J Comput Aided Mol Des (2012) 26:787–799
DOI 10.1007/s10822-012-9575-9

Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide

Matthew P. Repasky · Robert B. Murphy · Jay L. Banks · Jeremy R. Greenwood · Ivan Tubert-Brohman · Sathesh Bhat · Richard A. Friesner

Received: 28 February 2012 / Accepted: 24 April 2012 / Published online: 11 May 2012

Abstract Glide SP mode enrichment results for two preparations of the DUD dataset and native ligand docking RMSDs for two preparations of the Astex dataset are presented. Following a best-practices preparation scheme, an average RMSD of 1.140 Å for native ligand docking with Glide SP is computed. Following the same best-practices preparation scheme for the DUD dataset an average area under the ROC curve (AUC) of 0.80 and average early enrichment via the ROC (0.1 %) metric of 0.12 were observed. 74 and 56 % of the 39 best-practices prepared targets showed AUC over 0.7 and 0.8, respectively. Average AUC was greater than 0.7 for all best-practices protein families demonstrating consistent enrichment performance across a broad range of proteins and ligand chemotypes. In both Astex and DUD datasets, docking performance is significantly improved employing a best-practices preparation scheme over using minimally-prepared structures from the PDB. Enrichment results for WScore, a new scoring function and sampling methodology integrating WaterMap and Glide, are presented for four DUD targets, hivrt, hsp90, cdk2, and fxa. WScore performance in early enrichment is consistently strong and all systems examined show AUC > 0.9 and superior early enrichment to DUD best-practices Glide SP results.

Keywords Docking · Virtual screening · DUD · Astex · WScore · WaterMap · Glide

M. P. Repasky (✉) · R. B. Murphy
Schrödinger, LLC, 101 SW Main St., Portland, OR 97204, USA
e-mail: matt.repasky@schrodinger.com

R. B. Murphy
e-mail: rob.murphy@schrodinger.com

J. L. Banks · J. R. Greenwood · I. Tubert-Brohman · S. Bhat
Schrödinger, LLC, 120 West 45th St., New York, NY 10036, USA
e-mail: jay.banks@schrodinger.com

J. R. Greenwood
e-mail: jeremy.greenwood@schrodinger.com

I. Tubert-Brohman
e-mail: ivan.tubert-brohman@schrodinger.com

S. Bhat
e-mail: satesh.bhat@schrodinger.com

R. A. Friesner
Department of Chemistry, Columbia University, New York, NY 10036, USA
e-mail: rich@chem.columbia.edu

Introduction

Virtual screening in principle represents an efficient approach to lead discovery in structure-based drug discovery projects. By docking a large library of compounds into one or more high resolution structures of the target receptor, potentially active compounds can be identified and tested experimentally. Docking methods are also used in lead optimization for predicting structures of protein–ligand complexes and for estimating the binding affinity of these complexes. Both applications require sufficiently accurate binding mode and binding affinity prediction. The present paper is focused on an assessment of these calculations, albeit in particular applications, via the Glide docking program [1, 2].

We evaluate Glide SP (standard precision) for binding mode prediction using the Astex [3] set of self-docking tests, and enrichment in a virtual screening context using the DUD dataset [4]. The datasets, evaluation metrics, and experiments performed were prescribed by conditions of the Docking and Scoring Symposium at the 241st ACS

Meeting in Anaheim. Glide SP is quite successful for a high fraction of the test cases in this exercise, but enrichment calculations for a subset of DUD targets display significant underperformance. We select four of the more difficult cases for Glide and employ ensemble docking combined with a new scoring function that incorporates information about water sites in the binding pocket from a WaterMap [5] simulation. This combination of methods shows significant improvement in enrichments in all four cases examined.

The paper is organized as follows. We do not discuss the Glide SP methodology, as this has been presented in previous work [1, 2]. In the "Glide SP Calculations" section we focus on the application of Glide SP to the Astex and DUD sets, describing methods and results for symposium-provided (hereafter referred to as "as-provided") protein and ligand structures. The Schrödinger best-practices procedure for structure preparation and results when using said structures is compared to the as-provided approach. A small number of DUD best-practices preparations were modified from results presented at the ACS Symposium to eliminate obvious errors in preparation that arose from a variety of errors and misunderstandings which are enumerated in detail. Elimination of these problems, none of which have anything to do with actual Glide SP performance, improves the best practices results considerably. In "Ensemble docking with a new WaterMap-based scoring function" section, we briefly outline our new scoring and sampling model incorporating WaterMap and discuss its application to a subset of DUD targets using ensemble docking. As the method became available for testing only weeks before the Symposium, four targets of particular interest were examined. The results reported are those presented at the Symposium and provide an indication of what can be expected as these techniques are further refined and evaluated on a wider set of data. Finally, in "Discussion" section, we summarize our results and discuss future directions.

Glide SP calculations

Methods

Two sets of input structures were prepared for the current virtual screening experiments. The "as-provided" Astex and DUD set protein and ligand structures were used as prepared by the ACS Docking and Scoring Symposium organizers. The "best-practices" Astex and DUD protein and ligand datasets were prepared utilizing Schrödinger's best-practices preparation process that has been developed within the Schrödinger Drug Discovery and Applications Group.

Preparation of Astex as-provided protein structures

Symposium organizers prepared the as-provided Astex protein structures employing the preparation method summarized here. Heavy atom positions of protein–ligand complexes were refined using Refmac5 [6] (version 5.5.0109) with, in most cases, deposited protein coordinates and re-fit ligand coordinates. The ligand and protein were separated and protein hydrogen atoms were added using REDUCE [7]. All atoms were then re-refined using phenix.refine [7]. For ten systems with covalently modified protein structures (1gpk, 1hvy, 1hww, 1k3u, 1l7f, 1lrh, 1n1m, 1q4g, 1r1h, and 1r55) phenix.refine failed and hydrogen atoms were added using REDUCE and minimized using OpenEye's SZYBKI [8] application with the MMFF94S force field [9]. For 22 systems, cofactors or other crystallographic precipitants did not include hydrogen atoms. In these cases hydrogen atoms were added via Maestro's [10] add hydrogen feature and manually corrected when necessary. All cocrystallized ligand sites within the crystal structures were prepared leading some complexes to have up to eight prepared protein–ligand sites for docking. All prepared protein structures were converted to Maestro format prior to virtual screening.

Preparation of DUD as-provided protein structures

ACS Docking and Scoring Symposium organizers provided DUD protein and ligand structures derived from DUD v2 [11]. Hydrogen atoms were added to complexes using the REDUCE application. The ligand and protein were separated and protein hydrogen atoms were added using REDUCE. All atoms were then re-refined using phenix.refine. All water molecules were then stripped except where the DUD authors included active site water molecules. All prepared protein structures were converted to Maestro format prior to use in virtual screening.

Preparation of Astex and DUD best-practices protein structures

PDB structures for the IDs provided by organizers were extracted from the RCSB (www.pdb.org) [12]. Bond orders and hydrogen atoms were added using the Maestro Protein Preparation Wizard. These were manually corrected to ensure optimum cocrystallized-ligand protonation and tautomeric states. Any discrepancies between PDB structures and the literature were addressed, such as accounting for missing ligand heavy atoms. Some examples include 1ckp with missing heavy atoms, 1ba8 and 2fzi which are both clearly covalently bound, 1cx2 where there is a poor rotamer for the phenylsulfonamide and 1pxi, 1pxj, 1pxk, 1pxl, 1pxm, 1pxn, 1pxo, and 1pxp where a proximal water

molecule has been incorporated into the ligand structure. Obvious protein side-chain conformation errors were corrected such as when two or more conformations can satisfy the electron density equally well, but the conformation in the PDB is less consistent with the protein environment. An example of such a change is flipping conformations of amino acid side chain amides by 180 degrees to swap oxygen and nitrogen atom positions. Zero-order bonds and formal charges were manually assigned for protein metal ions. Missing side chains were rebuilt using Prime [13]. Low energy locations of hydrogen atoms for the entire complex structure including waters, cofactors and the cocrystallized ligand were assigned using protassign.py. The hydrogen bond network generated by protassign.py was manually evaluated and suboptimal hydrogen atom placements were corrected. The complex was relaxed to a structure suitable for enrichment experiments using impref with opls2005 force field and a tolerance of r = 0.3 Å. Finally, water molecules and extraneous HET groups were removed.

For three Astex dataset complexes the alternative ligand conformations included in the as-provided dataset were not prepared (1ig3, 1sg0, and 1tz8). One of the 40 DUD targets was not prepared for the best-practices experiments because it was derived from a homology model and deemed ill-suited for the current comparison (pdgfrb).

Preparation of as-provided DUD ligands

Structures taken from the DUD v2 site were provided by Symposium organizers. They were sorted by absolute formal charge and only a single unique molecule per compound was retained by ID. The MOE database viewer "wash" feature [14] was used to deprotonate strong acids/protonate strong bases when necessary. 3D structures for accessible low-energy ionization and tautomeric states were generated using LigPrep [15] selecting the following options: "opls2005 force field", "determine chiralities from 3D structure", "retain original state", and deselect "desalt" and "generate tautomers". There were a small number of ligands that required manual correction including seven tautomer errors, several amidine tautomer errors from the MOE wash step (see http://wiki.bkslab.org/index.php/DUD:Errata), correction of tadalafil analog chiralities to be consistent with Ref [16], and selection of the single stereoisomer most consistent with the original DUD configuration in 39 cases where the original configuration wasn't detected and two stereoisomers were generated. Additionally, 24 decoy ligands were not processed by LigPrep. These ligands were instead minimized with the OPLS-AA force field in MOE.

Ligands were processed by Epik [17] to facilitate optimal application of the GlideScore scoring function which includes a state penalty to account for the energy required to generate a given tautomeric/isomeric form of a ligand. State penalties were generated at pH of 7.0 ± 2.0 and only the provided ligand ionization/tautomeric state was retained. For cases where ligands are known to interact with a metal ion in the binding site (ace, ada, comt), the metal binding option was employed in Epik to ensure appropriate state penalties were assigned.

Preparation of best-practices Astex and DUD ligands

For best-practices runs, canonical SMILES strings were generated from the as-provided ligands using the uniquesmiles utility in Maestro. Ligands were filtered to ensure no duplicates. Next, LigPrep was run with default settings (pH = 7.0 ± 2.0) to convert ligands from SMILES representation to 3D-prepared structures for docking, retaining existing stereochemistry. For cases where ligands are known to interact with a metal ion in the binding site (ace, ada, comt), the metal binding option was employed in Epik to generate states amenable to interacting with metals and appropriate state penalties. These structures were then input into Macromodel [18] where the bmin cleanup application and a conformational search were run with the opls2005 and mmffs forcefields, keeping one lowest energy conformation per force field. In this fashion, all ligand states for known actives and decoys are provided with two different starting conformations for use in docking.

Virtual screening experiments

All Glide [19] grid generation and docking experiments were run on a cluster of 2.2 GHz AMD Opteron 2427 processors using the Schrödinger Suite 2011 release.

Glide grids for as-provided and best-practices Astex and DUD protein structures were created by an automated procedure that for each complex recognized the ligand, computed the center of the grid as the arithmetic mean of all ligand atom coordinates, and determined the outer box dimension from the maximum distance between any two atoms in the cocrystallized ligand geometry (outer box dimension = 14 + 0.8 × (max separation in Å)). Default Glide options for grid generation were employed including use of a 10 Å cubic inner box. It should be emphasized that these are the defaults employed in Maestro when setting up a Glide grid and no effort was made to tweak grid generation settings to boost Glide performance.

Glide docking experiments were performed for Astex as-provided and best-practices complexes with default options. A random conformation of the cocrystallized ligand provided by Symposium organizers was used as the input structure for all Astex docking experiments. For each

input ligand state, the Emodel scoring function is used to rank-order possible docked poses of that ligand state. Reported RMSDs are for ligand heavy atoms only and are computed via comparison to the organizer-provided cocrystallized ligand geometry in the as-provided experiment and to the Schrödinger-prepared cocrytallized ligand geometry in the best-practices experiment.

Glide docking experiments were performed for DUD as-provided and best-practices inputs with all default options. For the best-practices runs, multiple ionization/tautomeric states per ligand were docked for each ligand and for each ligand state two inputs were docked that were minimized with different force fields. For each input state/minimized geometry docked, the lowest Emodel pose was retained. All lowest Emodel poses were then sorted by DockingScore keeping only the lowest DockingScore state per ligand (identified by unique titles). Here, the DockingScore is a sum of the GlideScore and Effective Epik State Penalty.

Enrichment metrics are used to gauge the ability of a virtual screening experiment to separate known active compounds from those that are presumed to be inactive in a rank-ordered list of compounds. Four common enrichment metrics based on receiver-operator characteristic curves are employed including the area under the ROC curve (AUC) and the value on the y-axis at the 0.1, 1.0, and 2.0 % points of the x-axis of the ROC plot, referred to hereafter as ROC (0.1 %), ROC (1.0 %), and ROC (2.0 %), respectively. The ROC (1.0 %) and related metrics are used to judge early enrichment. A value of ROC (1.0 %) = 0.15 indicates that 15 % of the known actives were recovered by the time 1.0 % of known decoy compounds are screened. Selecting compounds randomly has an expectation value for ROC (1.0 %) of 0.01.

Results

Pose accuracy in the Astex dataset

Results in Table 1 show the performance of Glide in reproducing native cocrystallized ligand poses using Schrödinger's best-practices preparations versus what is observed when using the as-provided protein and ligand structures. RMSDs for each complex are provided in supplementary material. For the top-Emodel ranked pose, the best-practices average and median RMSDs are 35 % smaller than those with the as-provided inputs. When selecting the lowest RMSD pose found in the 400 refined poses of Glide, the best-practices average and median RMSDs were 24 % and 31 % improved relative to the as-provided RMSDs. These results demonstrate that in an absolute sense Glide does an excellent job of reproducing experimental bound ligand poses and that there is a significant benefit from using the best-practices preparation of input structures as compared to the as-provided inputs.

Table 1 Aggregate RMSD metrics (in Angstroms) from self-docking docking experiments using best-practices and as-provided Astex structures

	Best-practices		As-provided	
	Best Emodel	Best refined	Best Emodel	Best refined
Average	1.14	0.59	1.75	0.77
Median	0.64	0.37	0.99	0.54
SD	1.33	0.61	1.76	0.60
Maximum	8.21	4.22	8.42	3.35

The Best Emodel columns show results from the lowest Emodel pose while the Best refined columns show results of the lowest RMSD pose found among the best 400 refined poses

In Fig. 1 RMSDs for the top-ranked by EModel pose and lowest-RMSD refined pose are plotted by Astex target and ordered by increasing RMSD in the top-ranked by EModel pose. Also shown are the percent of cases that satisfy different RMSD cutoff criteria. The difference between the top-ranked Emodel and lowest-RMSD refined pose histograms in Fig. 1 reflects scoring difficulties in the Emodel function. Figure 1 illustrates for the overwhelming majority of systems a low (<2.0 Å) RMSD pose is found among the refined poses, but in some targets it is not identified as the lowest Emodel pose. Based on these results, a focus on addressing remaining Emodel scoring failures is likely to improve overall performance more significantly than addressing the remaining sampling failures.

Symposium organizers note four targets with poor electron densities for residues or cofactors near the binding site, 1j3j, 1meh, 1u1c, and 1v4s. Best-practices docking shows the best Emodel pose in all ligand dockings for these targets to be smaller than 1.5 Å RMSD. RMSDs were averaged over the multiple binding sites provided for 1u1c and 1j3j. The average RMSDs over all four targets are 0.77 Å for best practices and 1.22 Å for the as-provided structures. These averages are better than average RMSD across the entire dataset and so it can be concluded that for this dataset at least, poor electron density does not negatively effect the ability of Glide to predict high-quality docked poses.

Seven targets where crystal packing interactions are expected to play a role were noted by Symposium organizers (1gkc, 1hvy, 1jje, 1r55, 1t40, 1v48, and 2bsm). Neither as-provided inputs nor best-practices inputs include crystal mates in grid generation. Thus, Glide is unaware of the crystal mates in these experiments. Average RMSDs for a protein were computed over all available binding sites (1jje, 1gkc, and 1hvy have more than one site). For this subset of targets the average RMSDs are 1.36 and 1.86 Å

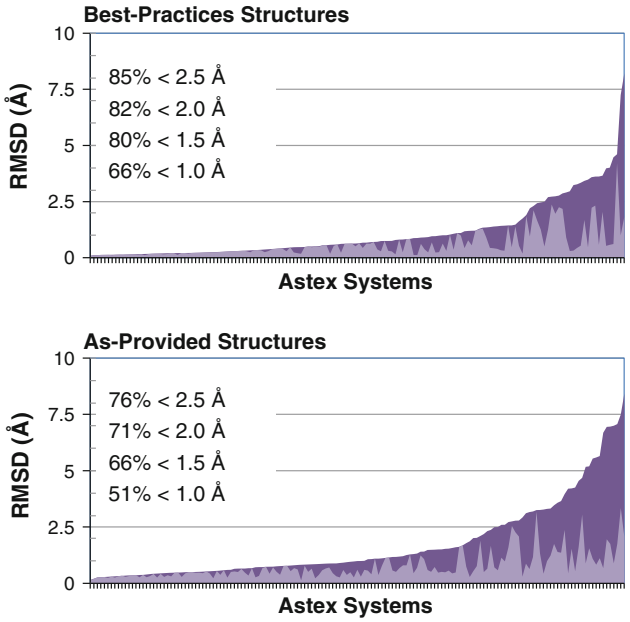

Fig. 1 Plots of RMSDs from Astex docking experiments for the top-ranked Emodel pose (*blue*) and lowest-RMSD refined pose (*light blue-gray*). The percentage of top-ranked Emodel poses that meet different RMSD cutoffs is shown

for the best-practices and as-provided experiments, respectively. While the average RMSD for cases expected to exhibit crystal packing effects is larger than the average RMSD over all Astex complexes, Glide's absolute performance on this set remains very good.

Only 1jje and 1hvy targets in the crystal packing dataset exhibited >2.0 Å RMSDs using best-practices inputs. We investigated the structural basis of their misdocking to evaluate whether the lack of crystal packing or other effects lead to the large RMSDs. For 1jje chain B, the docked ligand pose is flipped in the binding site leading to a very high RMSD as shown in Fig. 2a. The center of the ligand interacts in a similar fashion to the native ligand geometry with two acid groups interacting with protein metal ions. The poor RMSD ligand pose interacts with a solvent-exposed protein hydroxyl group, Ser 80, and obtains a slightly better Emodel energy than the native ligand. In this case positioning of the protein hydroxyl, which is interacting with solution, significantly impacts the best-Emodel pose. Another case with RMSD >2.0 Å in the best-practices docking are the four chains of 1hvy. This ligand has a flexible alkyl chain with two carboxylates that interact with water molecules in the crystal structure. These water

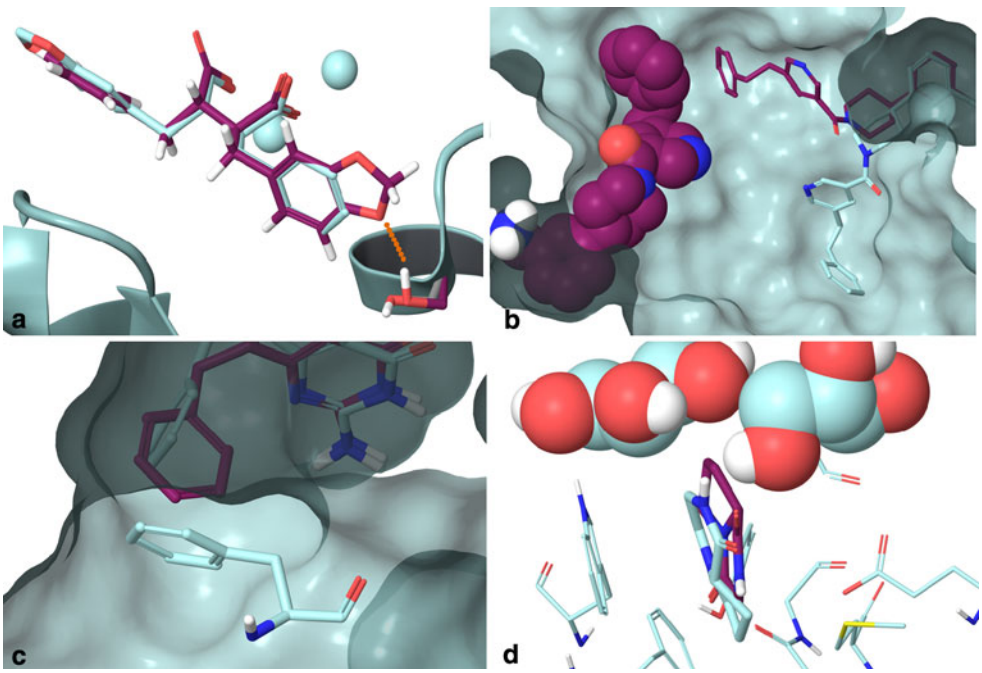

Fig. 2 Illustration of select structural differences in docked poses between as-provided (*turquoise*) and best-practices (*magenta*) Astex and DUD experiments. Protein structures have been aligned. **a** In 1jje chain B the best-practices docked pose is flipped 180 degrees relative to the low RMSD as-provided docked pose. The Ser80 hydroxyl contributing to the best-practices misdocking is shown. **b** In 2bm2 the best-practices docked pose interacts with the bound ligand from an adjacent binding site (in CPK). The as-provided molecular surface shows the cleft into which the low RMSD as-provided pose docks when the adjacent site bound ligand isn't included. **c** For the DUD as-provided pnp system (1b8o) the crystal mates Phe 459 (tube) residue limits volume accessible to the docked ligands. The best-practices docked pose for ligand ZINC 5887 is shown. **d** In 1wlp for the as-provided docking experiment two glycerol groups (GOL, shown in CPK) limit volume accessible to the ligand. The GOL groups were removed as part of the best-practices preparation. Note overlap between the best-practices docked pose and one of the GOL molecules

molecules are not present in the best-practices protein structure. Without the water molecules to interact with in the best-practices docking experiment, the carboxylates both interact with the surface Lys 77 and a backbone NH on the surface of the pocket leading to the larger RMSD. Docked poses have RMSDs between 2.43 and 2.87 Å depending on the chain.

The dependence of Glide docking on fine details of the binding site was investigated by looking at the spread in RMSDs for proteins with more than one symmetry-related binding site. As can be seen in Fig. 3, Glide shows limited variation in RMSDs for the best-practices top-Emodel pose identified across the thirty-five proteins examined. By far the largest standard deviation identified was for 1jje where the chain A docking generates a low RMSD pose and the chain B docking generates the flipped pose noted previously.

While the average top-ranked Emodel RMSD when running with best-practices inputs was smaller by 35 % than using the as-provided inputs, for a small number of cases the as-provided inputs generated lower RMSD predictions. Systems where the best-practices RMSD was >2.5 Å but the as-provided RMSD was <2.5 Å and the difference between best-practices and as-provided RMSDs was >0.5 Å were examined: 1jje, 1wlp, 1gpk, and 2bm2. The basis for 1jje chain B best-practices misdocking have already been discussed. For 1wlp two glycerol (GOL) cocrystallization cofactors were omitted from the best-practices protein structure prior to docking. These are present in the as-provided input. The best-practices docked pose overlaps significantly with one of the GOL groups as shown in Fig. 2d. A smaller RMSD pose is obtained with the as-provided inputs because the available volume for docking is restricted by the GOL cofactors. For 1gpk the reference ligand pose provided by organizers for the as-provided runs has incorrect stereochemistry for a fused ring while the best-practices reference ligand has the correct stereochemistry. Both as-provided and best-practices dockings find the same qualitative pose and the difference in RMSD arises from the different reference ligand stereochemistry. For 2bm2 we included the co-crystallized ligand in a very close adjacent binding site as seen in Fig. 2b. The docked ligand forms strong van der Waals interactions with the bound ligand in the nearby binding site, leading to identification of a higher RMSD pose with a lower Emodel energy than the native ligand geometry. Without a deeper understanding of the factors important to protein–ligand binding particular to these cases, it is hard to see how the impact of the structural differences observed could have been recognized and corrected during protein/ligand structure preparation in a prospective fashion.

Enrichment performance in the DUD dataset

Aggregate enrichments over the entire DUD set are shown in Table 2 for best-practices and as-provided inputs. Over the 39 and 40 systems examined using best-practices and as-provided inputs average AUCs of 0.80 and 0.74 were obtained. The median AUCs were 0.82 and 0.76, respectively, suggesting limited effect from outliers. Virtual screening experiments with best-practices inputs recover 12, 25, and 34 % of known actives in screening the top-ranked 0.1, 1, and 2 % of decoy compounds. With as-provided inputs, 7, 21, and 29 % of known actives are recovered in screening the top-ranked 0.1, 1 and 2 % of decoy compounds.

In Tables 3 and 4 the early and complete enrichment for the best-practices and as-provided preparations are shown,

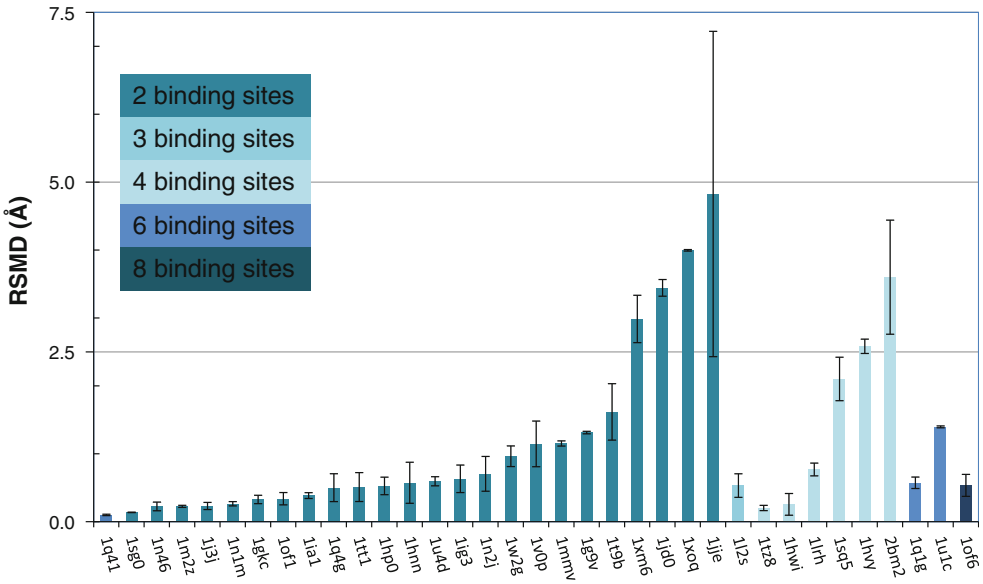

Fig. 3 Average best-practices Glide SP docked pose variability in multi-binding site Astex protein targets. *Error bars* reflect Glide dependence on fine structural details of the binding site. *Histogram bars* are *colored* by the number of binding sites per target

Table 2 Early and complete enrichment results using best-practices and as-provided structures for the DUD dataset of 40 systems

	ROC (0.1 %)	ROC (1 %)	ROC (2 %)	AUC
Best-practices structures				
Average	0.12	0.25	0.34	0.80
Median	0.06	0.19	0.27	0.82
SD	0.17	0.22	0.24	0.14
Maximum	0.82	0.88	0.88	0.98
As-provided structures				
Average	0.07	0.21	0.29	0.74
Median	0.03	0.15	0.23	0.76
SD	0.12	0.19	0.23	0.19
Maximum	0.53	0.67	0.76	0.99

There are only 39 best-practices systems as pdgfrb was not prepared using best-practices methods

respectively. The AUC metric shows Glide is superior to random selection (AUC = 0.5) in 97 % of the 39 systems examined using our best-practices input preparation. We did not prepare the pdgfrb system for best-practices experiments as a homology model was used to create the as-provided protein structure for this target. In 33 % of the best-practices systems examined AUC values >0.9 were generated. These systems included folate enzymes (gart and dhfr), one serine protease (trypsin), several nuclear hormone receptors (ar, er_agonist, gr, rxr, and mr), a cyclooxengenase inhibitor (cox2), and four systems unclassified by protein family (sahh, na, parp, and pnp). In 56 % of the best-practices systems examined, AUC > 0.8 were found, in 74 % of systems AUC > 0.7 was found, and in 95 % of cases AUC > 0.6 was found. Looking at early and full enrichment by protein family in Table 5 its clear that Glide performs well across a wide range of targets with weakest average performance in kinases where protein flexibility is known to be vitally important and in metalloenzymes which are known to be challenging to docking algorithms in general. Early enrichment is very good on a per-protein family basis and in screening only the top 0.1 % of decoy compounds Glide is able to recover between 4 and 22 % of known actives by protein family on average. This is well above random where 0.1 % of known actives would be recovered. When screening the top 1 % of decoy compounds, between 17 and 38 % of known actives are recovered by protein family on average.

Impact of structure preparation errors on enrichments

For four systems, src, comt, gpb, and pnp, enrichment with as-provided inputs significantly outperformed the initial best-practices inputs. Manual investigation of these systems found differences in protein preparation, protein structure employed, and the ionization/tautomeric states of

Table 3 Early and full enrichment results using best-practices structures for each DUD target

System	ROC (0.1 %)	ROC (1 %)	ROC (2 %)	AUC
Serine proteases				
fxa	0.06	0.23	0.27	0.75
throm	0.06	0.31	0.42	0.89
trypsin	0.00	0.09	0.34	0.95
Kinases				
cdk2	0.02	0.34	0.44	0.76
egfr	0.02	0.19	0.32	0.84
fgfr1	0.14	0.14	0.16	0.42
hsp90	0.00	0.00	0.00	0.65
p38	0.00	0.04	0.06	0.61
pdgfrb				
src	0.36	0.53	0.58	0.89
tk	0.00	0.05	0.09	0.82
vegfr	0.05	0.11	0.19	0.66
Metalloenzymes				
ace	0.00	0.04	0.12	0.73
ada	0.00	0.04	0.22	0.67
comt	0.33	0.50	0.50	0.82
pde5	0.06	0.18	0.22	0.74
Nuclear hormone receptors				
ar	0.12	0.20	0.22	0.93
er_agonist	0.10	0.22	0.40	0.93
er_antagonist	0.00	0.31	0.56	0.88
gr	0.05	0.09	0.09	0.92
mr	0.07	0.36	0.50	0.97
ppar	0.00	0.30	0.33	0.86
pr	0.04	0.04	0.04	0.51
rxr	0.00	0.40	0.75	0.91
Folate enzymes				
dhfr	0.12	0.66	0.87	0.98
gart	0.05	0.10	0.38	0.94
Other				
ache	0.00	0.01	0.06	0.68
alr2	0.08	0.08	0.12	0.74
ampc	0.00	0.00	0.05	0.66
cox1	0.04	0.12	0.16	0.75
cox2	0.43	0.73	0.80	0.97
gpb	0.06	0.08	0.09	0.82
hivpr	0.06	0.21	0.25	0.68
hivrt	0.08	0.15	0.25	0.83
hmgr	0.31	0.43	0.43	0.79
inha	0.24	0.34	0.35	0.63
na	0.37	0.59	0.65	0.95
parp	0.33	0.61	0.73	0.97
pnp	0.07	0.12	0.23	0.93
sahh	0.82	0.88	0.88	0.96

The pdgfrb system was not prepared using best-practices methods

Table 4 Early and full enrichment results using as-provided structures for each DUD target

System	ROC (0.1 %)	ROC (1 %)	ROC (2 %)	AUC
Serine proteases				
fxa	0.04	0.26	0.32	0.78
throm	0.03	0.31	0.37	0.88
trypsin	0.07	0.27	0.75	0.93
Kinases				
cdk2	0.06	0.16	0.36	0.74
egfr	0.05	0.12	0.19	0.63
fgfr1	0.07	0.10	0.12	0.42
hsp90	0.00	0.00	0.00	0.66
p38	0.00	0.00	0.02	0.56
pdgfrb	0.06	0.06	0.06	0.48
src	0.30	0.44	0.50	0.68
tk	0.00	0.00	0.05	0.87
vegfr	0.09	0.15	0.23	0.57
Metalloenzymes				
ace	0.00	0.12	0.20	0.67
ada	0.00	0.00	0.00	0.53
comt	0.00	0.00	0.27	0.73
pde5	0.00	0.18	0.20	0.75
Nuclear hormone receptors				
ar	0.14	0.26	0.34	0.96
er_agonist	0.01	0.63	0.70	0.97
er_antagonist	0.00	0.33	0.54	0.83
gr	0.05	0.12	0.13	0.95
mr	0.53	0.53	0.53	0.99
ppar	0.02	0.43	0.49	0.88
pr	0.04	0.04	0.07	0.48
rxr	0.00	0.50	0.60	0.93
Folate enzymes				
dhfr	0.02	0.04	0.06	0.40
gart	0.00	0.19	0.57	0.96
Other				
ache	0.00	0.01	0.02	0.67
alr2	0.04	0.08	0.12	0.72
ampc	0.00	0.00	0.00	0.29
cox1	0.00	0.08	0.12	0.80
cox2	0.27	0.67	0.76	0.95
gpb	0.02	0.06	0.12	0.87
hivpr	0.09	0.21	0.23	0.42
hivrt	0.03	0.13	0.20	0.74
hmgr	0.29	0.34	0.34	0.76
inha	0.01	0.04	0.06	0.45
na	0.04	0.31	0.43	0.91
parp	0.12	0.39	0.61	0.92
pnp	0.00	0.20	0.24	0.96
sahh	0.39	0.48	0.55	0.89

ligands lead to the large enrichment differences. These enrichment differences illustrate the importance of understanding the physics of binding in each system studied in order to achieve maximum benefit from a virtual screening experiment. This is common in real-world application of virtual screening to a particular target, but is limited when examining tens of systems in a short timeframe. We believe this is a common weakness in many docking methodology comparison papers.

For best-practices runs of src the 2src PDB structure noted in the DUD paper was employed. Best-practices enrichments for 2src were significantly worse than those obtained using the as-provided structures. Protein sequence homology shows the as-provided protein was actually 1y57. The 1y57 protein was prepared following the best-practices procedure and significantly improved early and overall enrichment results were obtained. AUC improved from 0.40 to 0.87 when the protein was changed from 2src to 1y57 with the best-practices preparation. Early enrichment significantly benefited when using 1y57 instead of 2src with ROC (0.1 %), ROC (1 %) and ROC (2 %) improvements of 3,489, 786, and 724 %. Comparison of 1y57 best-practices and as-provided results shows the updated best-practices preparation to be superior with AUC, ROC (0.1 %), ROC (1 %) and ROC (2 %) improvements of 28, 20, 21, and 15 %, respectively.

Investigation of the comt system revealed an aromatic 1,2-diol moiety on most of the DUD known actives was not provided the necessary ionization state to interact with a protein metal ion. This was because of failure in Epik enumeration of protonation/tautomerization states in the suite2011 release to identify the appropriate state. This weakness was largely addressed in the suite2011 update release. Decoy and known active ligands were reprepared via LigPrep using the suite2011 update within the best-practices procedure. Now that Glide has the opportunity to form and score a physically reasonable protein–ligand interaction with the 1,2-diol moiety, enrichment performance improves from 0.55 to 0.82 AUC. Early enrichment significantly benefits with ROC (0.1 %), ROC (1 %) and ROC (2 %) results improved over the original best-practices results by 266, 177, and 85 %. The suite2011 update best-practices preparation is clearly superior to the as-provided enrichments, showing significant early enrichment of 0.33, 0.5, and 0.5 for ROC (0.1 %), ROC (1 %), and ROC (2 %) while the as-provided run generates 0.0, 0.0, and 0.27, respectively.

A similar basis for differences in as-provided and preliminary best-practices enrichments was found in investigation of the gpb system. Here investigation of the docked poses for known actives revealed incorrect chiralities for some sugar moieties. This error was introduced by a bug in LigPrep that was fixed for the suite2012 release. Using a

Table 5 Aggregate best-practices DUD dataset enrichment as a function of protein family

Family	Number	AUC		ROC (0.1 %)		ROC (1.0 %)	
		Average	Median	Average	Median	Average	Median
Nuclear hormone receptors	8	0.86	0.92	0.05	0.04	0.24	0.26
Kinases	9	0.71	0.71	0.07	0.02	0.17	0.13
Serine proteases	3	0.86	0.89	0.04	0.06	0.21	0.23
Metalloenzymes	4	0.74	0.74	0.10	0.03	0.19	0.11
Folate enzymes	2	0.96	0.96	0.09	0.09	0.38	0.38
Other	13	0.82	0.82	0.22	0.08	0.33	0.21

near-release candidate for suite2012, active and decoy DUD ligands were re-prepared. Docking was then performed with the suite2011 release into the original best-practices protein structure. This addressed the chirality issue and improved AUC by 93 %. Early enrichment also benefited going from ROC (0.1 %), ROC (1 %), and ROC (2 %) of 0.0, 0.02, 0.06 with the suite2011 preparations to 0.06, 0.08, and 0.09 with the suite2012 preparations.

Finally, it was found the best-practices protein created for pnp did not include a key crystal mate of the PDB structure illustrated in Fig. 2c. This crystal mate closed the binding site through residue Phe 459. This is expected to lead to improved enrichment due to reduced sampling space. Additionally there were a small number of known actives for pnp where LigPrep flipped chirality in a sugar group. The protein was re-prepared including the entire key crystal mate with the suite2011 update release. Using a near-release candidate for suite2012, active and decoy DUD ligands for pnp were reprepared. Grids were regenerated and docking was run with the suite2011 release. The AUC metric for pnp improved from 0.82 to 0.92 with the updated best-practices structures. Very early enrichment with ROC (0.1 %) improved from 0.00 to 0.07 while ROC (1.0 %) and ROC (2.0 %) decreased from 0.16 to 0.12 and from 0.36 to 0.23. Since enrichment for pnp is acceptable in an absolute sense, we chose not to pursue this further.

All best-practices DUD Glide SP enrichment results presented in this article include the four modifications described above. We believe the modified results represent a more accurate picture of what a user of Glide can expect with the current release version when care is taken to properly prepare the receptor and input ligands for docking. Besides these four systems with obvious preparation issues, no changes in protein or ligand preparation, Glide scoring, or docking were made to improve retrospective results.

In the following section we present preliminary results for four problematic targets in which conformational flexibility and failure to properly account for desolvation and strain effects are addressed. We believe these are the two principal remaining problems to be addressed in virtual screening experiments. These results are unchanged from what was presented at the ACS Symposium. They represent a snapshot of our advanced development efforts and provide a preview as to what can be expected when this effort is completed.

Ensemble docking with a new WaterMap-based scoring function

Methods

While most of the DUD test cases yield good enrichments using Glide SP, in a subset of cases inadequate enrichment performance for use in practical virtual screening experiments is obtained. An AUC score of about 0.8 indicates that there are a substantial number of decoys outscoring the known DUD actives and truly robust performance is indicated by an AUC score that is in excess of 0.9. In 44 % of targets the Glide SP best-practices AUC is below 0.8 and for 23 % of targets it is greater than or equal to 0.8 and less than 0.9.

We believe there are two principal reasons for degraded performance in a virtual screening exercise following the present protocol. First, some subset of the active compounds may simply not be able to dock in a reasonable pose into the single receptor conformation selected for docking. Given the well known importance of induced fit effects in protein–ligand binding [20, 21], and our observation with src described above, such a result would neither be surprising nor does it indicate anything about the quality of the scoring function employed in rank ordering the compounds. Any scoring function that is going to reject decoys effectively must have a "hard" component defining the shape of the binding site. Some softening of the potential function is possible to allow for minor variation in ligand size and shape, but major changes in either side chain or backbone positioning can render such expedients ineffective in properly docking ligands that natively bind to a substantially different protein conformation. Secondly, the scoring function may have flaws that fail to recognize strong binders or reject nonbinding compounds. Our

experience with the Glide SP scoring function to date is that the problem is primarily the latter, that is, the scoring function does not always properly penalize various effects, principally significant strain energy in the ligand and desolvation effects.

In this section, we present preliminary results addressing both the effects described above. Rather than docking into a single receptor conformation, we have selected three to four receptor conformations for each target and docked the actives and decoys into all members of the ensemble. The use of ensemble docking permits a variety of ligand chemotypes to be accommodated, while increasing computation time only by a linear factor. The final score of a compound is computed by taking the best score across the ensemble, and then reducing this score by a preset penalty if drastically more positive scores are obtained in all other ensemble members. Active compounds generally obtain decent (if not optimal) scores in at least one other ensemble member in addition to the top scoring complex, whereas decoys often display only one reasonable score, accruing substantial penalties in the other receptor conformations. This is indicative of a poor fit of the ligand, leading to strain and/or decreased entropy, and we penalize this situation accordingly.

The second modification made is to incorporate quasi-localized water sites from a WaterMap analysis of a molecular dynamics simulation into the sampling algorithm and scoring function. To generate a WaterMap for each receptor, the ligand is removed, a molecular dynamics simulation is carried out, and regions of water density that substantially exceed bulk density are identified as quasi-localized water sites [22–28]. The free energy required to displace these waters is estimated from a modified version of inhomogeneous solvation theory (IHT) [29–31]; various computational benchmark calculations and comparison with experimental data have established that the IHT approximation is a reasonably good one. The waters are then placed within the grid used for Glide docking, and interaction of the ligand with these waters is incorporated into the scoring function. We used the Glide XP [32] scoring function as a starting point, but there has been sufficient modifications that it is more accurate to think of the resulting scoring function as a new construction that retains a subset of features of the older approach. Details of the new scoring function, which we designate WaterMap-score or WScore, will be presented in a subsequent publication. It should be noted the WScore method has not yet been publically released. The key penalty terms are briefly described below.

1. Strain energy terms based on short intra-ligand contacts and torsions where conjugation of two or more aromatic groups is disrupted when conjugation is possible in solution. Development of such terms requires the docking of active compounds to be able to generate binding modes that are reasonably close to those in their cognate receptor conformations. The ensemble docking approach, assuming a suitable ensemble, allows these terms to be optimized for imperfect, but relatively minor, cross docking effects.
2. Desolvation penalties are applied when WaterMap water sites in hydrophobically enclosed regions are displaced by ligand groups that leave a polar protein group with an unsatisfied hydrogen bond.
3. If an unstable WaterMap water site is contacted, but not displaced, by a hydrophobic group of the ligand, the water may have increased difficulty making hydrogen bonds, and hence may lose enthalpy and/or entropy. Such contacts are not seen in a database of native cocrystallized structures from the PDB containing around 600 complexes. Consequently, we penalize structures that exhibit contacts of this type.

Results

The four systems we chose to study using ensemble docking and WScore are listed in Table 6, along with the number of members of each ensemble and the PDB ID codes of receptor structures employed. The receptor conformations were selected based on previous experience with these targets and display a good ability to accommodate a relatively wide range of ligand chemotypes. In the case of hsp90, there is a large conformational change (flipping of Leu 107) between an "open" and "closed" form of the receptor. As the ligands that exceeded our minimum standards for potency (<10 μM binding affinity) dock only into the closed form of the receptor, several closed form receptors were selected for the ensemble. Interestingly, overall results are improved if an exclusively closed form ensemble is employed. We present results with such an ensemble as well as with one that is balanced between open and closed forms to illustrate differences in performance that arise from this choice. Of the other receptors, factor Xa is a relatively rigid protein but does require some structural diversity in the receptor ensemble

Table 6 Ensembles used in WScore docking

System	Num. members	PDB IDs
cdk2	4	1h1r, 1h1s, 1oiu, 2bpm
hivrt	3	1c0u, 1hni, 1rt1
hsp90 (all/closed form)	4/2	1osf, 1yc4, 2fwy, 2h55/1yc4, 1osf
fxa	4	1fjs, 1lpk, 1z6e, 2bok

Table 7 Early and full enrichment performance for WScore and Glide SP docking and scoring functions when using all DUD active ligands or a subset of actives with <10 μM experimental binding affinity for cdk2, hivrt, hsp90, and fxa

Methodology	Num. active ligands	ROC (0.1 %)	ROC (1 %)	ROC (2 %)	AUC
cdk2					
WScore	50	0.08	0.34	0.48	0.88
WScore (<10 μM)	41	0.10	0.37	0.49	0.92
Glide SP	50	0.02	0.34	0.44	0.76
hivrt					
WScore	40	0.08	0.35	0.38	0.85
WScore (<10 μM)	35	0.09	0.40	0.43	0.90
Glide SP	40	0.08	0.15	0.25	0.83
hsp90					
WScore	24	0.00	0.04	0.29	0.89
WScore (<10 μM)	11	0.00	0.00	0.36	0.91
WScore closed-form	24	0.08	0.29	0.33	0.84
WScore closed-form (<10 μM)	11	0.18	0.64	0.73	0.92
Glide SP	24	0.00	0.00	0.00	0.64
fxa					
WScore	142	0.13	0.68	0.78	0.93
WScore (<10 μM)	122	0.16	0.62	0.71	0.94
Glide SP	142	0.06	0.23	0.27	0.75

The closed-form ensemble of hsp90 is comprised of 1yc4 and 1osf

in order to accommodate ligands of different size. Hivrt is very flexible though the DUD ligands exhibit only a few chemotypes and low RMSD binding modes to several of the "normal" receptor conformations selected for our ensemble. As in the case of factor Xa we suspect a single receptor is inadequate to accommodate the range of variability seen in the known actives data set. It should be remembered that the NNRTI binding site targeted by ligands in the DUD set is an allosteric pocket, which does not exist in the apo form of the enzyme. Cdk2 lies somewhere in between factor Xa and the other two receptors in plasticity, but the ensemble we have selected again has coverage of the ligand chemotype space that is well suited to the DUD ligands. Even better results could likely be obtained by optimization of the receptor conformations to the DUD active compounds, but this is a direction that we did not pursue, beyond ensuring that the ensembles we choose produced suitable binding modes for a high fraction of known actives.

WScore enrichment results are shown for all four systems in Table 7 along with the best-practices single receptor Glide SP results. Given that this methodology remains under development these results should be considered preliminary. ROC data for 0.1, 1, and 2 % of the database is shown along with the standard AUC metric. Substantial improvements for WScore over Glide SP enrichment can be noted in all cases. Most critically, the AUC with WScore is above 0.9 for all four targets, and the ROC at 2 % of the database is quite substantial, ranging between 50 and 80 % of the seeded actives. We believe this is a realistic cutoff in terms of what a typical project in the pharmaceutical industry would be willing to survey experimentally. In contrast, the single receptor Glide SP results have AUC values which are all below 0.9 (and three of four are below 0.8), and the 2 % ROC values range from 0.0 to 0.44. These results suggest that the two factors identified above, use of a single receptor conformation and failure to penalize decoy ligands sufficiently, account for a very high fraction of the problems observed. Furthermore, the combination of ensemble docking and improvements to the scoring function go a significant way towards remedying these deficiencies.

In Table 7, we show results for the entire set of DUD actives and a subset with smaller than 10 μM affinity, a typical threshold used in high throughput screening experiments to separate hits from inactive compounds. Actives with less than 10 μM binding affinities will be more difficult to identify and also are of less interest in a real-world lead discovery program. Furthermore, when dealing with very weak binders, the question of how many compounds in the "decoy" set are actually more active than the very weak seeded compounds becomes significant. This is particularly relevant in the case of DUD decoys where the similarity-based method of selecting difficult decoy ligands may enrich the decoy set in weakly active compounds. Differences between the two data sets are not enormous in general, but do have an impact particularly in the ROC (2 %) metric when using a closed form ensemble of receptors for hsp90.

In the present paper, we do not attempt to distinguish how much of the improvement reported here is due to the use of ensemble docking, and what percentage can be attributed to the use of a different (presumably improved) scoring function. This is clearly an important issue, and we plan to examine it in detail in future work, when the development of WScore has been finalized and tested on a much larger and diverse set of targets, active compounds, and decoys. What we have established is that there is a path towards creating a virtual screening protocol that can robustly deliver superior performance for a wide range of targets, by docking into multiple structures and by working on improvements to scoring. How far one can go in this direction is an interesting question which remains to be determined in subsequent publications.

Discussion

Virtual screening has proven to be a very successful tool for identifying active compounds at low cost. The current best-practices results effectively illustrate this utility. Glide reproduces crystal complex geometries in 85 % of Astex cases with <2.5 Å RMSD. Glide is also effective in enriching screened databases in known actives, beating random selection in 97 % of targets and giving an average AUC of 0.80 across 39 target systems. Early enrichment performance shows on average 12, 25, and 34 % of known actives are recovered in screening the top-ranked 0.1, 1, and 2 % of recovered decoys. This is particularly impressive given that docking a ligand with Glide in SP mode takes only about 10 s using modern hardware.

Our results clearly show dataset preparation makes a large difference in pose prediction and enrichment experiments for Glide. This is not surprising, given that both Emodel and GlideScore are relatively hard scoring functions with limited softening of non-bonded interactions. We believe that virtual screening experiments using best-practices input preparations are more realistic and representative of practical application given that all low-energy ionization/tautomeric states of all ligands are docked into a manually prepared protein structure. For enrichment at least, the best-practices experiment is conceptually more difficult then the as-provided experiment as only in the best-practices test is the docking algorithm and scoring function asked to find and give a good score to a low RMSD pose with the ideal ionization/tautomeric state of a known active ligand while avoiding giving a good score to any of the multiple ionization/tautomeric states of decoy compounds.

Enrichment performance is consistently good across a wide range of targets with the exception of some kinases that generate worse than random AUC enrichment. In these cases, protein flexibility is assumed to be the culprit. This conclusion is supported by the improved results obtained for a kinase (cdk2) and the heat shock protein (hsp90) using ensemble docking and the WScore scoring function. The hsp90 case is particularly dramatic; the initial structure provided by the organizers for docking DUD ligands for this target is not well-suited to accommodate the dominant chemotype in most DUD active ligands. From the structural issues identified in this paper, there is room for improvement in addressing limitations in structure preparation, both ligand and protein.

The combination of Glide SP and ensemble docking/WScore results provide a fair picture of the current state of the art in docking and scoring. If one wants to rapidly perform a virtual screen with the minimal amount of effort, the protocol used for the Glide SP calculations, in which a single receptor conformation is employed and the calculations per ligand are extremely rapid, delivers strong performance across a diverse array of targets and ligands. However, such an approach cannot be guaranteed to succeed 100 % of the time, nor will it always locate a high fraction of the more weakly binding low micromolar/high nanomolar ligands that are typically going to be present in prospective virtual screening of an in house or external purchasable deck of drug-like compounds. This is quite different from what is measured in retrospective studies in which many ligands have low nanomolar/picomolar binding affinities because they are products of elaborate medicinal chemistry programs where interactions with the receptor have been optimized. Improved performance in real-world application requires much more intensive effort where multiple receptor conformations are routinely employed and more computationally expensive docking approaches such as WScore are used at the final stages of filtering compounds. An extensive evaluation of the possibilities inherent in this latter approach will be provided in subsequent publications.

Supplementary Material

RMSDs for best-practices and as-provided Astex experiments are provided.

Acknowledgments The authors would like to thank Nick A. Boyles, Amy R. Rask, K. Shawn Watts, and Carolyn M. McQuaw of Schrödinger, LLC for their collection and curation of ligand binding affinities from the DUD and Astex ligand sets.

References

1. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin P (2004) J Med Chem 47(7):1739–1749

2. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) J Med Chem 47(7):1750–1759
3. Hartshorn MJ, Verkonk ML, Chessari G, Brewerton SC, Mooij WTM, Mortenson PN, Murray CW (2007) J Med Chem 50(4):726–741
4. Huang N, Shoichet BK, Irwin JJ (2006) J Med Chem 49(23):6789–6801
5. Abel R, Young T, Farid R, Berne BJ, Friesner RA (2008) J Am Chem Soc 130(9):2817–2831
6. Vagin A, Steiner R, Lebedev A, Potterton L, McNicholas S, Long F, Murshudov G (2004) Acta Cryst D60:2284–2295
7. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) J Mol Biol 285(4):1735–1747
8. Wlodek S, Skillman AG, Nicholls A (2010) J Chem Theory Comput 6(7):2140–2152
9. Halgren TA (1999) J Comp Chem 20(7):720–729
10. Maestro v9.2. Schrödinger LLC. New York
11. DUD v2. http://dud.docking.org/r2/
12. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
13. Prime v3.0. Schrödinger LLC. New York
14. MOE. Chemical Computing Group, Inc. Montreal
15. LigPrep v2.5. Schrödinger LLC. New York
16. Jorissen RN, Gilson MK (2005) J Chem Inf Model 45(3):549–561
17. Epik v2.2. Schrödinger LLC. New York
18. Macromodel v9.9. Schrödinger LLC. New York
19. Glide v5.7. Schrödinger LLC. New York
20. Sherman W, Beard HS, Farid R (2006) Chem Biol Drug Des 67:83–84
21. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R (2006) J Med Chem 49:534–554
22. Young T, Abel R, Kim B, Berne B, Friesner RA (2007) PNAS 104:808–813
23. Abel R, Young T, Farid R, Berne B, Friesner RA (2008) J Am Chem Soc 130:2817–2831
24. Beuming T, Farid R, Sherman W (2009) Protein Sci 18:1609–1619
25. Pearlstein R, Hu Q, Zhou J, Yowe D, Levell J, Dale B, Kaushik V, Daniels D, Hanrahan S, Sherman W, Abel R (2010) Proteins 78:2571–2586
26. Higgs C, Beuming T, Sherman W (2010) ACS Med Chem Lett 1:160–164
27. Robinson D, Sherman W, Farid R (2010) Chem Med Chem 5:618–627
28. Abel R, Salam N, Shelley J, Farid R, Friesner RA, Sherman W (2011) Chem Med Chem 6:1049–1066
29. Lazaridis T (1998) J Phys Chem B 102:3531–3541
30. Lazaridis T, Karplus M (1996) J Chem Phys 105:4294–4316
31. Lazaridis T, Paulatis ME (1992) J Phys Chem 96:3847–3855
32. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) J Med Chem 49(21):6177–6196