	20.51	29.17	8.78	51.28	65.07	32.56	38.78	67.62	40.28	59.09	69.39	42.72
	0	76.92	19.44	11.71	15.38	17.81	25.58	14.29	20.48	44.44	18.18	26.53	20.36
FlexX	2	7.69	18.06	45.61	25.64	16.44	16.28	65.31	9.47	40.28	68.18	36.73	27.31
	1	66.67	81.94	20	41.03	62.33	23.26	30.61	40.31	27.78	18.18	61.22	38.42
	0	25.64	0	34.39	33.33	21.23	60.47	4.08	50.22	31.94	13.64	2.04	34.27
ICM	2	17.95	50	79.51	51.28	28.08	37.21	73.47	12.11	34.72	9.09	24.49	41.29
	1	76.92	41.67	2.93	5.13	57.53	41.86	6.12	65.42	54.17	81.82	71.43	40.72
	0	5.13	8.33	17.56	43.59	14.38	20.93	20.41	22.47	11.11	9.09	4.08	17.99

[a] The part of "0", "1", and "2" poses are presented in the table as percentage of all the active compounds. Between brackets is the number of active compounds for each system.

Overall, global enrichment was in the favor of 3D-LBVLS methods for 8 out of the 11 systems evaluated in this study. This is particularly striking for ER and TK, where all the 3D-LBVLS methods outperform all the docking-BVLS methods. We observed a mixed performance for both types of methods, with a slightly better outcome of the 3D-LBVLS methods for NA, TRP, HIVRT, and DHFR.

There were exceptions to this statement: for CDK2 and FXA, docking-BVLS methods globally outperformed the 3D-LBVLS methods and for P38, all the methods exhibited a poor enrichment.

Chemotype Enrichment Analysis by 3D-VLS Methods. A critical assessment of the robustness of the methods could be performed by evaluating their ability to retrieve structures with different chemotypes and thus to perform "scaffold-hopping".[42,43] Similarly to the work published by Good,[38] we thus assessed the structural diversity of the different DUD-own active compounds by assigning them to their respective clusters according to a defined Tanimoto similarity distance descriptor threshold (Td < 0.55).

We defined cluster enrichment via a simple procedure: a cluster is counted as present if at least one of its members is retrieved within the subset sampled. Results are presented on Figure 3 using cluster enrichment graphs. Overall, the enrichment in active compounds and in clusters follows a similar trend for 3D-LBVLS methods. Slight differences can be noticed in the individual performance of the methods for some systems. Surflex-sim performed efficiently in cluster enrichment for CDK2, P38, NA, and TK and was one of the best methods for these systems. Similarly, FlexS displayed a good performance on cluster enrichment for TRP, HIVRT, P38, NA, and FXA. Regarding docking-BVLS methods, their global performance was good for cluster enrichment. This was particularly striking on ER, FXA, and CDK2, where all docking-BVLS methods displayed a very good performance. By comparing the relative individual performance of docking-BVLS methods, it seems that the global trend was conserved for most of the systems used in this study. For 8 out of 11 systems, a method that displayed the best performance in active compounds enrichment on a defined target also displayed the best performance in cluster enrichment. Striking differences in cluster enrichment could be observed with FlexX that displayed much better performance for DHFR and with ICM that displayed much better performance for THR and TRP. Overall, both 3D-LBVLS and docking-BVLS methods retrieved more than 50% of the clusters in the top 10% for 6 out of the 11 systems.

To highlight global trends between 3D-LBVLS methods and docking-BVLS methods in enrichment, we created mean enrichment graphs using the means of the ranks obtained after screening the DUD-own databases. As the performance of each individual method varies depending on the target, an advantage of such representation is to enhance the differences in the global performance for 3D-LBVLS methods and docking-BVLS methods. Mean enrichment graphs for cluster enrichment (Td > 0.55) and active compounds enrichment are presented on Figure 4.

As expected, global trends in enrichment of combined docking-BVLS methods and combined 3D-LBVLS methods were similar to the trends of individual methods. The global performance of docking-BVLS methods was relatively poor and similar to 3D-LBVLS methods in cluster enrichment for HIVRT and THR. Overall docking-BVLS methods displayed the best performance on 5 out of the 11 targets, it was thus equivalent on this test to 3D-LBVLS methods.

Impact of Molecular Alignment on Enrichment. It is obvious that, at least for 3D-LBVLS methods, there should

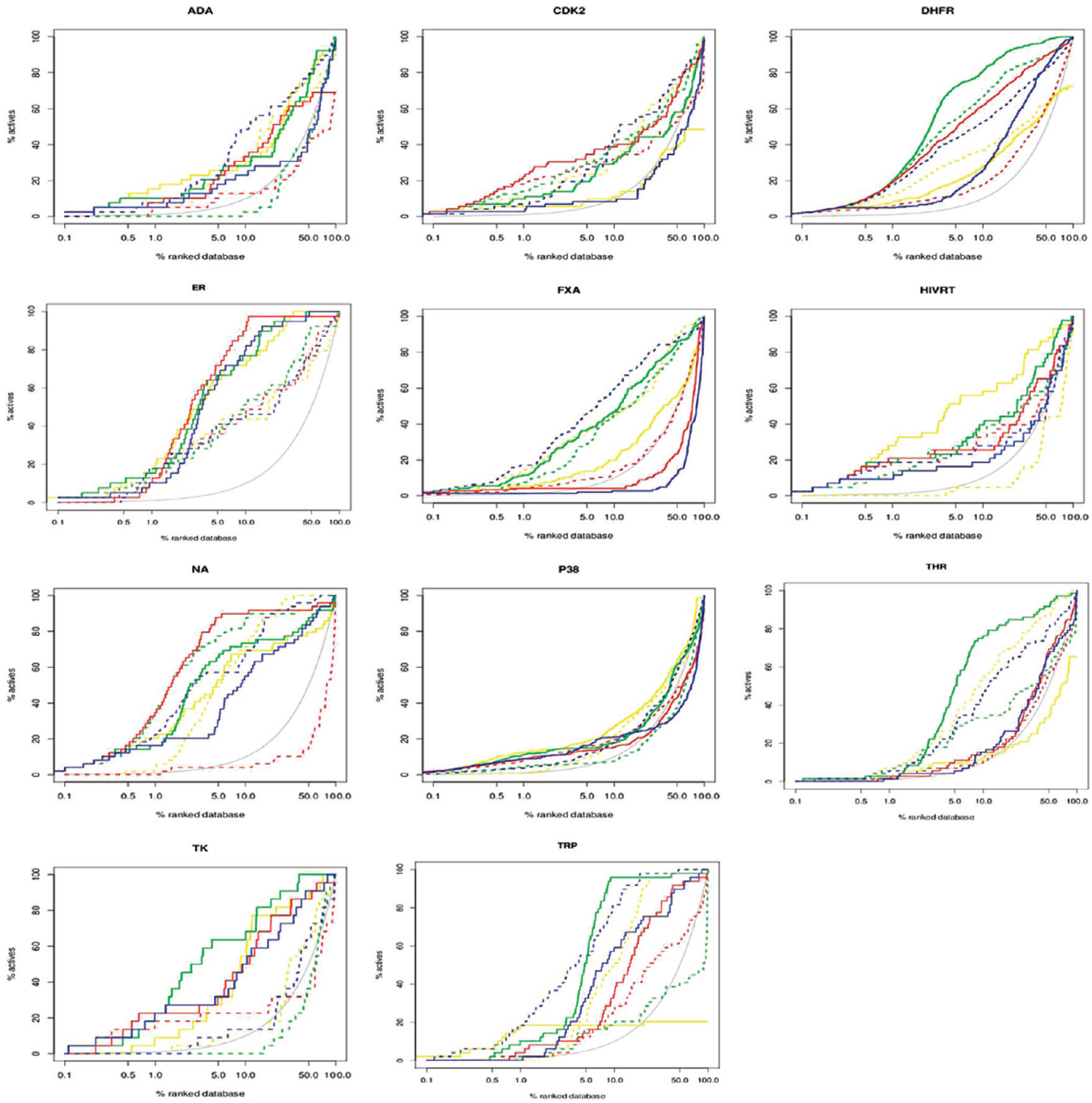

Figure 2. Enrichment graphs with docking-BVLS methods (dotted lines) and 3D-LBVLS methods (plain lines). The gray line represents random enrichment. Each color is for a package: blue for Surflex-dock/Surflex-sim, green for ICM/ICM-sim, yellow for FlexX/FlexS, and red for FRED/ROCS.

be a direct correlation between the quality of the alignment of the active compounds on the reference compound and their rank in the database. To highlight this, we plotted mean enrichment graphs for the acceptable and unacceptable poses of the active compounds of the corresponding DUD-own database provided by the methods evaluated in this study. These mean enrichment graphs are presented in Figure 5.

It clearly appears from the analysis of these graphs that, for 3D-LBVLS methods, compounds that are accurately superimposed to the reference active compound provided by the DUD are present in the early ranks. For 8 out of the 11 systems (all but HIVRT, P38 and NA), more than 95% of the active compounds available in early enrichment (1% threshold) are accurately superimposed (visual score = 2). At a 10% threshold, most of the active compounds retrieved are superimposed in an acceptable manner (visual score = 1 and 2) but up to 50% of these actives are superimposed with a visual score of 1. This highlights that the less accurate the superposition provided by the methods the lower the score. It also shows that the scoring functions used for the shape-similarity search methods evaluated in this study are efficient and reliable.

Concerning the docking-BVLS methods, at a 1% threshold, more than 95% of the active compounds were accurately positioned in the binding site only for DHFR. For all systems but TK, CDK2, and TRP, a non-negligible part of the ligands that ranked in the top 10% of the corresponding DUD-own database displayed an inaccurate positioning, the most

Table 4. Early (1%) and Late (10%) Enrichments for the DUD-own Active Compounds[a]

enrichment	Surflex-sim		ROCS		FlexS		ICMsim		DOCK	
all actives	1.00%	10.00%	1.00%	10.00%	1.00%	10.00%	1.00%	10.00%	1.00%	10.00%
ADA	5.13	23.08	7.69	33.33	15.38	30.77	10.26	28.21	15	40
CDK2	2.78	8.33	22.22	38.89	5.56	11.11	9.72	30.56	10	25
DHFR	4.88	25.61	19.02	61.95	8.05	27.8	20.24	80.73	25	60
ER	10.26	82.05	10.26	89.74	15.38	71.79	17.95	76.92	10	20
FXA	1.37	2.74	4.11	4.11	5.48	19.86	11.64	46.58	5	35
HIVRT	9.3	18.6	20.93	25.58	27.91	58.14	18.6	39.53	5	25
NA	16.33	55.1	34.69	89.8	20.41	69.39	16.33	73.47	10	60
P38	9.03	21.15	8.81	15.42	12.56	26.21	11.23	17.84	2	25
thrombin	1.39	15.28	0.75	15.28	2.78	12.5	2.78	76.39	5	35
TK	22.73	50	22.73	50	9.09	54.55	22.73	63.64	0	25
trypsin	0	59.18	4.08	34.69	18.37	18.37	10.2	95.92	0	30
mean	7.56	32.83	14.12	41.71	12.82	36.41	13.79	57.25	7.91	34.55
st. dev.	6.99	24.91	10.48	28.79	7.57	22.72	5.85	25.77	7.35	13.87
median	5.13	23.08	10.26	34.69	12.56	27.8	11.64	63.64	5	30
enrichment	Surflex-dock		FRED		FlexX		ICM		DOCK	
all actives	1.00%	10.00%	1.00%	10.00%	1.00%	10.00%	1.00%	10.00%	1.00%	10.00%
ADA	7.69	48.72	5.13	12.82	5.13	33.33	0	2.56	15	40
CDK2	6.94	44.44	18.06	31.94	11.11	29.17	13.89	38.89	10	25
DHFR	17.56	52.93	5.85	17.56	11.95	36.83	19.02	66.83	25	60
ER	17.95	43.59	12.82	48.72	15.38	43.59	17.95	53.85	10	20
FXA	17.12	59.59	4.11	13.7	14.38	44.52	7.53	45.21	5	35
HIVRT	18.6	27.91	18.6	32.56	0	4.65	11.63	37.21	5	25
NA	22.45	67.35	0	4.08	4.08	69.39	34.69	85.71	10	60
P38	3.74	18.06	7.93	19.6	1.1	21.59	4.63	7.49	2	25
thrombin	5.56	47.22	1.39	11.11	6.94	52.78	5.56	33.33	5	35
TK	0	13.64	18.18	22.73	0	13.64	0	0	0	25
trypsin	18.37	81.6	2.04	26.53	0	51.02	2.04	18.37	0	30
mean	12.36	45.91	8.56	21.94	6.37	36.41	10.63	35.4	7.91	34.55
st. dev.	7.63	20.32	7.13	12.5	5.96	18.7	10.42	27.16	7.35	13.87
median	17.12	47.22	5.85	19.6	5.13	36.83	7.53	37.21	5	30

[a] Corresponding results with DOCK from the DUD original paper (Huang et al. J. Med. Chem. 2006) are also presented for information.

striking examples being for THR and ER where 10.45% and 6.45%, respectively, of the active compounds retrieved in the top 10% were docked inaccurately.

Overall, using docking-BVLS, 0.57% of the active compounds retrieved in the top 1% of the ranked database were inaccurately positioned. This rate reached 3.38% when considering the top 10% of the ranked database. Concerning 3D-LBVLS methods, the results were respectively 0.13% for 1% of the ranked database and 1.11% for 10% of the ranked database.

We could also observe that out of the active compounds ranked late in enrichment (from 10% to 100%), 16.15% and 10.82% were positioned accurately by docking-BVLS and 3D-LBVLS, respectively, methods. This highlights that even though 3D-VLS methods provide accurate molecular alignments, a non-negligible part of the active compounds can still be missed because of scoring.

DISCUSSION

Molecular Alignment and Docking Accuracy. From the observation of the results, we can conclude that both 3D-LBVLS and docking-BVLS methods are able to produce conformations similar to the bioactive conformation. Docking-BVLS methods are more challenging since the conformational space of a binding pocket is more complex than the pseudoreceptor defined by the reference ligand in 3D-LBVLS methods. But they display a good performance in terms of docking accuracy even on difficult targets where artifacts because of the construction of the DUD may sometimes affect the results. In particular, when the protein is considered as rigid (as in all the methods evaluated in this study), it may be very difficult for some ligands to be docked into the binding site of the structure provided by the DUD. For example, in the P38 structure, there are 82 active compounds (i.e., 18% of the P38 active compounds) that cannot be docked into the binding site because of major clashes occurring in the current protein conformation. We retrieved from the PDB database structures corresponding to some of the 82 active compounds that caused problems with the P38 structure, and it seems that there is a major shift of F169 that caused this docking issue (see figure 6). Similar problems can be observed with the structures of FXA, ADA, and TK, where some of the active compounds cannot bind properly to the currently provided structure because of the orientation of some key side chains.

Concerning the relative performance of the methods, it seems that for both docking-BVLS and 3D-LBVLS, ICM displays the best performance in terms of molecular alignment and docking accuracy. This may be caused by a more effective treatment of the flexibility of the compounds in ICM via its biased probability Monte Carlo procedure.[44] FRED displays an overall very good performance despite being the fastest docking method. It is even the most effective docking method on DHFR, P38, and ER, which pose problems to the other docking methods. These targets display the largest binding sites, making the conformational search within these

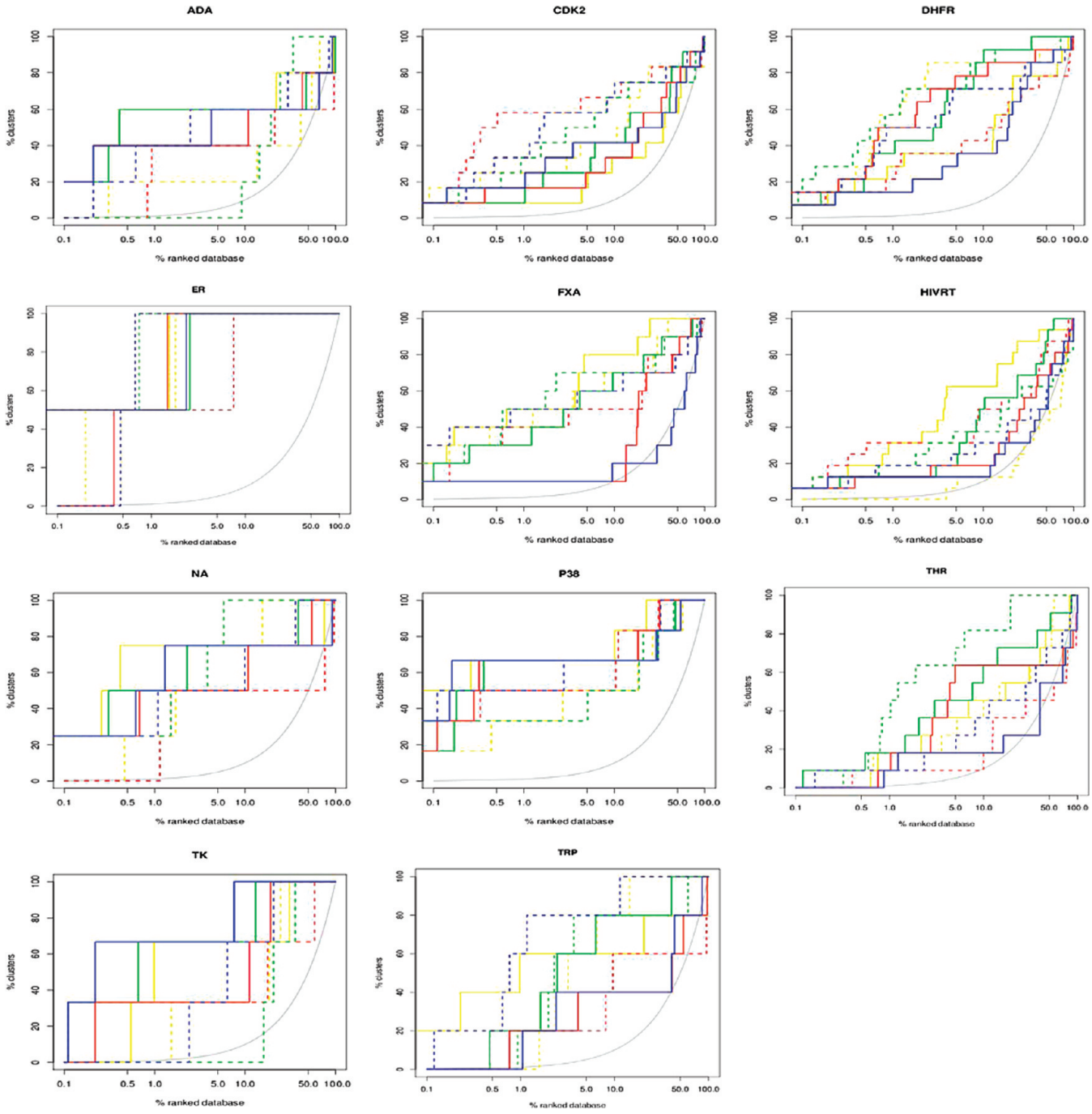

Figure 3. Cluster enrichment graphs with docking-BVLS methods (dotted lines) and 3D-LBVLS methods (plain lines). Each color is for a package: blue for Surflex-dock/Surflex-sim, green for ICM/ICM-sim, yellow for FlexX/FlexS, and red for FRED/ROCS.

sites more difficult. One of the possible reasons for the good performance of FRED on these 3 systems could be that FRED focuses on the optimization of only translational and rotational degrees of freedom during the calculation, which seems more effective for exploring such large binding sites than using "fully flexible" ligands. For FRED calculation, the torsional degrees of freedom are estimated using multiconformer libraries generated by OMEGA. Also, FRED calculations were performed using shape-gauss to evaluate the ligand poses, which is known to be relatively clash-tolerant when using default settings.[34] This could also be a possible reason for its success compared to the other methods for P38, taking into consideration that a non-negligible part of the DUD-active compounds cannot bind to the provided structure because of the conformation of F169.

Surflex-dock appears to be the best on 3 systems (TK, NA, TRP) despite displaying the least overall docking accuracy (59.93% of acceptable poses), principally, because of its very poor performance on P38 and THR.

Overall the different programs evaluated in this study are able to perform acceptable molecular alignments that lead to an accurate superimposition of the active compounds on the reference for 3D-LBVLS methods and acceptable docking poses for docking-BVLS methods.

Comparative Performance in Enrichment. The quality of virtual ligand screening methods is evaluated by their

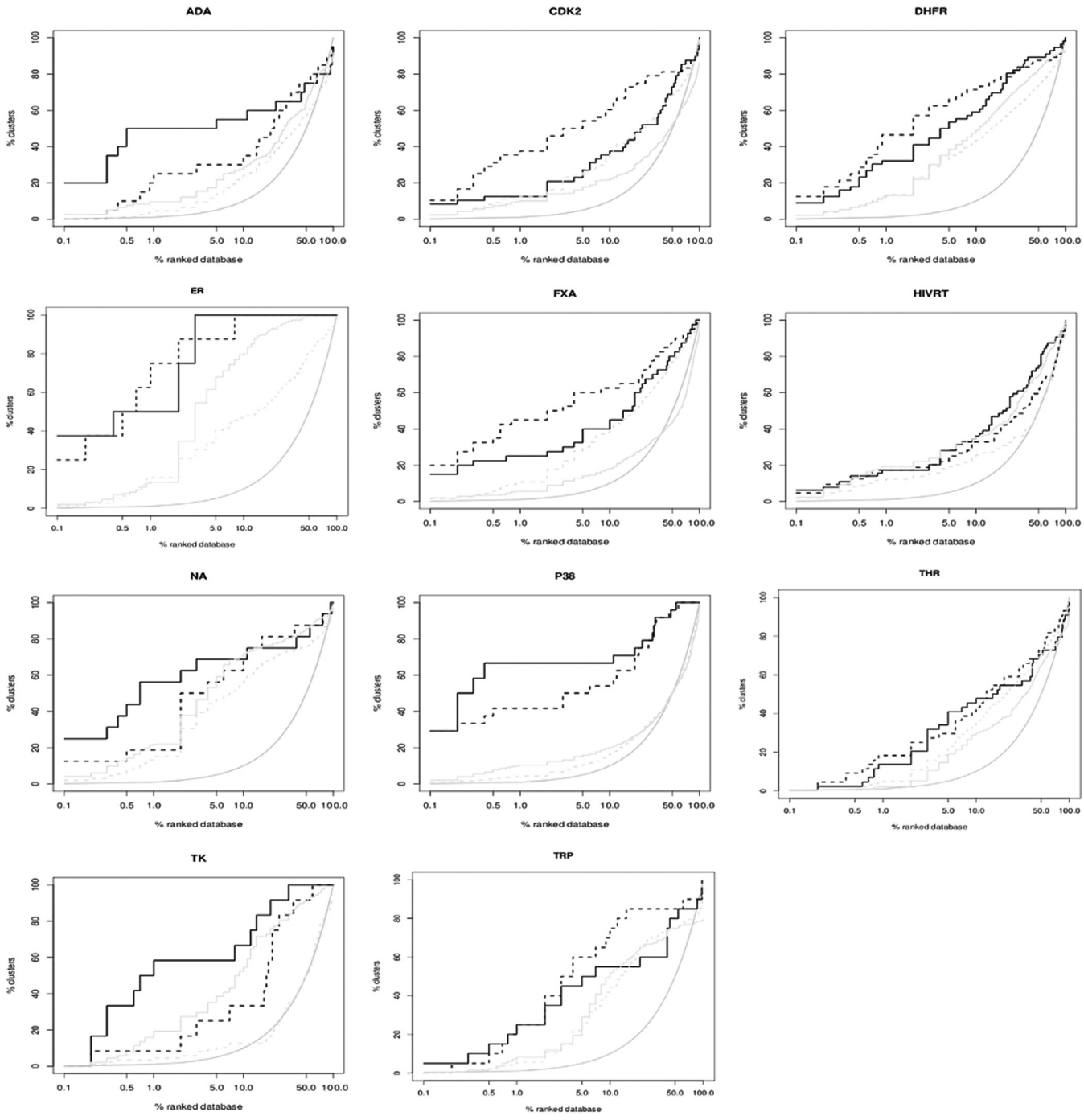

Figure 4. Mean enrichment graphs (gray) and mean cluster enrichment graphs (black) obtained by docking-BVLS methods (dotted lines) and 3D-LBVLS methods (plain lines).

ability to retrieve active compounds within the early ranked database, the so-called enrichment.

Overall, mean early and late enrichments are clearly in favor of 3D-LBVLS methods (9.48%/34.91% for docking-BVLS versus 12.07%/42.05% for 3D-LBVLS). This is particularly striking for ER where systematically all the 3D-LBVLS methods outperform docking-BVLS methods. Von Korff[45] suggested that this should not be surprising because of the construction of the DUD database as, decoys are mostly topologically dissimilar to the active compounds.[21] However, as shown by Cleves,[36] the impact of this bias is important principally for 2D similarity search methods. An opposite opinion is given by Kirchmair[17] when he suggests that docking-BVLS methods are in fact favored in the DUD considering that docking has, more likely, data present on the dimensions of the active site. We believe that shape-similarity search methods are less sensitive to the possible bias highlighted by Von Korff and Cleves as several features directly related to the topology such as the bond and atom types are not primarily considered in 3D-LBVLS methods and, as Kirchmair stated,[17] that ligand-based methods have to represent the binding site properties using a single active compound, which is already very challenging on its own.

Even considering possible bias, docking-BVLS methods are still very effective on 6 out of the 11 systems evaluated in this study (NA, TRP, HIVRT, DHFR, CDK2 and FXA), the most striking performance being with CDK2 and FXA.

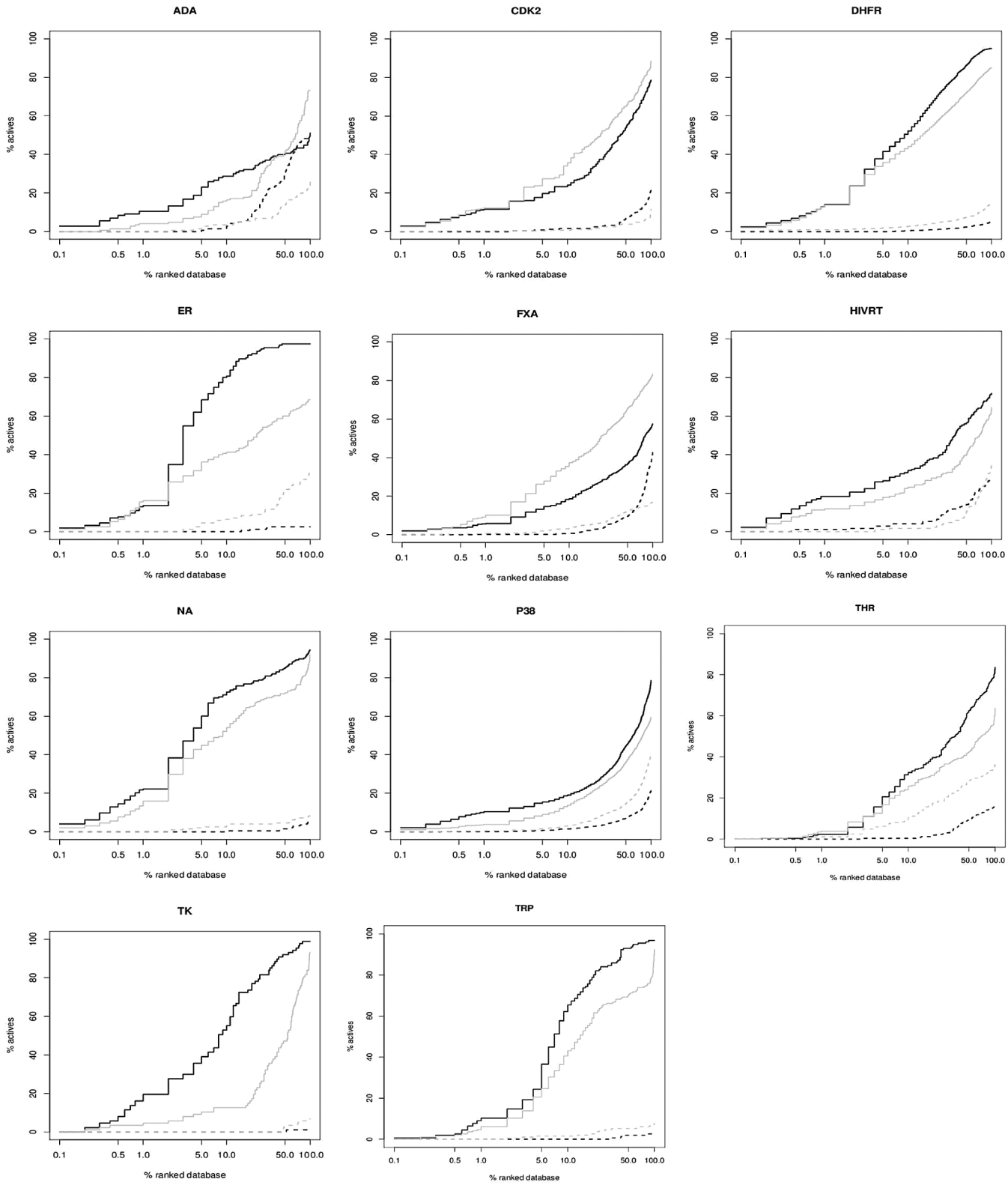

Figure 5. Molecular alignment of the active compounds using 3D-LBVLS methods (plain lines) and docking-BVLS methods (dotted lines). Light gray line shows the random enrichment. Black lines are for acceptable poses and gray lines for unacceptable poses.

Surflex-dock appears to be the most effective docking-BVLS method on enrichment. It also outperforms most of the 3D-LBVLS methods in several systems. By comparing our results with Cross et al,[20] the Surflex-dock ringflex parameter seems clearly valuable at least for serine proteases where we could see a significative impact (data not shown). It improves the enrichment for 5 out of the 11 systems evaluated in this study (ADA, CDK2, ER, FXA, and THR).

FRED appears to show its limits in enrichment except on the systems where the more complex docking methods fail, such as HIVRT, P38, and TK. This shows that in particularly difficult systems, multiconformer rigid-body docking methods can perform at least, as well as flexible docking. As for docking accuracy, one possible explanation can be the clash tolerance of the shape-gauss scoring function implemented in FRED for these targets where several clashes can occur

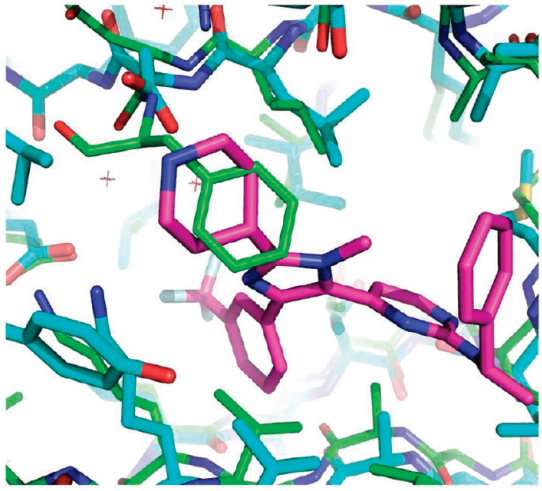

Figure 6. Illustration of a clash between F169 of the P38 structure provided by the DUD (green) and one of the compounds from the P38 DUD-own active set, ZINC03815672, positioned using ICM (pink). The blue structure represents an alternate PDB that could have accepted this compound.

with a non-negligible part of the active compounds of the corresponding DUD-own data set, for example, in the currently provided structure for P38 (see the docking-accuracy section), and in the small and buried binding site of TK and HIVRT.

The variance observed between the results is important within the DUD targets and, as observed by von Korff,[45] highlights the importance of multiple systems for benchmarking studies.

Comparison with Other Studies. On 2 out of the 11 targets evaluated in this study (ADA and DHFR), DOCK results from the original DUD paper[21] display a better performance in early enrichment compared to the other docking methods. For the other targets, its performance is comparable to FlexX.

As expected, our results with ROCS are almost identical to those obtained by Kirchmair.[17] The slight differences we observe can be explained by the different settings used for the multiconformer generation of the DUD-own sets with OMEGA but seem to have a very low impact on the final performance of the program. It is reensuring to have similar results using the same data sets and encouraging for using standardized data sets, such as the DUD for benchmarking studies. As proposed by the creators of the DUD,[21] a broadened use of such data sets will be beneficiary for the whole community to compare their benchmarking results.

By comparing the ROC AUC with the early and late enrichment rates, we can draw several observations. ROC AUC is a good measure of the global discriminating performance of the different methods since we use the same conditions for testing. In general, similar conclusions can be drawn from the analysis of ROC curves/ROC AUC and enrichment graphs around 10% or 20% of the ranked database but in early enrichment it is not necessarily the case.[17] For "real life" project decisions, the important point about the methods is the definition of the selection threshold to use, that is, the number of compounds to test experimentally after a virtual screening. This measure, in our opinion, cannot be necessarily provided using a metric, such as ROC AUC or even the analysis of ROC curves but rather by classical enrichment values comparisons. For instance, in the case of ADA, by observing the ROC curves and AUC, the best method seems to be ROCS but in fact in early and late enrichment, it appears that Surflex-dock is the best method to use even if its ROC AUC is 10% lower. This highlights the point stated by Kirchmair[17] that both ROC AUC and enrichment rates need to be considered in method evaluation and thus in the definition of new drug-discovery protocols. Once more, the use of standardized metrics in benchmarking studies will benefit the whole community.

Comparative Performance in Scaffold Enrichment. As expected, methods that display a good performance in enrichment also display a good performance in scaffold-enrichment. It is widely accepted that docking methods should be appropriate to retrieve compounds with different molecular scaffolds, as a protein's binding site is more permissive than a pseudoreceptor derived from ligand-based hypothesis. Indeed, docking-BVLS methods perform well in scaffold enrichment, retrieving in average more than 50% of the different clusters in the top 10% for 7 out of the 11 systems studied here. However, for TK, NA, ADA, and P38, 3D-LBVLS methods outperform significantly docking-BVLS methods, highlighting their good performance in scaffold-hopping. By analyzing the trends from the mean enrichment graphs, we can conclude that overall, 3D-VLS methods have a similar acceptable performance in scaffold enrichment, depending on the system. It is thus of importance to underscore that, in the systems tested in the present study, docking-BVLS methods do not perform systematically better in scaffold-hopping than 3D-LBVLS methods. This makes 3D-LBVLS methods also reliable for retrieving diverse molecular scaffolds.

As a confirmation, we also tested the ability to retrieve compounds within the same cluster as the reference active compound provided by the DUD (thus more similar compounds i.e., the inner ability of LBVLS methods). As expected, in this trivial test for ligand-based methods, 3D-LBVLS methods and ROCS in particular display a much better overall performance, outperforming systematically docking-BVLS methods (11 out of the 11 systems studied, data not shown).

Molecular Alignment and Enrichment. It should be obvious that, using 3D-VLS methods, compounds that are retrieved in the top scores should be positioned accurately in the binding site/superimposed accurately on the reference compound (i.e., accurate molecular alignment). We thus analyzed the molecular alignment of the different active compounds after 3D-VLS.

As expected, for the 3D-LBVLS programs evaluated in this study, more than 80% of the active compounds retrieved in the top 10% are positioned in an acceptable manner, while a very low number display unacceptable poses: 19/338 for Surflex-sim, 8/500 for ROCS, 10/712 for ICM-sim, and 19/399 for FlexS. Only few exceptions occur for FXA and CDK2 with Surflex-sim and for ADA with ROCS. This is not surprising since the scoring in 3D-LBVLS methods is directly based on the molecular alignment and not on a (pseudo) binding energy evaluation as in docking-BVLS methods.

In the case of the docking-BVLS methods, all the programs but FRED retrieve more than 80% of the active compounds in the top 10% with an acceptable positioning but we can

observe that a non-negligible part of the active compounds retrieved in the top 10% display unacceptable binding modes: 66/576 for Surflex-dock, 39/268 for FRED, 36/518 for ICM, 37/467 for FlexX, and particularly in the case of THR and P38. One possible reason might be that the active compounds of these systems, where the positioning is difficult, display high flexibility and are to be positioned in relatively large binding sites which makes the conformational search more difficult for docking-BVLS methods.

This illustrates one of the surprising current paradoxes in virtual screening that is retrieving active compounds in the top scores while displaying an inaccurate molecular alignment. This paradox is much more frequent with docking-BVLS methods compared to 3D-LBVLS methods as the conformational space to explore is wider in a "real" binding pocket relatively to a pseudobinding site derived from a ligand-based hypothesis and thus more susceptible to errors. It is difficult to pinpoint a specific cause for this paradox but there might be several factors including (1) the weight of the different parameters in scoring functions which generally do not take correctly into account solvation/desolvation and entropy and (2) the rigid-body treatment of the atoms of the binding pocket or the limited handling of flexibility of the compounds during the simulation (especially for highly flexible compounds).

CONCLUSION

We have investigated the performance of 8 different 3D-VLS programs on 11 DUD systems using their corresponding DUD-own database. The use of the DUD-own database is more challenging compared to the general DUD data set and is more appropriate for benchmarking studies. The programs evaluated in this study are all able to perform acceptable molecular alignments with standard parameters. In this exercise, ICM appeared to be the most efficient underscoring the good performance of the probability biased Monte Carlo procedure used for simulating the compounds' flexibility. We identified several problems with some of the structures provided by the DUD. In particular, for P38, a non-negligible part of the ligands provided as active compounds appeared unable to bind the provided structure because of major clashes biasing negatively the results for docking-BVLS methods evaluation.

Concerning the performance of retrieving active compounds among decoys, that is, the enrichment, it was overall acceptable but variable depending on the target and the program used. Surflex-dock and ICM had the best global performance in enrichment. In terms of scaffold enrichment, 3D-LBVLS methods showed a comparable performance to docking-BVLS methods, which is quite surprising as docking-methods are generally considered to be the methods of reference for retrieving diverse molecular scaffolds. This is very interesting for the definition of new drug-discovery projects since 3D-LBVLS methods use very limited computational resources compared to the more sophisticated docking-BVLS methods (excluding FRED) thus opening new questions about which method to use when the structure of a small-molecule/protein target complex is known. The variance observed between the results is important within the DUD targets and highlights the importance of multiple systems for benchmarking studies.

Finally this study illustrates the molecular alignment paradox in enrichment that consists in finding a non-negligible number of active compounds displaying good scores, and thus good ranks but with an inaccurate positioning into the binding pocket. Considering that docking methods are in most of the cases able to produce accurate binding modes, this highlights the current limits of the scoring functions used in docking-BVLS methods that still need to be optimized to avoid such problems.

Abbreviations. 3D-LBVLS, 3D ligand-based virtual ligand screening; docking-BVLS, docking-based virtual screening; 3D-VLS, 3D virtual ligand screening; Tc, Tanimoto coefficient; ADA, adenosine deaminase; CDK2, cyclin dependent kinase 2; DHFR, dihydrofolate reductase; ER, antagonists for estrogen receptor; FXA, coagulation factor Xa; HIVRT, HIV reverse transcriptase; NA, neuraminidase; P38, mitogen-activated protein; THR, thrombin; TK, thymidine kinase; TRP, trypsin.

ACKNOWLEDGMENT

We are very grateful to Openeye Scientific Software for generously providing ROCS, FRED and OMEGA. We thank Prof Jain for generously providing the Surflex package. We would like to also thank Molsoft LLC and BiosolveIT for providing academic licenses for the ICM suite and the FlexX/FlexS package. HG is recipient of a CIFRE fellowship from ANRT.

REFERENCES AND NOTES

(1) Alvarez, J. C. High-throughput docking as a source of novel drug leads. *Curr. Opin. Chem. Biol.* **2004**, *8* (4), 365–370.
(2) Lajiness, M. S.; Johnson, M. A.; Maggiora, G. M. Implementing drug screening programs using molecular similarity methods. *Prog. Clin. Biol. Res.* **1989**, *291*, 173–176.
(3) Johnson, M.; Lajiness, M.; Maggiora, G. Molecular similarity: A basis for designing drug screening programs. *Prog. Clin. Biol. Res.* **1989**, *291*, 167–171.
(4) Guner, O. F.; Hughes, D. W.; Dumont, L. M. An integrated approach to three-dimensional information management with MACCS-3D. *J. Chem. Inf. Comput. Sci.* **1991**, *31* (3), 408–414.
(5) Good, A. C.; Hermsmeier, M. A.; Hindle, S. A. Measuring CAMD technique performance: A virtual screening case study in the design of validation experiments. *J. Comput.-Aided Mol. Des.* **2004**, *18* (7−9), 529–536.
(6) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. Scaffold-hopping by topological pharmacophore search: A contribution to virtual screening. *Angew. Chem., Int. Ed. Engl.* **1999**, *38* (19), 2894–2896.
(7) Kirchmair, J.; Ristic, S.; Eder, K.; Markt, P.; Wolber, G.; Laggner, C.; Langer, T. Fast and efficient in silico 3D screening: Toward maximum computational efficiency of pharmacophore-based and shape-based approaches. *J. Chem. Inf. Model.* **2007**, *47* (6), 2182–2196.
(8) Putta, S.; Eksterowicz, J.; Lemmen, C.; Stanton, R. A novel subshape molecular descriptor. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (5), 1623–1635.
(9) Haigh, J. A.; Pickup, B. T.; Grant, J. A.; Nicholls, A. Small molecule shape-fingerprints. *J. Chem. Inf. Model.* **2005**, *45* (3), 673–684.
(10) Kroemer, R. T. Structure-based drug design: Docking and scoring. *Curr. Protein Pept. Sci.* **2007**, *8* (4), 312–328.
(11) Dror, O.; Shulman-Peleg, A.; Nussinov, R.; Wolfson, H. J. Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. *Curr. Med. Chem.* **2004**, *11* (1), 71–90.
(12) Bowman, A. L.; Lerner, M. G.; Carlson, H. A. Protein flexibility and species specificity in structure-based drug discovery: Dihydrofolate reductase as a test system. *J. Am. Chem. Soc.* **2007**, *129* (12), 3634–40.
(13) Rush, T. S., 3rd; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein−protein interaction. *J. Med. Chem.* **2005**, *48* (5), 1489–95.
(14) Bologa, C. G.; Revankar, C. M.; Young, S. M.; Edwards, B. S.; Arterburn, J. B.; Kiselyov, A. S.; Parker, M. A.; Tkachenko, S. E.;

Savchuck, N. P.; Sklar, L. A.; Oprea, T. I.; Prossnitz, E. R. Virtual and biomolecular screening converge on a selective agonist for GPR30. *Nat. Chem. Biol.* **2006**, *2* (4), 207–212.
(15) Freitas, R. F.; Oprea, T. I.; Montanari, C. A. 2D QSAR and similarity studies on cruzain inhibitors aimed at improving selectivity over cathepsin L. *Bioorg. Med. Chem.* **2008**, *16* (2), 838–853.
(16) Bostrom, J.; Berggren, K.; Elebring, T.; Greasley, P. J.; Wilstermann, M. Scaffold hopping, synthesis, and structure−activity relationships of 5,6-diaryl-pyrazine-2-amide derivatives: a novel series of CB1 receptor antagonists. *Bioorg. Med. Chem.* **2007**, *15* (12), 4077–4084.
(17) 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* (3), 678–692.
(18) Moffat, K.; Gillet, V. J.; Whittle, M.; Bravi, G.; Leach, A. R. A comparison of field-based similarity searching methods: CatShape, FBSS, and ROCS. *J. Chem. Inf. Model.* **2008**, *48* (4), 719–729.
(19) Lemmen, C.; Lengauer, T. Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14* (3), 215–232.
(20) 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* (6), 1455–1474.
(21) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49* (23), 6789–6801.
(22) Rarey, M.; Kramer, B.; Lengauer, T. Time-efficient docking of flexible ligands into active sites of proteins. *Proc.−Int. Conf. Intell. Syst. Mol. Biol.* **1995**, *3*, 300–308.
(23) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: A method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41* (23), 4502–4520.
(24) Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8* (3), 243–256.
(25) Klebe, G. The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. *J. Mol. Biol.* **1994**, *237* (2), 212–235.
(26) Jain, A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46* (4), 499–511.
(27) Jain, A. N. Ligand-based structural hypotheses for virtual screening. *J. Med. Chem.* **2004**, *47* (4), 947–961.
(28) Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **1996**, *3* (6), 449–462.
(29) Jain, A. N. Morphological similarity: A 3D molecular similarity method correlated with protein−ligand recognition. *J. Comput.-Aided Mol. Des.* **2000**, *14* (2), 199–213.
(30) Abagyan, R.; Totrov, M.; Kusnetsov, D. ICM (internal coordinate mechanics). *J. Comput. Chem.* **1994**, *15*, 488.
(31) Schapira, M.; Totrov, M.; Abagyan, R. Prediction of the binding energy for small molecules, peptides, and proteins. *J. Mol. Recognit.* **1999**, *12* (3), 177–190.
(32) Totrov, M. Atomic property fields: Generalized 3D pharmacophoric potential for automated ligand superposition, pharmacophore elucidation, and 3D QSAR. *Chem. Biol. Drug Des.* **2008**, *71* (1), 15–27.
(33) OpenEye Scientific Software. http://www.eyesopen.com.
(34) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68* (1), 76–90.
(35) Irwin, J. J.; Shoichet, B. K. ZINC−A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45* (1), 177–182.
(36) Cleves, A. E.; Jain, A. N. Effects of inductive bias on computational evaluations of ligand-based modeling and on drug discovery. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3−4), 147–159.
(37) Hawkins, P. C.; Warren, G. L.; Skillman, A. G.; Nicholls, A. How to do an evaluation: pitfalls and traps. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3−4), 179–190.
(38) 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* (3−4), 169–178.
(39) Sing, T.; Sander, O.; Beerenwinkel, N.; Lengauer, T. ROCR: Visualizing classifier performance in R. *Bioinformatics* **2005**, *21* (20), 3940–3941.
(40) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.
(41) Fischer, B.; Fukuzawa, K.; Wenzel, W. Receptor-specific scoring functions derived from quantum chemical models improve affinity estimates for in-silico drug discovery. *Proteins* **2008**, *70* (4), 1264–1273.
(42) Perez-Nueno, V. I.; Ritchie, D. W.; Rabal, O.; Pascual, R.; Borrell, J. I.; Teixido, J. Comparison of ligand-based and receptor-based virtual screening of HIV entry inhibitors for the CXCR4 and CCR5 receptors using 3D ligand shape matching and ligand-receptor docking. *J. Chem. Inf. Model.* **2008**, *48* (3), 509–533.
(43) 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* (20), 5912–5931.
(44) Abagyan, R.; Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **1994**, *235* (3), 983–1002.
(45) von Korff, M.; Freyss, J.; Sander, T. Comparison of ligand- and structure-based virtual screening on the DUD data set. *J. Chem. Inf. Model.* **2009**, *49* (2), 209–231.

CI900507G