JOURNAL OF CHEMICAL INFORMATION AND MODELING

LETTER

pubs.acs.org/jcim

Virtual Decoy Sets for Molecular Docking Benchmarks

Izhar Wallach*,†,‡ and Ryan Lilien*,†,‡,§

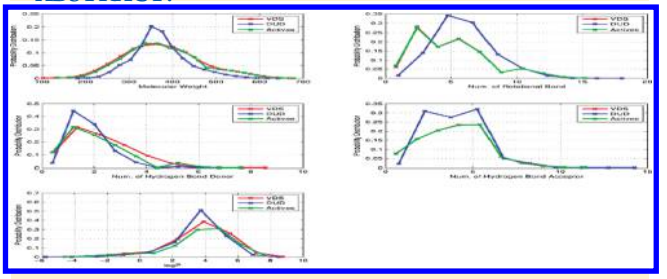
†Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

‡Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada

§Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada

Supporting Information

ABSTRACT:



Virtual docking algorithms are often evaluated on their ability to separate active ligands from decoy molecules. The current state-of-the-art benchmark, the Directory of Useful Decoys (DUD), minimizes bias by including decoys from a library of synthetically feasible molecules that are physically similar yet chemically dissimilar to the active ligands. We show that by ignoring synthetic feasibility, we can compile a benchmark that is comparable to the DUD and less biased with respect to physical similarity.

Drug discovery pipelines typically utilize the virtual docking of compound libraries as a mechanism for identifying active molecules. Scoring functions used by these algorithms approximate the underlying physical forces governing receptor−ligand interactions, and accurate predictions of binding affinities continue to be a challenge.[1] It is common to evaluate docking algorithms by their ability to enrich a set of top scoring solutions with active ligands. For a screening library, consisting of a small set of active ligands and a large set of nonbinding decoys, the enrichment factor of the top k% ranked ligands is the ratio between the frequency of active ligands within the top k% and the frequency throughout the entire library. Because the enrichment factor is based on the relative ranking of actives and decoys, great attention should be given to the design of the decoy sets. Desirable decoy sets should only permit virtual docking algorithms to separate actives from decoys on the basis of predicted chemical interaction with the receptor. Huang et al.[2] showed that decoys having different physical properties than those of their corresponding active ligands may yield biased enrichments. To provide a nonbiased benchmark for virtual screening algorithms, they published the Directory of Useful Decoys (DUD)—a comprehensive compilation of decoy sets for a diverse collection of protein targets and active ligands, which preserves physical similarity.

The DUD is currently the gold standard benchmark for virtual docking algorithms; however, it suffers from several limitations.[3] First, all DUD decoys are selected from the ZINC database[4], and thus, they only span a small synthetically feasible subset of small-molecule space. Second, physical similarity between active ligands and their corresponding decoys may not always be achieved. This is because each DUD decoy set consists of the 36 ZINC compounds that are most physically similar yet chemically different from the active ligand (and from each other). As a result, the quality of the decoy sets may drop for active ligands whose physical properties are poorly represented in the ZINC database. Third, in the event that a set of 36 decoys may be too small for enrichment studies, it may not be possible to identify a larger high quality decoy set. Finally, with only a single decoy set, the community runs the risk of inadvertently tuning algorithms and scoring function to perform well on the single benchmark. To reduce the risk of overfitting docking algorithms, it is beneficial to be able to generate multiple decoy sets of similar quality. However, under the current DUD methodology, generating additional decoy sets will likely result in sets containing decoys that are not as physically similar to the active ligands as are the decoys in the first set.

In this work, we rationalize the use of in silico generated decoys in virtual docking benchmarks. We argue that decoys do not have to be synthetically feasible to be useful. In fact, the restriction to synthetically feasible compounds can yield biased benchmark sets. The advantages of in silico decoy generation are the following: (i) decoys may be generated for any active ligand, (ii) the physical features of the decoys can be guaranteed for any set size, (iii) decoys may be generated on-the-fly, and (iv) the risk of overfitting may be reduced by generating and using multiple decoy sets of similar quality. In this work, we present a virtual decoy set (VDS) of molecules that are chemically correct but are not necessarily synthetically feasible. We demonstrate the utilization of the VDS as a molecular docking benchmark and compare it to the DUD data set. Our results suggest that the VDS is comparable to the DUD set and, in several aspects, is a more reliable benchmark.

We generated decoy sets using the same five physical descriptors used to generate the DUD data set.[2] Given a reference active molecule, we calculated the following five physical descriptors using the OpenBabel[5] chemical toolbox: (i) molecular weight, (ii) number of rotational bonds, (iii) number of hydrogen bond donors (HBD), (iv) number of

Published: January 05, 2011

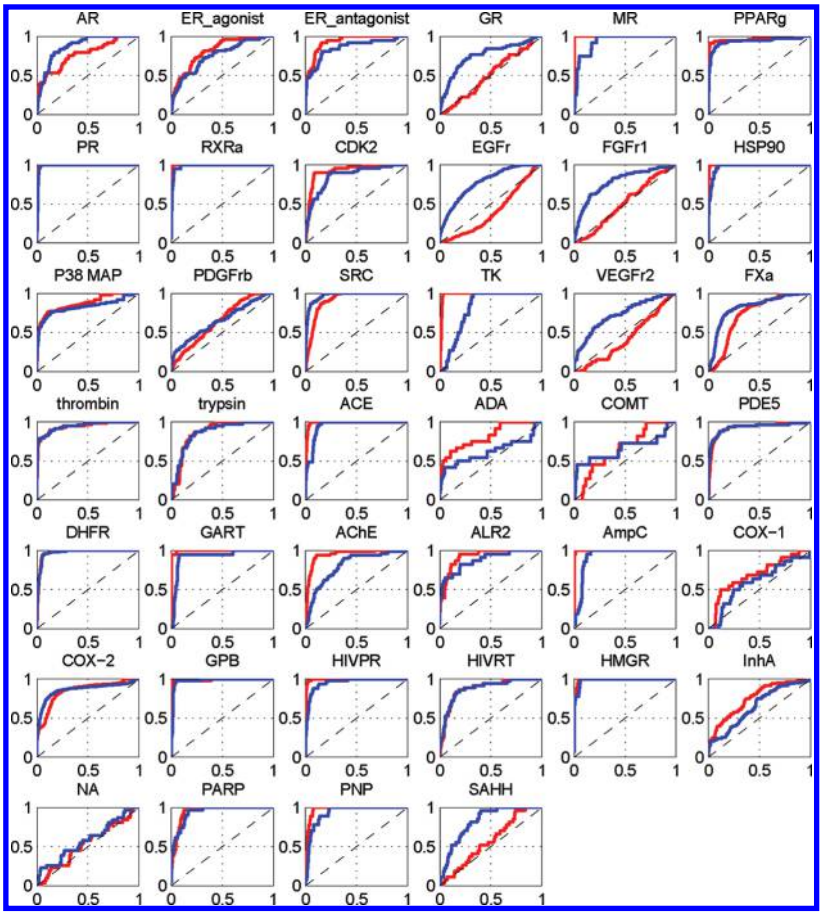


Figure 1. Comparison of the enrichments attained using the VDS and DUD decoy sets over the 40 DUD protein targets. Enrichment is measured using the receiver operator characteristic (ROC) curves. Each curve shows the fraction of selected decoys (*x*-axis) versus the fraction of selected ligands (*y*-axis) ranked by their docking score. A decoy set that shows worse enrichment than that of the other (smaller area under the curve) is generally considered better for benchmarking. The VDS and DUD enrichments are illustrated with red and blue curves, respectively.

hydrogen bond acceptors (HBA), and (v) the octanol−water partition coefficient (CLogP). We note that although it was previously suggested that this set of physical descriptors could be further extended,[3] we elected not to extend the descriptor set in order to fairly benchmark the VDS and the DUD sets. Because the aim of this study is to establish the admissibility of using virtual decoys, we do not focus here on details of compound generation. Similar to the methods utilized in de novo ligand design,[6−8] we implemented an algorithm to generate ligands that satisfy physical constraints. Our ligand generator uses a library of chemical building blocks and bridges to iteratively generate chemically plausible molecules (see Supporting Information for a description of the algorithm). It is important to emphasize that although we used our in-house application for generating molecules, one may use any generator and filtering method that identifies molecules with the desired physical properties. Similar to the DUD, for every active ligand we generated a set of 36 decoys. In our algorithm, a new molecule is added into the decoy set if the molecule's five physical descriptors are similar to the reference molecule and its chemical similarity to the reference is less than 0.9 (measured by the Tanimoto coefficients using the MACCS fingerprints). To be considered similar, the physical descriptors of the decoy must satisfy the following constraints with respect to the active ligand: (i) molecular weight of ±40 Da, (ii) exact same number of rotational bonds, HBDs, and HBAs, and (iii) CLogP of ±1.0. In a few cases where full decoy sets could not be generated within several hours, we relaxed the constraints of the rotational bonds, HBDs, and HBAs to be within ±1 and the CLogP to be within ±1.5. The VDS is publicly available at http://compbio.cs.toronto.edu/VDS.

We benchmarked the VDS and DUD using similar methodology to Huang et al. We built two screening sets for the DUD protein targets. Each set included either the DUD or the VDS decoys, and both shared the same collection of active ligands (from the DUD data set). We screened the two sets using the eHiTS docking algorithm[9] and calculated enrichments for every protein target. We used receiver operator characteristic (ROC) curves to plot the fraction of selected decoys versus the fraction of selected ligands (Figure 1). Assuming similar physical properties, a decoy set that shows worse enrichment than that of the other is generally considered better for benchmarking. For eight systems, the area under the curve (AUC) of the VDS sets was more than 10% smaller than the AUC of the DUD sets. For five systems, the AUC of the DUD sets was more than 10% smaller than the VDS sets, and for the rest of the targets, both decoy sets were comparable (Table 1). Following Huang et al.,[2] for every target, we computed the enrichment factors at 1% of the data set (EF_1), 20% of the data set (EF_{20}), and the maximal enrichment over the whole data set (EF_{max}). The enrichment results are summarized in Table 1. The average enrichments for the VDS and the DUD data sets using the EF_{max}, EF_1, and EF_{20} cutoffs

dx.doi.org/10.1021/ci100374f |*J. Chem. Inf. Model.* 2011, 51, 196–202

Table 1. Comparison of Different Enrichments Attained Using VDS and DUD Decoy Sets over the 40 DUD Protein Targets[a]

	AUC		EF_{max}		EF_1		EF_{20}	
Target	VDS	DUD	VDS	DUD	VDS	DUD	VDS	DUD
AR	0.78	0.88	36.0	40.5	20.0	8.1	2.6	3.9
ER_agonist	0.82	0.75	36.0	18.8	12.6	9.4	2.9	2.6
ER_antagonist	0.94	0.86	36.0	29.4	27.7	21.4	4.5	3.9
GR	0.51	0.76	1.1	11.3	0.0	7.7	1.0	3.0
MR	1.00	0.93	36.0	13.8	36.0	7.4	4.9	4.5
PPARg	0.96	0.94	36.0	26.1	34.8	19.5	4.6	4.5
PR	0.99	0.99	27.0	3.2	21.6	18.1	4.9	4.8
RXRa	1.00	0.99	36.0	24.7	30.9	24.7	4.9	4.8
CDK2	0.93	0.86	14.4	30.8	9.5	17.3	4.4	3.5
EGFr	0.41	0.78	1.0	33.3	0.5	12.0	0.5	2.8
FGFr1	0.51	0.78	1.1	13.9	0.0	7.2	0.6	3.0
HSP90	1.00	0.98	30.0	29.2	22.5	29.2	4.9	4.9
P38 MAP	0.87	0.83	24.0	31.2	20.8	20.8	3.8	3.8
PDGFrb	0.65	0.65	7.2	31.1	3.3	13.6	1.6	2.0
SRC	0.91	0.97	11.0	33.7	10.2	27.0	4.3	4.8
TK	0.99	0.81	19.6	2.8	13.5	0.0	4.9	2.4
VEGFr2	0.43	0.73	1.0	30.7	0.0	12.0	0.6	2.7
FXa	0.73	0.82	2.3	7.9	0.0	3.0	2.0	3.6
thrombin	0.94	0.95	36.0	34.2	27.4	29.1	4.4	4.4
trypsin	0.87	0.86	5.2	22.3	2.6	12.9	4.0	3.7
ACE	0.99	0.95	36.0	32.7	36.0	25.2	4.9	4.9
ADA	0.82	0.62	22.9	22.2	20.0	12.7	3.2	2.4
COMT	0.65	0.66	2.3	11.7	0.0	0.0	2.2	2.7
PDE5	0.93	0.93	15.7	31.0	13.1	26.1	4.4	4.4
DHFR	0.97	0.97	36.0	35.8	19.2	27.3	4.8	4.8
GART	1.00	0.93	36.0	9.4	36.0	0.0	4.9	4.6
AChE	0.95	0.81	18.0	30.0	8.8	4.8	4.6	3.0
ALR2	0.93	0.88	36.0	37.2	20.0	33.0	4.2	3.4
AmpC	1.00	0.93	36.0	34.7	31.5	4.3	4.9	4.9
COX-1	0.69	0.60	3.7	2.0	0.0	0.0	2.4	1.5
COX-2	0.85	0.87	36.1	30.8	23.3	22.7	3.8	4.0
GPB	0.99	0.98	32.9	36.8	30.9	22.5	4.7	4.8
HIVPR	0.99	0.96	36.0	31.0	30.0	25.2	4.9	4.5
HIVRT	0.88	0.88	18.0	34.4	11.1	15.9	4.1	4.0
HMGR	0.99	0.99	36.0	55.6	36.0	44.5	4.9	4.9
InhA	0.75	0.65	35.6	31.4	9.5	11.6	2.5	1.5
NA	0.53	0.59	1.6	5.7	0.0	3.1	1.1	1.3
PARP	0.95	0.94	36.0	35.1	19.6	31.9	4.7	4.7
PNP	0.98	0.93	36.0	14.8	21.6	7.4	4.9	4.3
SAHH	0.57	0.82	1.5	11.0	0.0	7.0	1.1	2.7
Mean	0.84	0.85	22.7	25.6	16.5	15.6	3.6	3.7
STD	0.18	0.12	14.5	11.8	12.6	10.9	1.5	1.1

[a] AUC denotes the area under the ROC curve. Larger AUC values indicate better enrichment. In our case, data sets that produce smaller AUC values are generally considered better for benchmarking. EF_1, EF_{20}, and EF_{max} correspond to the enrichment factors at 1% of the decoys, 20% of the decoys, and the maximal enrichment over the whole set of decoys, respectively.

were 22.7 and 25.6, 16.5 and 15.6, and 3.6 and 3.7, respectively. These results suggest comparable performance between the two sets.

While the enrichment experiments show that the VDS is comparable to the DUD, an analysis of the differences between the two sets suggests that the VDS methodology may provide better control of the decoy properties. We compared the physical properties of the VDS and DUD with respect to the properties of the active ligands (Figure 2). For the VDS data set, we were able to generate decoys with physical properties that were highly similar to those of the active ligands. The DUD data set, however, was selected from a limited chemical space, and thus, the distributions of physical properties may significantly differ (Figure 2). Particularly, the distributions of the number of rotational bonds, the number of HBDs, and the molecular weight for the DUD decoys substantially deviate from the distributions of the active ligands. Furthermore, the physical similarity for individual targets may vary significantly (Table 1 of the Supporting Information).

In order to demonstrate the limitation of selecting decoys from a synthetically feasible library, we tested the ability to generate a DUD-like data set using the same physical constraints applied in the VDS. We used a collection of more than 20 million compounds from the ZINC database to select decoys satisfying the physical constraints (see the Self-Generated DUD paragraph in the Experimental Procedures section). In our results, for approximately 12% of the active ligands, it was not possible to generate a set with more than 10 decoys, and for more than 30% of the active ligands, we could not obtain sets with over 100 decoys (Figure 1 of the Supporting Information). Furthermore, for some targets such as SAHH and GART the generated sets for all active ligands were smaller than 100 decoys (Figure 2 of the Supporting Information). These results demonstrate the limitation of deriving decoys from a synthetically accessible domain.

Perhaps most important for generating a nonbiased decoy set is the minimization of any differences in physical properties between the decoys and the active ligands. If decoys and active ligands have different physical properties, the docking scores may be influenced by the discrepancy, and the enrichment may get arbitrarily better or worse. As demonstrated in the previous paragraph, the physical properties of the DUD decoys may substantially deviate from the properties of their corresponding active ligands (Figure 2). For example, the DUD decoys for the FXa target have, on average, 2.3 HBAs more than their corresponding active ligands (Table 1 of the Supporting Information). To test the effect such HBA deviations have on the enrichment, we generated four VDS sets for FXa that differed by their number of HBAs. In each set, decoys were generated with a number of HBAs that was offset by −2, −1, +1, or +2 HBA from the active ligands. Computing enrichments for these four decoy sets showed substantial differences (Figure 3). These results demonstrate the importance of having physical similarity between the active ligands and decoys and may explain some of the individual enrichment differences between the VDS and DUD. We demonstrated similar phenomena for other targets and physical properties (Figure 3 of the Supporting Information).

To further illustrate the robustness of the VDS approach and to demonstrate that the advantages of the VDS are independent of the algorithm used to generate the virtual decoys, we conducted two additional docking experiments using a different decoy generator. In the following, we used Tripos' EA-Inventor (http://tripos.com) to generate sets of decoys.

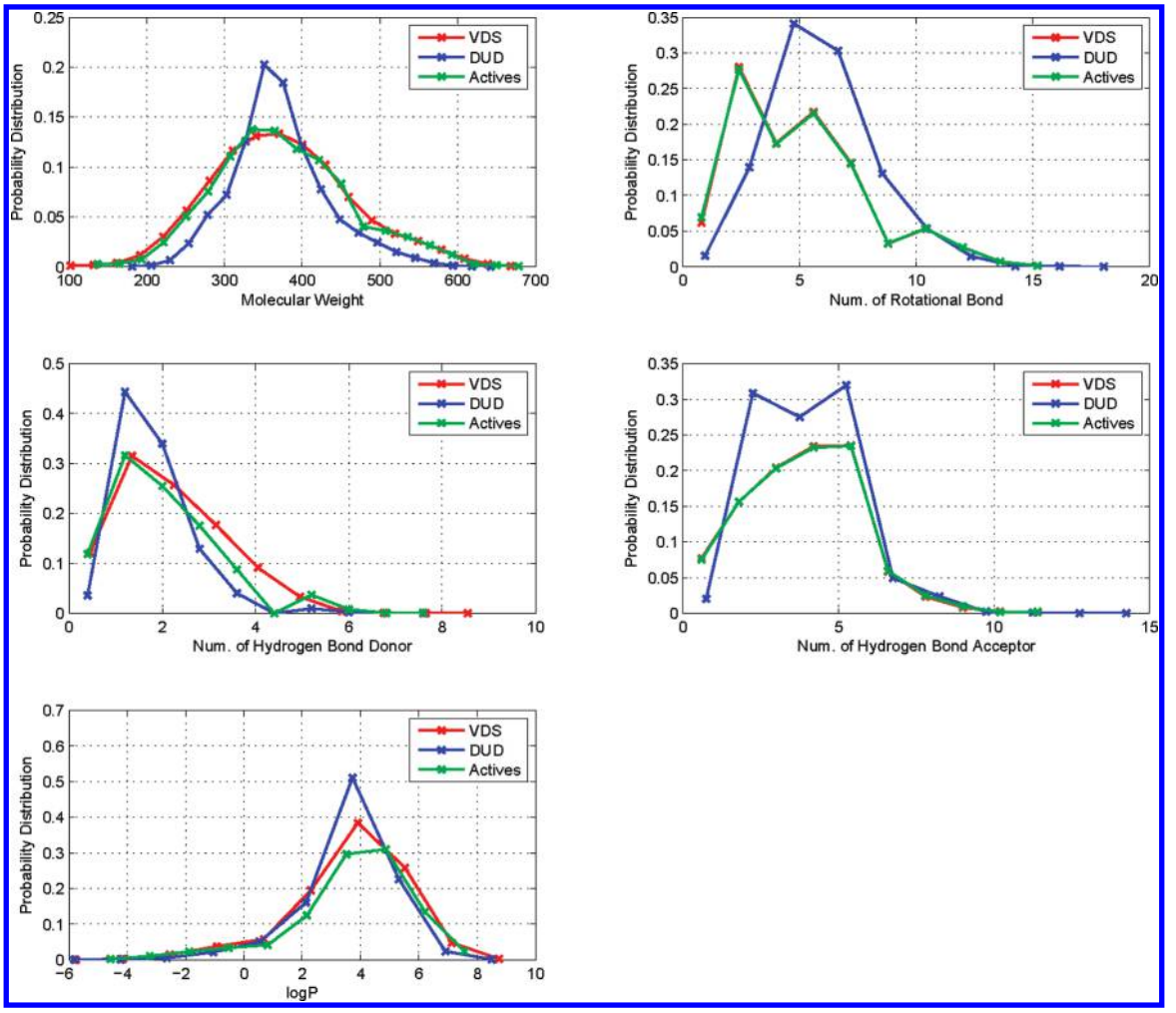


Figure 2. Comparison between the VDS and DUD data sets over five physical descriptors used as indicators of physical similarity between active ligands and decoys (summed across all the targets). The green line illustrates the distribution of the active ligands. The red and blue lines illustrate the distributions of the VDS and DUD, respectively.

The EA-Inventor uses a genetic algorithm to generate a population of chemically correct molecules (although not necessarily synthetically feasible) that minimize some fitness function. In this case, we used a fitness function with seven physical property terms. Following the suggestions of Irwin,[3] we added the formal charge and topological polar surface area (TPSA) descriptors to the set of five physical properties used previously. As a data set, we used 13 DUD targets that have more than 15 different active ligands based on Andrew Good's DUD clustering (http://dud.docking.org/clusters). In this set, active ligands were clustered on the basis of the reduced graph assembly algorithm[10], and the smallest ligand from each cluster was selected. To reduce computation effort, we did not use the whole set of DUD decoys but instead randomly sampled a reduced subset, while ensuring that the decoy set for each target was 10 times larger than the size of the active set. Jain[11] demonstrated that such a sampling is sufficient for ranking studies. We docked the data sets using the Glide[12] docking algorithm in addition to the eHiTS algorithm used earlier. The ranking results for both docking algorithms are illustrated in Figures 4 and 5. Because the size of the data sets are small compared to the whole DUD set used above, we measured early enrichment at 3% of the ranked decoys in addition to the late enrichment measured at 20% of the ranked decoys. The enrichment and the AUC results for the 13 targets are summarized in Tables 2 and 3 for eHiTS and Glide, respectively. Generally, the results are in agreement with the previous results obtained with the full DUD data set. For both docking algorithms, the average early enrichment of the VDS is smaller than that of the the DUD, and the late enrichment of the VDS is marginally better (smaller) using eHiTS and marginally worse using Glide. We note that enrichment values should not be directly compared with the full DUD results in Table 1 because the ratios between actives and decoys differ between the sets.

In this work, we rationalize the use of in silico generated decoys in virtual docking benchmarks. We claim that decoys do not have to be synthetically feasible and, hence, may be derived from the entirety of chemical space. The lack of a synthetic feasibility constraint allows us to generate virtual decoy sets that more closely match the physical properties of the active ligands. We first showed that the VDS is comparable to the current gold standard benchmark for virtual screening algorithms. We then demonstrated the advantages of the VDS over the DUD in providing a less biased benchmark and in providing decoy sets with an arbitrary number of molecules. Finally, we note that the

VDS methodology allows one to generate multiple benchmarks of equal quality and, thus, may reduce the likelihood of overfitting docking algorithms to a single benchmark. In silico decoy generation is not only useful for benchmarking docking algorithms, but it may also open the door for dynamic calibration of virtual docking parameters and the ranking of docking results. Previous work demonstrated the use of decoys to calibrate docking parameters and normalize docking scores.[13−15] Virtual docking applications that can easily generate a "calibration decoy set" for every input ligand may effectively utilize decoys for virtual screening applications rather than just for benchmarking.

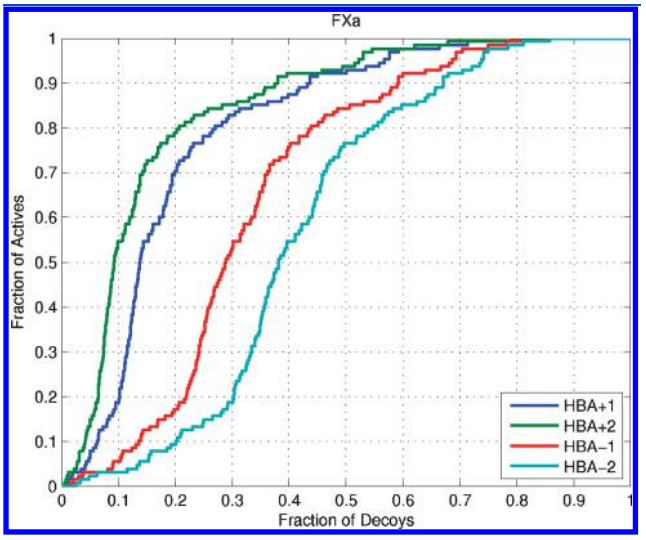


Figure 3. Example of arbitrary enrichment variations due to physical property differences between the active ligands and the decoys. We generated four decoy sets for the FXa target that differed by their number of HBA. In each set, decoys were generated with a number of HBA that was offset by −2, −1, +1, or +2 HBA from the active ligands. The enrichment curves show that if the physical properties of the decoys differ from the active ligand ones, it may result with an arbitrary increase or decrease in the enrichment.

EXPERIMENTAL PROCEDURES

Docking Procedure. The DUD data set was obtained from http://dud.docking.org (release 2). The DUD data set contains sets of active ligands and decoys for 40 protein targets. All targets structures were used without any modification. We removed redundant molecules from the active and decoy sets (Tanimoto coefficients of 1.0 using the Daylight fingerprints[16]). For every active ligand, a virtual decoy set of size 36 was generated using the physical constraints described above. All VDS decoys were minimized using the Ghemical force field[17], and their protonation state was fixed for pH 7.0 using an in-house software built with the OpenBabel chemical toolbox.[5] Docking was done using the eHiTS docking algorithm[9] (version 2009.0) with default parameters and a clipping box around the binding site (identified by the native binding ligand) of 10 Å. In cases where no docking solution was found, we assigned the ligand a score of zero. For docking the Andrew Good's DUD subset, we used eHiTS version 2009.1 and Glide version 201007. Docking was performed with the fast mode for eHiTS and SP mode for Glide. For Glide, receptors were prepared using prepwizard, and ligands were prepared using ligprep.

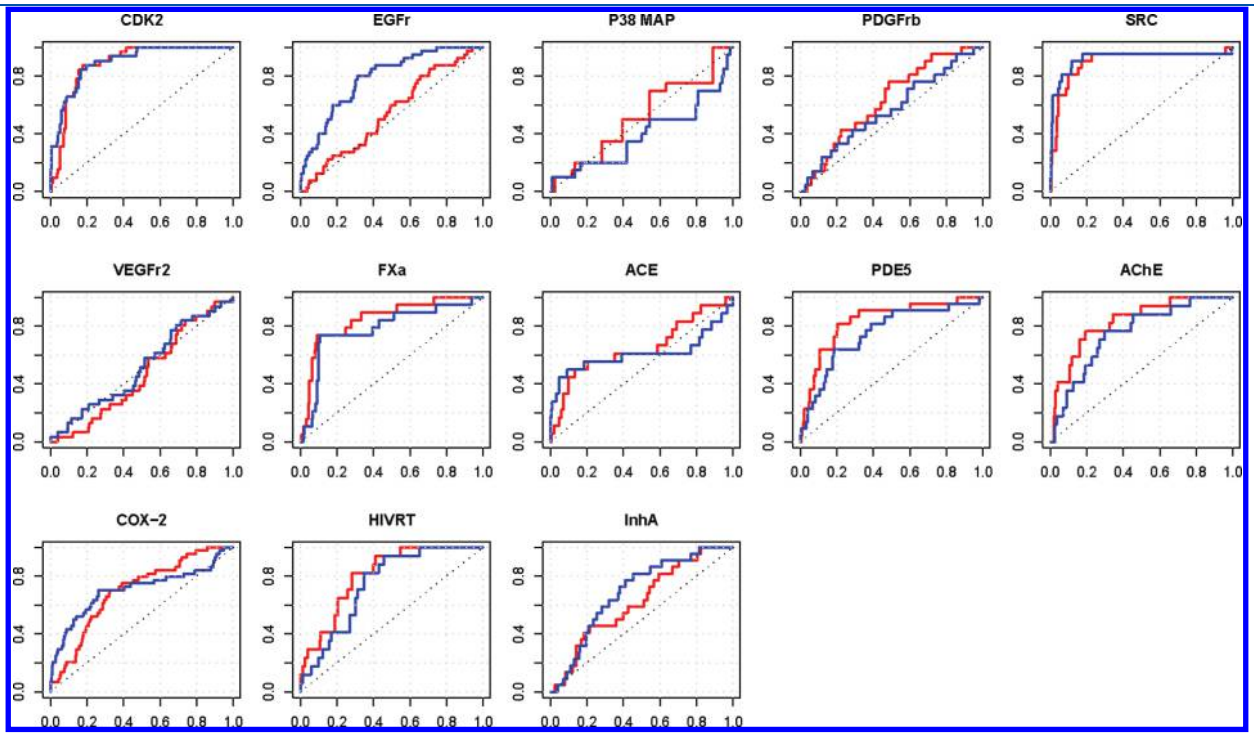


Figure 4. Using the eHiTS docking algorithm. A comparison of the enrichments attained using the VDS and DUD decoy sets over the 13 targets from Andrew Good's DUD clustering. Enrichment is measured using the receiver operator characteristic (ROC) curves. Each curve shows the fraction of selected decoys (*x*-axis) versus the fraction of selected ligands (*y*-axis) ranked by their docking score. A decoy set that shows worse enrichment than that of the other (smaller area under the curve) is generally considered better for benchmarking. The VDS and DUD enrichments are illustrated with red and blue curves, respectively.

dx.doi.org/10.1021/ci100374f |*J. Chem. Inf. Model.* 2011, 51, 196–202

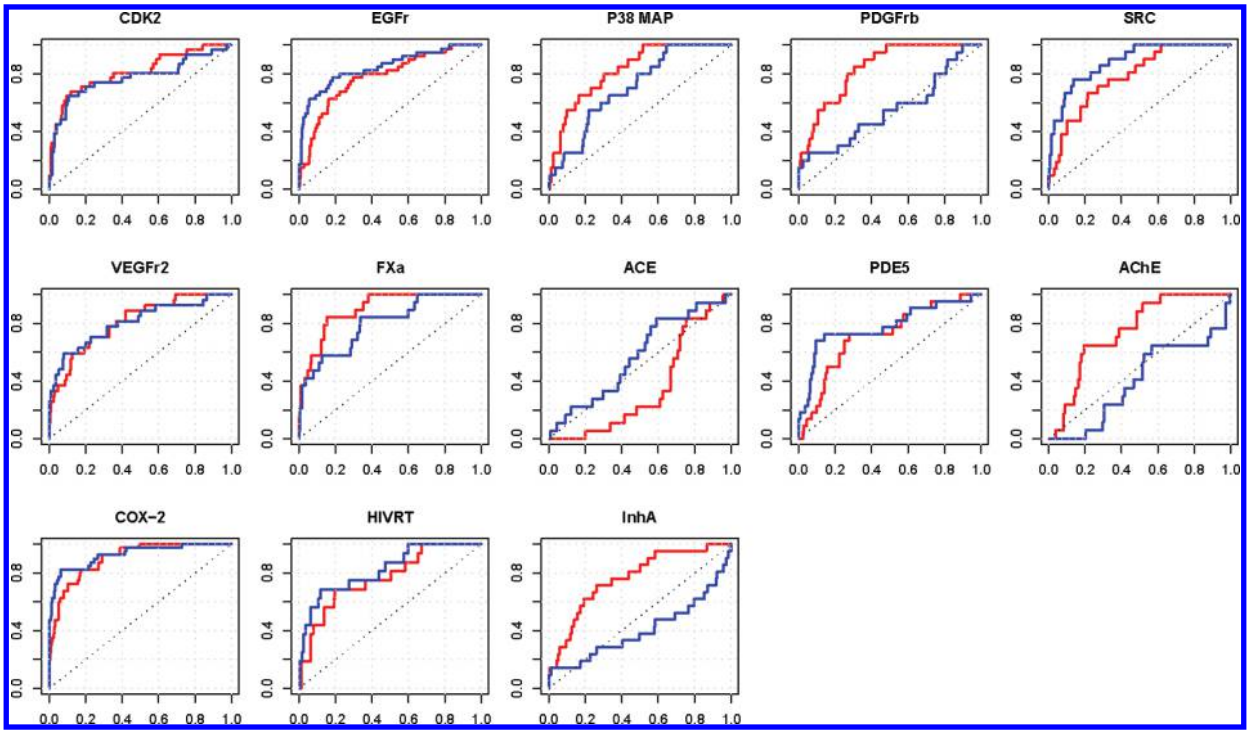


Figure 5. Using the Glide docking algorithm. A comparison of the enrichments attained using the VDS and DUD decoy sets over the 13 targets from Andrew Good's DUD clustering. Enrichment is measured using the receiver operator characteristic (ROC) curves. Each curve shows the fraction of selected decoys (*x*-axis) versus the fraction of selected ligands (*y*-axis) ranked by their docking score. A decoy set that shows worse enrichment than that of the other (smaller area under the curve) is generally considered better for benchmarking. The VDS and DUD enrichments are illustrated with red and blue curves, respectively.

Table 2. Using the eHiTS Docking Algorithm[a]

	AUC		EF_3		EF_{20}	
Target	VDS	DUD	VDS	DUD	VDS	DUD
CDK2	0.89	0.90	2.36	4.19	2.84	2.84
EGFr	0.55	0.79	0.00	4.50	1.20	2.49
P38 MAP	0.52	0.39	2.75	2.75	0.98	0.98
PDGFrb	0.63	0.57	0.00	1.43	1.54	1.54
SRC	0.90	0.92	5.87	7.33	3.37	3.49
VEGFr2	0.47	0.52	0.00	1.00	0.34	1.10
FXa	0.85	0.78	3.67	2.75	2.96	2.96
ACE	0.65	0.61	2.96	5.92	2.22	2.41
PDE5	0.84	0.75	4.40	3.23	3.06	2.56
AChE	0.83	0.75	5.92	3.85	3.09	2.31
COX-2	0.70	0.71	1.81	4.84	2.00	2.39
HIVRT	0.81	0.75	4.89	3.14	2.30	1.88
InhA	0.63	0.69	1.29	0.00	1.82	1.82
Mean	0.71	0.70	2.76	3.46	2.13	2.21
STD	0.15	0.15	2.12	2.00	0.93	0.73

[a] A comparison of different enrichments attained using the VDS and the DUD decoy sets over the 13 targets from Andrew Good's DUD clustering. AUC denotes the area under the ROC curve. EF_3 and EF_{20} correspond to the early enrichment at 3% of the decoys and late enrichment at 20% of the decoys.

Table 3. Using the Glide Docking Algorithm[a]

	AUC		EF_3		EF_{20}	
Target	VDS	DUD	VDS	DUD	VDS	DUD
CDK2	0.81	0.77	5.84	5.84	2.86	2.86
EGFr	0.78	0.85	3.83	6.64	2.58	3.03
P38MAP	0.82	0.71	5.00	2.75	2.65	1.80
PDGFrb	0.84	0.55	4.85	4.24	2.54	1.21
SRC	0.78	0.88	3.49	6.55	2.42	2.99
VEGFr2	0.81	0.80	5.30	5.58	2.57	2.69
FXa	0.90	0.80	6.29	5.92	3.26	2.47
ACE	0.35	0.57	0.00	1.69	0.30	1.10
PDE5	0.72	0.78	2.43	3.98	2.19	2.92
AChE	0.75	0.40	0.00	0.00	2.71	0.31
COX-2	0.90	0.92	5.34	6.01	2.92	2.92
HIVRT	0.77	0.82	4.29	6.24	2.80	2.80
InhA	0.76	0.41	3.26	3.26	2.59	0.94
Mean	0.77	0.71	3.84	4.52	2.49	2.16
STD	0.14	0.17	2.02	2.08	0.71	0.95

[a] A comparison of different enrichments attained using the VDS and DUD decoy sets over the 13 targets from Andrew Good's DUD clustering. AUC denotes the area under the ROC curve. EF_3 and EF_{20} correspond to the early enrichment at 3% of the decoys and late enrichment at 20% of the decoys.

Self-Generated DUD. A set of over 20 million molecules was downloaded form the ZINC[4] database (http://zinc.docking.org, set #10, 2009-07-03). We randomly shuffled the order of the molecules. Then, we computed the MACCS fingerprints for every molecule and used the SUBSET algorithm[18] to select a nonredundant subset of approximately 1.4 million molecules

(using Tanimoto coefficient threshold of 0.9). For every active ligand, a new decoy (from the nonredundant set) was added to the active ligand's decoy set if the decoy's physical properties were within the physical constraints mentioned above and the similarity between the new decoy and the active ligand was below 0.9 (using the Tanimoto coefficient of the MACCS fingerprints).

■ ASSOCIATED CONTENT

Supporting Information. Description of the decoy generation algorithm, supplementary figures and tables for the "self-generated DUD" experiment, and additional figures demonstrating enrichment variations for decoy sets with varied physical properties. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: izharw@cs.toronto.edu (I.W); lilien@cs.toronto.edu (R.L.).

■ ACKNOWLEDGMENT

We thank members of the Lilien lab for helpful discussions and comments on drafts. We thank Simulated Biomolecular Systems (Toronto, Ontario) for providing access to their eHiTS High Throughput Screening software. R.H.L. is supported by a Discovery grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

■ REFERENCES

(1) Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins?. *J. Chem. Inf. Model.* **2009**, *49*, 1568–1580.

(2) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.

(3) Irwin, J. J. Community benchmarks for virtual screening. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 193–199.

(4) Irwin, J. J.; Shoichet, B. K. ZINC: A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.

(5) Guha, R.; Howard, M.; Hutchison, G.; Murray-Rust, P.; Rzepa, H.; Steinbeck, C.; Wegner, J.; Willighagen, E. The Blue Obelisk-interoperability in chemical informatics. *J. Chem. Inf. Model.* **2006**, *46*, 991–998.

(6) Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discovery* **2005**, *4*, 649–663.

(7) Vinkers, H. M.; de Jonge, M. R.; Daeyaert, F. F.; Heeres, J.; Koymans, L. M.; van Lenthe, J. H.; Lewi, P. J.; Timmerman, H.; Van Aken, K.; Janssen, P. A. SYNOPSIS: SYN-thesize and OPtimize System in Silico. *J. Med. Chem.* **2003**, *46*, 2765–2773.

(8) Ekins, S.; Honeycutt, J. D.; Metz, J. T. Evolving molecules using multi-objective optimization: Applying to ADME/Tox. *Drug Discovery Today* **2010**, *15*, 451–460.

(9) Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S. B.; Johnson, A. P. eHiTS: A new fast, exhaustive flexible ligand docking system. *J. Mol. Graph. Model.* **2007**, *26*, 198–212.

(10) Barker, E. J.; Gardiner, E. J.; Gillet, V. J.; Kitts, P.; Morris, J. Further development of reduced graphs for identifying bioactive compounds. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 346–356.

(11) Jain, A. N. Bias, reporting, and sharing: Computational evaluations of docking methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 201–212.

(12) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.

(13) Irwin, J. J.; Shoichet, B. K.; Mysinger, M. M.; Huang, N.; Colizzi, F.; Wassam, P.; Cao, Y. Automated docking screens: A feasibility study. *J. Med. Chem.* **2009**, *52*, 5712–5720.

(14) Xie, L.; Li, J.; Xie, L.; Bourne, P. E. Drug discovery using chemical systems biology: Identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. *PLoS. Comput. Biol.* **2009**, 5.

(15) Vigers, G. P.; Rizzi, J. P. Multiple active site corrections for docking and virtual screening. *J. Med. Chem.* **2004**, *47*, 80–89.

(16) James, A. C.; Weininger, D.; Delany, J. *Daylight Theory Manual, Daylight 4.71*; Daylight Chemical Information Systems: Laguna Niguel, CA, 2000.

(17) Hassinen, T.; Peräkylä, M. New energy terms for reduced protein models implemented in an off-lattice force field. *J. Comput. Chem.* **2001**, *22*, 1229–1242.

(18) Voigt, J. H.; Bienfait, B.; Wang, S.; Nicklaus, M. C. Comparison of the NCI open database with seven large chemical structural databases. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 702–712.