Virtual Fragment Docking by Glide: a Validation Study on 190 Protein−Fragment Complexes

Márk Sándor, Róbert Kiss,† and György M. Keserű*

Discovery Chemistry, Gedeon Richter plc., P.O. Box 27, H-1475 Budapest, Hungary

Received January 29, 2010

The docking accuracy of Glide was evaluated using 16 different docking protocols on 190 protein−fragment complexes representing 78 targets. Standard precision docking (Glide SP) based protocols showed the best performance. The average root-mean-square deviation (rmsd) between the docked and cocrystallized poses achieved by Glide SP with pre- and postprocessing was 1.17 Å, and an acceptable binding mode with rmsd < 2 Å could be found in 80% of the cases. Comparison of the docking results produced by different protocols suggests that the sampling efficacy of Glide is adequate for fragment docking. The docking accuracy seems to be limited by the performance of scoring schemes, which is supported by the weak correlation between experimental binding affinities and GlideScores. Cross-docking experiments performed on 8 targets represented by 63 complexes revealed that Glide SP gave similar results to that of the computationally more intensive Glide XP. The average rmsd achieved by Glide SP with pre- and postprocessing was 2.06 Å, and an acceptable binding mode with rmsd < 2 Å could be found in 63% of the cases. These cross-docking results were improved significantly selecting the optimal X-ray structure for each target (average rmsd = 1.3 Å, success rate = 77%), indicating the importance of enrichment studies and the use of multiple X-ray structures in virtual fragment screening.

INTRODUCTION

Fragment-based drug discovery (FBDD) is an increasingly popular paradigm in early phase drug discovery.[1−3] One of its main advantages is that fragment space can be sampled more effectively than drug-like space. Fragment hits are expected to be more suitable starting points for hit-to-lead optimization due to their reduced complexity, which leaves more freedom to medicinal chemists for multidimensional property optimization. On the other hand, fragments are typically less potent and often promiscuous binders, which makes their identification challenging. Biophysical approaches, such as X-ray diffraction, nuclear magnetic resonance (NMR), mass spectrometry (MS), and surface plasmon resonance (SPR), however, were routinely used in detecting low-affinity fragments. Considering that the estimated size of the chemically relevant fragment space[4] is at least in the order of 10^8 and that the number of commercially available fragments is over 300 000,[5] the throughput and also the associated costs limit the application of brute force biophysical screening. Structure-based virtual screening (SBVS) of drug-like libraries has already been proved to be an efficient technology in hit discovery.[6,7] Furthermore, SBVS was integrated with experimental screening, providing drug-like hits for a large variety of targets.[8,9] Based on these positive experiences, one can conclude that SBVS could also support FBDD by docking and prioritizing fragments. However, the applicability of SBVS on fragment space is still to be fully investigated. Accuracy of binding mode prediction and ranking of fragments needs to be at least comparable to those of drug-like molecules. However, docking fragments to protein binding sites can be a challenging task for multiple reasons. First, there are a number of interaction sites on protein surfaces accommodating low molecular weight compounds, such as solvents,[10] that might result in false positives in virtual fragment screening. Second, the binding cavity of target proteins can be significantly larger than the molecular volume of fragments that might result in incorrect binding modes.[11] Third, most scoring functions have been developed to reproduce the binding energetics of drug-like compounds having much higher affinity than fragments. Since the development of reliable scoring schemes of drug-like molecules is still challenging,[12] scoring and ranking of fragments forming less interactions with the binding site can be even more problematic.

Although docking programs were thoroughly evaluated on a large number of data sets, including different targets and screening libraries,[13] investigation of their performance in virtual fragment screening is rarely reported. Most of the published studies were conducted on a single target only.[5,14,15] Up to now, only two studies have been reported on multiple targets. Marcou and Rognan[16] analyzed the performance of four docking programs (FlexX, Glide, Gold, and Surflex) on 42 protein−fragment complexes, comparing the predicted binding modes of fragments to those obtained by X-ray crystallography. They found that Gold, Surflex, FlexX, and Glide could predict the correct pose for 70, 60, 40, and 70% of the fragments, respectively; correct pose in this context means the reproduction of native interaction motifs in 60%, which roughly corresponds with the widely used root-mean-square deviation (rmsd) criteria of 2 Å. A recent paper published by Loving et al.[17] investigated the accuracy of

* Corresponding author. Telephone: +36-1-4314605. Fax: +36-1-4326002. E-mail: gy.keseru@richter.hu.

† Present address: Chemical Engineering, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, Scotland, U.K.

10.1021/ci1000407
Published on Web 05/11/2010

fragment docking by Glide. These authors demonstrated that Glide XP docks fragments accurately to the binding site of 12 targets within the rmsd limit of 1 Å relative to the experimental binding mode, as calculated for the top scoring pose.

Glide is one of the most popular docking tools routinely used in SBVS with significant success.[18] It can predict binding modes with acceptable accuracy and has contributed to the discovery of potent drug-like ligands against a wide variety of targets. The above-mentioned two studies—using limited and different sets of targets—suggested a different performance in fragment docking for Glide that prompted us to investigate its performance using a comprehensive set of 190 protein−fragment complexes. This set was compiled from X-ray structures cocrystallized with low molecular weight (MW) ligands (MW < 300) available in AffinDB,[19] in the Astex Diverse Set (AstexDS),[20] and in the FBDD literature.[1−3,21] In this paper, we report the evaluation of Glide in predicting both experimental binding modes and affinities for 190 fragment−protein complexes.

COMPUTATIONAL METHODS

Compilation of the Fragment Database. Our docking database contains protein−fragment complexes with ligands having MW < 300 from three different sources. AffinDB is an annotated database of complexes deposited in the Protein Data Bank (PDB) with measured binding affinities. This database contains several ligands cocrystallized with the same protein. We omitted structures with identical binding modes and finally added 88 complexes from this source. Experimental binding affinity data for 86 out of the 88 cocrystallized fragments were also available. The AffinDB data set was extended by eight complexes (1eqg, 1fv9, 1gwq, 1qwc, 1s39, 1yz3, 2adu, 2jjc) retrieved from PDB by Congreve et al.[2] The next source of protein−fragment complexes was Astex Diverse Set (AstexDS) compiled from the PDB for docking and scoring performance validation. All of the entries from this subset are high-resolution crystal structures, and no particular target is represented more than once. More importantly ligand binding modes in this subset are fully supported by ligand electron densities, and ligands are in contact with protein atoms from the given crystallographic subunit. Thirty-four complexes not represented in the AffinDB subset were selected from this source, of which 15 had experimental binding affinity data. Finally, 60 additional protein−fragment complexes from the current FBDD literature were also included, of which one had experimental binding affinity data. It is important to mention that this latter subset is redundant to the target since the same proteins with different ligands are typically cocrystallized during the hit-to-lead process. Altogether, our fragment docking database comprises 190 complexes: 96 complexes (50%) from AffinDB, 34 complexes (18%) from AstexDS, and 60 complexes (32%) from the FBDD literature (see the Supporting Information). In 102 out of the 190 cases, experimental binding affinity (K_i) data are also available. Our data set includes 78 targets and 177 unique fragments in total, from which 46 targets and 91 fragments are represented in the subset with affinity data. Our data sets are available from the authors upon request. Further information about the targets and ligands in the different subsets is shown in Table 1.

Table 1. Database Composition[a]

	AffinDB	AstexDS	FBDD literature	total
number of complexes	96	34	60	190
number of unique fragments	85 (89%)	33 (97%)	60 (100%)	177 (93%)
number of targets	42 (44%)	34 (100%)	15 (25%)	78 (41%)

[a] Unique structures among the complexes from different sources.

Physicochemical properties of fragments were calculated by the calculator plugins of ChemAxon.[22] Investigation of these properties revealed that our ligand database fulfills the criteria of rule-of-three.[23] Average values and standard deviations of relevant properties, such as molecular weight and lipophylicity and the number of hydrogen-bond acceptors and donors were MW = 211 ± 54, ClogP = 0.9 ± 1.9, HBA = 3.38 ± 1.9, and HBD = 1.72 ± 1.21, respectively.

Ligand and Protein Preparation. Ligands of the AstexDS subset were acquired from the Web site of Cambridge Crystallographic Data Centre[24] in prepared form. All the other ligands were directly downloaded from the Chemical Component Dictionary (CCD) of PDB[25] as SMILES strings. In this case, the pK_a plugin of ChemAxon[22] was used to generate the appropriate ligand protonation states dominant at the pH of crystallization, preserving the tautomeric form found in CCD. Finally, three-dimensional (3D) coordinates of all ligands (including the ligands in the AstexDS) were generated by the LigPrep[26] module in the Schrödinger Suite 2008 combined with the molconvert tool in the Chemaxon Suite,[22] which replaced the native conformations and served as starting geometries for docking.

Crystal structures derived from AffinDB and FBDD literature were subjected to preparation steps using the Protein Preparation Wizard in Maestro:[27] waters were removed, bond orders were assigned, hydrogens were added, and metal atoms were treated automatically. Next, the orientation of amide (Asn and Gln), hydroxyl (Ser, Thr, and Tyr), and thiol groups (Cys) and the protonation and tautomeric state of His residues were optimized using the exhaustive sampling option. Protein−ligand complexes from the AstexDS were downloaded in a prepared form with correct geometry and charges; in this case, only the metal atoms were treated to achieve their Maestro-compatible form.

In the final preparation step, restrained minimization of all complexes was applied as an option (0.3 Å rmsd constraint, OPLS_2001 force field),[28] referred to as the preprocessing step throughout the manuscript.

Receptor Grid Generation. For each protein structure, a grid box of 30 × 30 × 30 Å^3 with a default inner box (10 × 10 × 10 Å^3) was centered on the corresponding ligand. Default parameters were used, and no constraints were included.

Docking Protocols. Docking calculations involved 16 different docking protocols. These protocols were combined from two docking algorithms, including Glide[29] Standard Precision (SP), and Extra Precision (XP); two sampling options, including default and expanded sampling (indicated by -ES suffix), and four processing options, including both pre- and postprocessing (default), both pre- and postprocessing only, and lack of pre- and postprocessing.

In the case of expanded sampling, we used the sampling scheme described by Loving et al.[17] The number of poses

per ligand for the initial phase of docking was increased to 50 000, a wider scoring window of 500.0 kcal/mol was used for keeping initial poses, and the best 1000 poses per ligand were kept for energy minimization. Glide was set to bypass the rough score sorting step and to minimize all 50 000 poses on the Glide grid.

The term preprocessing refers to the restrained minimization of the receptor−ligand complex in the OPLS_2001 force field, which is part of the protein preparation process recommended by Schrödinger. Postprocessing refers to postdocking minimization introduced in Glide versions 4.5 and 5.0 for the SP and XP docking, respectively, and included as default in the docking process. To investigate their effect on the accuracy, we decided to apply them as separated components in the docking protocols.

In all protocols Glide was set to write out the 15 best poses per ligand. Docking runs and grid generations were performed automatically using the Dock and the GridGen classes available in the Glide Python module, respectively.

Docking Analysis. Heavy atom rmsd values between the native and docked conformations were calculated by the ConformerRmsd class available in the Schrödinger Python API applying default settings. Physicochemical properties (MW, logP, HBD, HBA, and number of rotatable bonds (rotbond)) of fragments were calculated by the calculator plugins of ChemAxon.[22] Identification of open binding sites was carried out using the SiteMap[30] application available in the Schrödinger Suite. SiteMap was instructed to write out the site volume (volume points with 0.7 Å grid spacing) around the ligand using the −keepvolpts command line parameter. A particular binding site was considered to be open if any of the ligand atoms in the native pose was positioned out of the site volume, i.e., its distance from the nearest site volume point was greater than 1.5 Å.

Cross-Docking Experiments. Our database contains more than five representative PDB structures with sequentially identical binding sites for nine targets. Except one case (Posphodiesterase 4D) which has poor self-docking statistics, these targets (beta secretase, cAMP-dependent protein kinase, cell division protein kinase 2, cytochrome p450cam, penicillin amidohydrolase, trypsin, tryptophan synthase, and tyrosine−protein kinase SRC) were subjected to cross-docking calculations. Because cross-docking accuracy can be significantly dependent on the target structure used, we decided to dock all cocrystallized ligands into all representative structures, omitting only the self-docking cases.

To obtain reference ligand positions for the cross-docking evaluation, the following alignment steps were performed. First, structures belonging to the same target were spatially aligned using the protein alignment tool in Maestro 9.[31] In this step, the whole structures were considered. Next, an additional alignment was performed by the align binding sites tool (Maestro 9) using the C-α atoms of the amino acids in 5 Å proximity of the aligned ligands. These amino acids were considered as the binding site, and the resulted ligand positions served as references.

To describe the diversity of binding site geometries and the diversity of ligands belonging to a particular target, we calculated the average binding site rmsd and the average Tanimoto coefficient of ligands. For the rmsd calculation, we used the same method described above. Tanimoto coefficients were calculated using the linear fingerprint provided by ChemAxon.[22] The length of the fingerprint was set to 1024 bit, the maximal linear path contained 7 atoms, and 3 bits were set for representing a given structural feature.

RESULTS AND DISCUSSION

The primary goal of virtual fragment screening is selecting active fragments from a set of decoys. To achieve this goal, fragment docking should provide energetically favorable binding modes that require adequate posing and accurate scoring of docking conformations. We, therefore, tested the docking accuracy of Glide in a fragment setup, and we subsequently analyzed the scoring functions considering their efficacy in binding free energy prediction.

Self-Docking Accuracy. Similar to other docking tools widely used for high-throughput virtual screening, Glide utilizes the rigid receptor approximation. This means that proper binding modes can hardly be identified for ligands binding to alternative receptor conformations. In such cases, the "induced fit" effect might have a significant impact on the docking performance. Docking ligands into their native binding site (self-docking), however, is free from such effects and, therefore, provides an unbiased setup for the evaluation of docking programs and scoring functions. In order to characterize the self-docking capabilities of Glide exhaustively, 190 fragments were docked into their native binding site with different docking protocols. The investigated protocols involved SP and XP runs with both default and expanded samplings optionally combined with pre- and/or postprocessing steps. Fragments and receptor structures were prepared by the automated preparation tools available in the Schrödinger Suite complemented with the pK_a calculator plugin of ChemAxon. Docking accuracy was evaluated in terms of: (i) the average rmsd values calculated between the positions of ligand atoms in the X-ray structure and the docked complex, and (ii) the percentage of accurate poses having an rmsd < 2 Å (success rate).

Glide demonstrated acceptable accuracy in the binding mode prediction of fragments, considering the whole database. Except one protocol (Glide SP with expanded sampling, SP-ES), the average rmsd remained within 2 Å, and the success rate varied between 68−82% (Figure 1 and Table 2). Analyzing the effect of different docking runs, one can conclude that pre- and postprocessing caused significant improvement in both rmsd and success rate. Expanded sampling did not show any significant impact on the accuracy, whereas changing the docking method to Glide XP resulted in a slight decrease in the accuracy (Table 3).

This result is similar to that obtained by Friesner et al. by self-docking drug-like compounds derived from 282 PDB complexes.[32] These authors found that self-docking drug-like compounds by Glide SP 2.5 with preprocessing provided correct binding modes (rmsd < 2 Å) in 71% of the cases. Very recently, comparable results have been reported by three groups on self-docking of small fragment data sets.[16,17,33]

Although Glide XP has a more sophisticated scoring function and uses a more extensive pose generation algorithm,[34] we could not observe any advantage of XP over SP in terms of posing accuracy (Figure 1 and Table 3).

This finding is in line with the observations of Joseph−McCarthy et al. on the prostaglandin D synthase.[33] In the XP scoring function, new energy terms were introduced to

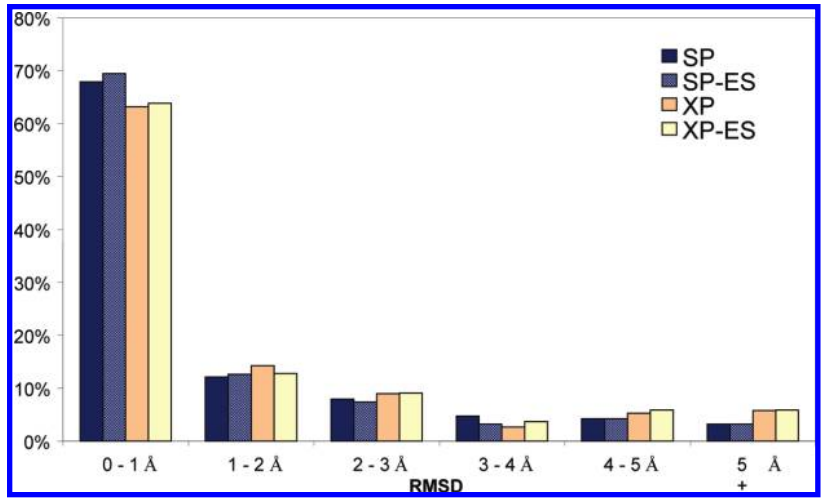

Figure 1. Rmsd distribution of the docking poses generated by SP and XP runs with standard (SP and XP) and expanded sampling (SP-ES and XP-ES) in a self-docking setup. All docking runs included both the pre- and postprocessing steps, corresponding to the default Glide protocols.

Table 2. Fragment Self-Docking Accuracy of the Protocols by Means of Average Rmsd and Success Rate[a]

	without processing		preprocessing	
method	rmsd (Å)	success rate	rmsd (Å)	success rate
SP	1.68	72%	1.43	76%
SP-ES	2.10	69%	1.64	75%
XP	1.78	70%	1.62	71%
XP-ES	1.84	68%	1.67	71%
	postprocessing		double processing	
method	rmsd (Å)	success rate	rmsd (Å)	success rate
SP	1.49	76%	1.17	80%
SP-ES	1.65	73%	1.10	82%
XP	1.61	72%	1.37	77%
XP-ES	1.62	74%	1.51	76%

[a] In default setup, Glide docking includes both pre- and postprocessing steps; these protocols are denoted as "double processing".

estimate the contribution of hydrophobic enclosures, special neutral−neutral single or correlated hydrogen bonds in a hydrophobic environment, and several types of charged−charged hydrogen bonds. Furthermore, new desolvatation and ligand strain energy terms were applied to reduce the number of false positives in virtual screening.[34] Since Glide 4.5, however, the rewarding energy terms are also included in GlideScore SP, and the two scoring functions differ only in the penalty terms. Although this difference can be important in virtual screening situations, our results suggest it is somewhat less important in the identification of correct binding modes of fragments.

The novel expanded sampling option available in Glide 5.5 offers more control over the sampling algorithm of Glide and allows the simple reproduction of the enhanced sampling setup used by Loving et al.[17] We used this setup in Glide 5.0, however, it had a minimal impact on the accuracy, suggesting that default sampling in Glide SP and XP is suitable for pose generation in fragment docking. Contrary to the before-mentioned protocol components, a significant increase in accuracy was observed when both pre- and postprocessing steps were applied. Restricted minimization of the receptor−ligand complexes prior to docking improved the average rmsd by 0.28 Å and the success rate by 4%. The rate of highly accurate poses was increased by 8 and 16% for poses within 1 and 0.5 Å, respectively. Postprocessing minimization had a similar effect on the accuracy (improvement in average rmsd and success rate were 0.28 Å and 5%, respectively), although the rate of highly accurate poses increased moderately (6 and 7% for 1 and 0.5 Å, respectively). This corresponds to the fact that during the preprocessing step protein−fragment complexes come closer to their OPLS force field minima force field minima, which can improve both the rank ordering and the top pose rmsd, whereas the postprocessing step affects primarily the rank order via the relaxation of slightly inaccurate poses. Combining pre- and postprocessing steps, accuracy improvements were found to be additive, yielding Glide SP with the default setup as one of the most accurate protocol (8% increase in success rate and 0.51 Å decrease in the average rmsd compared to the plain SP method).

In addition to impact of protocol components, docking accuracy on the different subsets was also investigated. Interestingly, the docking efficacy of Glide was dependent on the subset of the fragment database docked (Table 4).

Glide showed significantly better performance in terms of the average docking accuracy on the AstexDS subset, indicated by the average rmsd of the poses within 1 Å rmsd and the success rate over 90% in the case of the best protocols. Among the complexes collected from the FBDD literature, the accuracy dropped; average rmsd was 0.8 Å higher, and success rate was 20% lower than those obtained in the AstexDS subset. Investigating the possible reasons, we first compared the physicochemical properties of the different subsets (Table 5).

This analysis showed that the FBDD subset is more lipophilic and that its fragments contain less hydrogen-bond donors and acceptors than the other database parts. These observations suggest the simple explanation of the decreased accuracy. The moderated rate of polar groups can result in the absence of well recognizable interaction motifs, making it more difficult to find the correct binding mode for a docking algorithm. The fact that compounds from the AstexDS subset were less lipophilic and had more hydrogen-bond acceptors and donors seemed to be supporting this hypothesis. Although these trends are clearly identified from

Table 3. Impact of Protocol Components on the Self-Docking Accuracy of Glide[a]

	case	Δrmsd (Å)	Δ < 2 Å	Δ < 1 Å	Δ < 0.5 Å
average effect of preprocessing	SP	−0.40	6%	8%	18%
	XP	−0.17	3%	7%	14%
	average	−0.28	4%	8%	16%
average effect of postprocessing	SP	−0.36	4%	6%	5%
	XP	−0.20	5%	6%	8%
	average	−0.28	5%	6%	7%
average effect of expanded sampling	SP	0.18	−1%	−1%	2%
	XP	0.06	0%	0%	3%
	average	0.12	−1%	0%	3%
average effect of SP−XP change	average	0.10	−3%	−5%	−10%

[a] Improved accuracy is indicated by a decrease in the rmsd and an increase in the success rate.

Table 4. Accuracy of Fragment Self-Docking by Glide on Different Subsets[a]

subset	method	average rmsd (Å)	rmsd < 2 Å	rmsd < 1 Å	rmsd < 0.5 Å
total	SP	1.17	80%	68%	49%
	XP	1.37	77%	63%	42%
	SP-ES	1.10	82%	69%	57%
	XP-ES	1.51	76%	63%	44%
AstexDS	SP	0.65	91%	79%	65%
	XP	1.04	82%	68%	41%
	SP-ES	0.66	91%	79%	71%
	XP-ES	0.88	82%	74%	50%
AffinDB	SP	1.18	81%	70%	50%
	XP	1.17	82%	66%	49%
	SP-ES	1.14	83%	72%	56%
	XP-ES	1.50	78%	66%	50%
FBDD	SP	1.45	72%	58%	38%
	XP	1.87	67%	57%	32%
	SP-ES	1.28	75%	60%	50%
	XP-ES	1.86	68%	53%	30%

[a] Results for SP and XP runs with standard (SP and XP) and expanded sampling (SP-ES and XP-ES) are shown. All docking runs included both the pre- and postprocessing steps.

Table 5. Average Values and Standard Deviations of Physicochemical Properties of Fragments in the Different Subsets of the Database[a]

	MW	ClogP	HBA	HBD	rotbond
AstexDS	236 ± 40	0.5 ± 2.2	3.6 ± 1.7	2.2 ± 1.3	2.3 ± 1.6
AffinDB	202 ± 58	0.8 ± 2.0	3.5 ± 2.1	1.7 ± 1.3	2.2 ± 1.8
FBDD	214 ± 48	1.3 ± 1.5	3.2 ± 1.6	1.5 ± 0.9	2.2 ± 1.3
total	212 ± 54	0.9 ± 1.9	3.4 ± 1.9	1.7 ± 1.2	2.2 ± 1.6

[a] Molecular weight is MW, calculated logP is ClogP, number of hydrogen-bond acceptors is HBA and donors is HBD, and rotatable bonds is rotbond.

Table 5, we could not obtain statistically significant correlation between the physicochemical properties and the rmsd values.

Next, we investigated the localization of the binding sites for all the proteins represented in our fragment database. We found that open binding sites are significantly over-represented in the FBDD subset (relative abundance was 0.33 vs 0.17 for the whole data set). Moreover, fragments binding to open sites are typically more polar and were docked significantly less accurately indicated by the larger rmsd (1.78 Å) compared to that of the whole database (1.17 Å). Therefore, the decreased performance observed on the FBDD subset cannot be rationalized solely by the polarity of the ligands, however, the location of the binding site might also have some impact on the results. More thorough interpretation of docking accuracy in FBDD cases might include comparing targets included in the three different subsets and possibly requires analyzing the misdocked cases individually. Furthermore, results can be biased by complexes that are also included in the training set of Glide. From this point of view, docking fragments into the recently published structures represented in the FBDD subset can be more challenging than the use of structures included in the Astex Diverse Set or in the AffinDB. Such considerations, however, are out of the scope of this study.

Cross-Docking Experiments. Cross-docking studies were performed on a subset of our fragment database, consisting of eight different protein targets binding more than five different fragments to their sequentially identical binding site. These targets included beta-secretase (7 complexes), cAMP-dependent protein kinase (11), cell division protein kinase 2 (7), cytochrome p450cam (7), penicillin amidohydrolase (8), trypsin (10), tryptophan synthase (7), and tyrosine−protein kinase Src (6), representing almost one-third of the total data set (63 complexes). Our cross-docking setup involved docking all of these cocrystallized ligands into all representative structures, omitting only the self-docking cases. Consequently, cross-docking results were collected from 454 docking calculations performed by three protocols (Glide SP, SP-ES, and Glide XP, all with pre- and postprocessing). Results are shown in Table 6. To allow direct comparison of self-docking and cross-docking, statistics (average rmsd and success rate) were also calculated using the corresponding self-docking results. Comparative analysis of docking results revealed that the performance of Glide was somewhat moderate relative to self-docking, as expected. The average rmsd obtained by cross-docking fragments to all available X-ray structures was found to be about 1 Å higher than that of self-docking. Although the average success rates in these cases decreased by 20−25%, an acceptable binding mode (rmsd < 2 Å) could be identified in about 60% of the cases. Focusing our results to the particular X-ray structure having the lowest rmsd in cross-docking for the given target, cross-docking accuracy was significantly improved. The best performing protocol in this case was Glide SP providing the lowest average rmsd of 1.3 Å with minimal gain and the success rate of 77% comparable to that obtained in self-docking (average rmsd = 1 Å, success rate = 86%). These results support that multiple X-ray structures in enrichment studies are worth evaluating before setting up large-scale fragment docking.

Comparing self- and cross-docking results for this subset indicated that both rmsd values and success rates of Glide

Table 6. Fragment Cross-Docking Accuracy of the Protocols by Means of Average Rmsd and Success Rate[a]

target	no. of PDB	average resolution [Å]	avg. site rmsd [Å]	avg. Tanimoto coeff.	protocol	average rmsd [Å]		average success rate		best rmsd [Å]		best success rate	
						cross	self	cross	self	cross	self	cross	self
Beta secretase	7	2.40	0.635	0.411	SP	3.16	1.73	28.6%	71.4%	2.00	1.54	66.7%	83.3%
					SP-ES	2.94	1.35	38.1%	85.7%	2.23	1.47	66.7%	83.3%
					XP	3.02	1.74	35.7%	57.7%	2.68	1.49	50.0%	66.7%
cAMP-dep.protein kinase	11	2.08	1.107	0.340	SP	1.86	0.34	71.4%	100.0%	0.91	0.37	83.3%	100.0%
					SP-ES	3.01	1.79	59.5%	71.4%	1.40	0.66	83.3%	83.3%
					XP	2.02	0.68	66.7%	85.7%	0.96	0.76	83.3%	83.3%
CDC2 kinase	7	2.35	0.873	0.261	SP	2.77	0.67	59.5%	100.0%	0.84	0.68	100.0%	100.0%
					SP-ES	2.96	0.7	47.6%	100.0%	1.68	0.67	83.3%	100.0%
					XP	1.74	0.61	76.2%	100.0%	0.60	0.51	100.0%	100.0%
P450cam	7	1.83	0.412	0.290	SP	2.15	1.66	41.9%	57.1%	1.97	1.35	40.0%	60.0%
					SP-ES	2.05	1.59	42.9%	57.1%	1.50	1.30	60.0%	66.7%
					XP	2.06	0.8	43.3%	85.7%	1.35	0.88	60.0%	83.3%
penicillin amidohydrolase	8	0.40	0.398	0.475	SP	1.55	0.91	75.9%	75.0%	0.86	0.65	85.7%	85.7%
					SP-ES	3.16	0.67	57.1%	87.5%	0.80	0.65	85.7%	85.7%
					XP	1.95	0.74	66.1%	87.5%	1.10	0.72	85.7%	85.7%
trypsin	10	1.65	0.507	0.352	SP	0.97	0.65	90.0%	90.0%	0.84	0.67	88.9%	88.9%
					SP-ES	0.96	0.78	90.0%	90.0%	0.82	0.77	88.9%	88.9%
					XP	1.14	0.83	88.9%	90.0%	0.89	0.72	88.9%	88.9%
tryptophan synthase	7	2.16	0.336	0.439	SP	1.06	1	90.5%	100.0%	0.85	1.00	100.0%	100.0%
					SP-ES	0.92	0.6	92.9%	100.0%	0.58	0.53	100.0%	100.0%
					XP	1.15	0.71	90.5%	100.0%	1.00	0.76	100.0%	100.0%
tyrosine−protein kinase Src	6	1.83	0.445	0.261	SP	2.94	1.66	45.6%	71.1%	2.19	1.70	50.0%	66.7%
					SP-ES	2.84	1.57	40.0%	71.1%	1.99	1.61	50.0%	66.7%
					XP	2.79	2.24	42.8%	71.1%	2.52	2.17	66.7%	83.3%
average on 8 target	63	1.84	0.589	0.354	SP	2.06	1.08	62.9%	83.1%	1.31	1.00	76.8%	85.6%
					SP-ES	2.36	1.13	58.5%	82.9%	1.38	0.96	77.2%	84.3%
					XP	1.98	1.04	63.8%	84.7%	1.39	1.00	79.3%	86.4%

[a] Best rmsd values and best success rates are related to the X-ray structure producing the best performance (lowest rmsd) in cross-docking. All docking runs included pre- and post-processing.

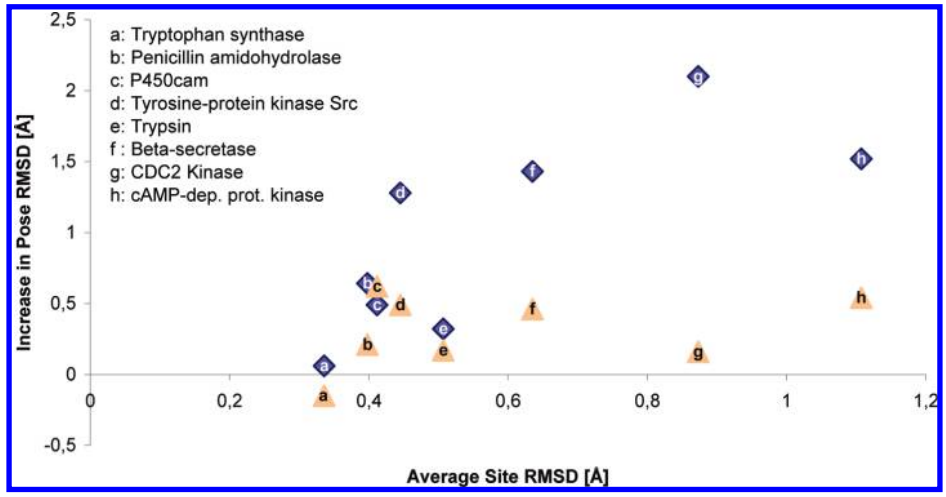

Figure 2. The impact of binding site diversity (measured by the average binding site rmsd) on the cross-docking accuracy of Glide SP for the eight targets. (Both pre- and postprocessing were used.) Squares represent the difference in the average pose rmsd resulted by cross-docking relative to self-docking using all X-ray structures, whereas triangles represent the same value using only the best X-ray structure produced the lowest average cross-docking rmsd for a given target.

SP and XP runs are similar. Average rmsd was around 2 Å, while average success rates were about 63% for both protocols, which were improved with 0.6−0.75 Å and about 15%, respectively, using the optimal X-ray structure of the targets. Considering that Glide XP calculations improved the average performance marginally but are computationally more intensive, we can conclude that the use of SP is more reasonable from a resource point of view. Similar to that found in self-docking expanded sampling did not improve the performance of Glide SP. The docking accuracy of Glide SP was analyzed in terms of the average resolution of the protein structures, average rmsd calculated for the binding site residues, and average Tanimoto-coefficients of fragments docked. Resolution and ligand similarity showed no significant impact on the docking performance. On the other hand, we found that Glide SP typically performs better on targets with less flexible binding sites, represented with X-ray structures having smaller binding site rmsd. Analyzing cross-docking results, we found a clear trend between the average binding site rmsd values and the average pose rmsd shift relative to self-docking (Figure 2). Using the X-ray structures having the lowest average cross-docking pose rmsd, this shift decreases significantly indicating that it is possible to select proper X-ray structures for cross-docking to flexible binding sites. Our result suggests that similar to the case of drug-like molecules binding site flexibility has a significant impact on fragment docking. From a practical point of view, this means that it is reasonable to test multiple protein structures

Table 7. Linear Correlation Coefficients and Rmsd Obtained between Experimental Binding Affinities and GlideScores for the Entire Data Set and Complexes Docked with Rmsd < 2 Å

	SP	XP
R (total)[a]	0.49	0.50
rmsd (total)	2.51	2.71
R (well-docked)[b]	0.55	0.51
rmsd (well-docked)	2.52	2.71

[a] R is the linear correlation coefficient. [b] Well docked denotes rmsd < 2 Å.

in enrichment studies before selecting the protein conformation best suited for virtual fragment screening.

Experimental vs Predicted Binding Free Energy. In addition to ranking accuracy, the precision of GlideScore in binding free energy prediction was evaluated using the complexes with measured affinity data. Eighty-six complexes from AffinDB, 15 from AstexDS, and 1 from the FBDD subsets were included in the calculation, affinities fall in the range of −18.2 to −2.03 kcal/mol (0.5 fM and 33 mM in terms of K_i) with an average value and standard deviation of −7.59 (3 μM) and 2.85 kcal/mol, respectively. Correlation analysis was performed between GlideScore SP and GlideScore XP, produced by the standard SP and XP protocols and the measured binding free energy. Our analysis resulted in weak linear correlation between scores and experimental binding affinity data (Table 7), which is in line with previously published results suggesting that scoring functions can only predict binding affinity with large errors.[35]

It was demonstrated that the XP scoring function was able to reproduce experimental binding affinities for a set of 198 drug-like complexes with standard deviations of 2.26 and 1.73 kcal/mol over all and well-docked ligands, respectively.[34] We obtained a rmsd of 2.71 kcal/mol for the well-docked fragments as well as for the whole database. This result suggests that scoring fragments can be more challenging than scoring drug-like compounds and that more accurate energy terms might be required for fragment-like molecules.

CONCLUSIONS

Fragment-based drug discovery gained significant popularity due to its advantages in the lead discovery phase. Identification of low-affinity fragment hits, however, typically needs biophysical detection techniques, such as X-ray crystallography, nuclear magnetic resonance (NMR) or surface plasmon resonance (SPR) that requires a high upfront investment and a significant screening budget. On the other hand, these biophysical approaches usually have limited throughput. Typical applications screen several hundred to thousand fragments leaving the majority of fragment space unexplored. Virtual screening is a low cost, high-throughput methodology that can be effectively integrated with experimental approaches. Therefore, virtual fragment screening might be an efficient tool to target specific fragment libraries suitable for subsequent experimental screening. In this study we investigated the performance of Glide as a docking tool on a series of protein−fragment complexes collected from different sources [AffinDB, AstexDS, and recent fragment-based drug discovery (FBDD) literature]. Our comprehensive validation study revealed that Glide is well suited for accurate fragment docking and that is one of the conditions to be fulfilled for successful virtual fragment screenings. Out of the 16 different docking protocols tested in self-docking, we found that the default Glide SP protocol combined with expanded sampling showed the best performance. Average root-mean-square deviation (rmsd) on all the 190 complexes was 1.1 Å, and this protocol was able to identify binding modes within 2 Å in 82% of the cases. On the other hand, the gain of expanded sampling to this performance was not significant (average rmsd of 1.1 Å and success rate of 80% was obtained in case of default sampling), and in the context of all protocols, its advantage could not be demonstrated. Interestingly, the Glide XP docking method could not improve the docking accuracy.

A cross-docking study performed on 8 targets and 63 structures confirmed these results. After the selection of proper X-ray structures, Glide yielded comparable results to those of self-docking efficacy producing 1.31 Å average rmsd with 77% success rate in case of default Glide SP, which was the most accurate protocol in the term of the average rmsd.

Glide showed acceptable performance in fragment docking but had some limitations, especially in the docking of ligands from the FBDD subset. Since expanded sampling did not affect these results, we suggest that this limitation is rather due to suboptimal ranking schemes. The lack of acceptable correlation between experimental fragment affinities and GlideScore values gives further support for this hypothesis, because GlideScore is also applied in the pose ranking method used both in Glide SP and Glide XP. One possible option to overcome this limitation would be developing fragment-specific scoring functions. There are two points to consider for designing a fragment-specific scoring function. First, fragments typically have less hydrogen-bond donors and acceptors than drug-like compounds used for the parametrization of currently available scoring schemes. Enthalpic terms associated with the desolvation and the formation of protein interactions might be rather different for fragments and drug-like molecules. Accurate prediction of the fewer specific interactions formed by fragments might have a more significant impact on ranking performance. Second, fragments have significantly less rotatable bonds than drug-like molecules. Rotatable bonds are used for the estimation of entropic contributions in Glide, and therefore, it was hypothesized that entropic effects might be underestimated for fragments.[17] Basically rigid fragments are less subjected to conformational changes upon binding, consequently the penalty applied in Glide for freezing rotatable bonds in fragments would be smaller than that of drug-like compounds. Fragments have more translational and rotational freedom relative to druglike ligands, which suggests that both translational and rotational entropy changes are underestimated during fragment docking. These theoretical considerations could contribute to improving fragment scoring in general, however, the investigation of their impact on docking accuracy requires the detailed analysis of misdocked cases that is in progress.

Although further improvements of the scoring scheme is still possible by considering the above-mentioned aspects, Glide was validated in fragment docking using 190 different protein-fragment complexes. Further evaluation of Glide in large-scale virtual fragment screening is in progress. We think that the protein−fragment database

compiled and used in this study would be a suitable starting point of further detailed investigation and improvement of fragment scoring.

Supporting Information Available: All 190 complexes from AffinDB and AstexDS and from the FBDD literature. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Alex, A. A.; Flocco, M. M. Fragment-based drug discovery: What has it Achieved so Far. *Curr. Top. Med. Chem.* **2007**, *7*, 1544–1567.
(2) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent developments in fragment-based drug discovery. *J. Med. Chem.* **2008**, *51*, 3661–80.
(3) Schulz, M. N.; Hubbard, R. E. Recent progress in fragment-based lead discovery. *Curr. Opin. Pharmacol.* **2009**, *9*, 1–7.
(4) Fink, T.; Bruggesser, H.; Reymond, J. L. Virtual exploration of the small-molecule chemical universe below 160 Da. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 1504–1508.
(5) Chen, Y.; Shoichet, B. K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2009**, *5*, 358–364.
(6) Polgár, T.; Keser"u, G. M. Structure-Based Virtual Screening. *Front. Drug Des. Discovery* **2007**, *3*, 477–502.
(7) Tuccinardi, T. Docking-Based Virtual Screening: Recent Developments. *Comb. Chem. High Throughput Screening* **2009**, *12*, 303–314.
(8) Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discovery* **2002**, *1*, 882–894.
(9) Fara, D. C.; Oprea, T. I.; Prossnitz, E. R.; Bologa, C. G.; Edwards, B. S.; Sklar, L. A. Integration of virtual and physical screening. *Drug Discovery Today: Technol.* **2006**, *3*, 377–385.
(10) English, A. C.; Groom, C. R.; Hubbard, R. E. Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Eng.* **2001**, *14*, 47–59.
(11) Nayal, M.; Honig, B. On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins* **2006**, *63*, 892–906.
(12) Leach, A. R.; Schoichet, B. K.; Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *J. Med. Chem.* **2006**, *49*, 5851–5855.
(13) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
(14) Teotico, D. G.; Babaoglu, K.; Rocklin, G. J.; Ferreira, R. S.; Giannetti, A. M.; Shoichet, B. K. Docking for fragment inhibitors of AmpC β-lactamase. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 7455–460.
(15) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793–806.
(16) Gilles, M.; Didier, R. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Comput. Sci.* **2007**, *47*, 195–207.
(17) Loving, K.; Salam, N. K.; Sherman, W. Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 541–554.
(18) Zhou, Z.; Felts, A. K.; Friesner, R. A.; Levy, R. M. Comparative Performance of Several Flexible Docking Programs and Scoring Functions: Enrichment Studies for a Diverse Set of Pharmaceutically Relevant Targets. *J. Chem. Inf. Model.* **2007**, *47*, 1599–1608.
(19) Block, P.; Sotriffer, C. A.; Dramburg, I.; Klebe, G. AffinDB: a freely accessible database of affinities for protein-ligand complexes from the PDB. *Nucleic Acids Res.* **2006**, *34*, D522–526.
(20) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T.; Mortenson, P. N.; Murray, C. W. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J. Med. Chem.* **2007**, *50*, 726–41.
(21) Murray, C. W.; Rees, D. C. The rise of fragment-based drug discovery. *Nature Chem.* **2009**, *1*, 187–192, and references therein.
(22) *Marvin*, version 5.2; Chemaxon: Budapest, Hungary, 2009.
(23) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'Rule of Three' for fragment-based lead discovery. *Drug Discovery Today* **2003**, *8*, 876–877.
(24) Cambridge Crystallographic Data Centre, Astex Diverse Set; Cambridge Crystallographic Data Centre: Cambridge, U.K.; http://www.ccdc.cam.ac.uk/products/life_sciences/gold/validation/astex_diverse/. Accessed January, 2009.
(25) Worldwide Protein Data Bank, Chemical Component Dictionary; European Bioinformatics Institute: Cambridge, U.K.; http://www.wwpdb.org/ccd.html. Accessed January, 2009.
(26) *LigPrep*, version 2.2; Schrödinger, LLC: New York, NY, 2005.
(27) *Maestro*, version 8.5; Schrödinger, LLC: New York, NY, 2008.
(28) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
(29) *Glide*, version 5.0; Schrödinger, LLC: New York, NY, 2008.
(30) *SiteMap*, version 2.2; Schrödinger, LLC: New York, NY, 2008.
(31) *Maestro*, version 9; Schrödinger: LLC, New York, NY, 2009.
(32) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
(33) Kawatkar, S.; Wang, H.; Czerminski, R.; Joseph-McCarthy, D. Virtual fragment screening: an exploration of various docking and scoring protocols for fragments using Glide. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 527–539.
(34) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
(35) Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. Comparative Assessment of Scoring Functions on a Diverse Test Set. *J. Chem. Inf. Model.* **2009**, *49*, 1079–1093.

CI1000407