

Virtual fragment screening: an exploration of various docking and scoring protocols for fragments using Glide

Sameer Kawatkar · Hongming Wang ·
Ryszard Czerminski · Diane Joseph-McCarthy

Received: 23 March 2009 / Accepted: 7 May 2009 / Published online: 3 June 2009
© Springer Science+Business Media B.V. 2009

Abstract Fragment-based drug discovery approaches allow for a greater coverage of chemical space and generally produce high efficiency ligands. As such, virtual and experimental fragment screening are increasingly being coupled in an effort to identify new leads for specific therapeutic targets. Fragment docking is employed to create target-focussed subset of compounds for testing along side generic fragment libraries. The utility of the program Glide with various scoring schemes for fragment docking is discussed. Fragment docking results for two test cases, prostaglandin D2 synthase and DNA ligase, are presented and compared to experimental screening data. Self-docking, cross-docking, and enrichment studies are performed. For the enrichment runs, experimental data exists indicating that the docking decoys in fact do not inhibit the corresponding enzyme being examined. Results indicate that even for difficult test cases fragment docking can yield enrichments significantly better than random.

Keywords Virtual screening · Structure-based drug design · Enrichment rate · Fragment libraries · Prostaglandin D synthase · DNA ligase

Introduction

Experimental fragment screening is increasingly being used to identify new leads for specific therapeutic targets. In addition to fragment screening companies such as ASTEX [1], a number of pharmaceutical and biotech companies have recently published successful examples (e.g., [2–6]) of its use. The approach is being more widely integrated into the drug discovery process as a complementary approach to high throughput screening (HTS) and virtual screening (VS) for lead generation. Fragment screening is more routinely undertaken in therapeutic areas such as infectious disease, for which HTS has traditionally resulted in low hit rates [7]. For those therapeutic areas in particular, the risks associated with a fragment screening campaign in terms of resource usage are outweighed by the potential gains. Fragment screens require experimental testing of large numbers of compounds either by biochemical assay or a biophysical technique such as NMR, biacore (or surface plasmon resonance), or mass spectrometry at high compound concentrations. These methods can be labor intensive as is the subsequent followup of weak fragment hits. For example, most Medicinal Chemistry teams are unwilling to pursue a high micromolar or low millimolar fragment lead unless an X-ray structure of the fragment bound to the target of interest can be obtained. Even for protein targets with robust crystallographic systems in place, the ability to obtain fragment-protein complex structures is not guaranteed. The fragment hits may bind but not induce a conformational change necessary for crystallization or they may be too weak to be observed in the crystal system.

In general, however, while HTS generates leads for most targets, the leads may have relatively low ligand

Electronic supplementary material The online version of this article (doi:10.1007/s10822-009-9281-4) contains supplementary material, which is available to authorized users.

S. Kawatkar · H. Wang · R. Czerminski ·
D. Joseph-McCarthy (✉)
AstraZeneca, R&D Boston, 35 Gatehouse
Dr., Waltham, MA 02451, USA
e-mail: diane.joseph-mccarthy@astrazeneca.com

efficiencies. The ligand efficiency is commonly defined as the free energy of binding of the ligand to a given target divided by the number of heavy atoms in the ligand [8–11], and as such can be thought of as the average free energy of binding per heavy atom. At the end of the optimization process, in addition to the desired potency levels, the molecular weight and other physical properties need to be within an acceptable range for a candidate drug [12–15]. Typically, as a lead series is optimized, the molecular weight and lipophilicity increase [14]. It can be quite difficult during the optimization process to remove functional groups to transform a low ligand efficiency ligand into a high ligand efficiency ligand. Ligands with high ligand efficiency have the potential for a greater improvement in binding affinity through the addition of a heavy atom or small substituents.

The advantages of a fragment based lead discovery approach are that fragments typically have high ligand efficiencies and that screening fragments allows for a greater coverage of chemical space since chemical space is proportional to the size and complexity of the “small molecules” being considered. Fragments are generally considered to be small molecules with molecular weight less than 300 Da that obey more of a “rule of three” [16] rather than “Lipinski’s rule of five” [12]. To increase the likelihood of success of a fragment screen, in addition to screening general fragment libraries, a virtual screen is often carried out for each target to select a focused subset of fragments to be screened experimentally for that target. Since a fragment screen is generally only carried out for structure-based projects, a structure-based virtual screen is usually performed. Typically molecular docking programs and scoring functions optimized for screening larger small molecules are utilized. The effectiveness of these approaches at accurately docking fragment-like small molecules is less well studied and may be somewhat more challenging [17].

In this paper, the use of the docking program Glide [18, 19] with a number of different scoring schemes [20] for docking fragments is explored. Standard Glide docking protocols with and without MM-GBSA re-scoring, as well as an expanded funnel protocol expected to enhance the sampling of fragment-like molecules binding in a given protein target binding site are examined. Self-docking, cross-docking, and enrichment are investigated for a prostaglandin D synthase (PGDS) fragment dataset [21]. Self-docking refers to docking each ligand back into its native protein structure, while cross-docking involves docking the ligand into the protein structure from a different complex. Enrichment studies including the use of constraints are also carried out for a larger in-house dataset on a bacterial NAD⁺ dependent DNA ligase (ligase) target.

Methods

PGDS test set

PGDS carries out the isomerization of prostaglandin H₂ to prostaglandin D₂, an allergic and inflammatory mediator, in the presence of the cofactors glutathione (GSH) and Mg²⁺. Hohwy et al. [21] previously published a fragment screening study in which ~2,000 fragments were screened by NMR spectroscopy resulting in 24 confirmed binders. Of the 24 hits, X-ray complex structures were solved for six bound to PGDS (PDB ids: 2VCQ, 2VCW, 2VCX, 2VCZ, 2VD0, and 2VD1) with resolutions ranging from 1.95 to 2.25 Å. The six ligands are fragment-like with molecular weight (MW) less than 350 Da (Fig. 1). There is also a seventh publicly available X-ray complex structure of PGDS with a slightly larger ligand (2CVD, see Fig. 1) bound with resolution 1.45 Å [22]. This set of seven structures along with the screening data for the entire library was chosen as a test set for this docking and enrichment study.

Ligase test set

Ligase is a multi-domain protein that catalyzes DNA joining during replication and repair similar to the eukaryote ATP-dependent ligases [23]. The residues involved in NAD⁺ binding are largely conserved across different bacterial species. An enrichment study was carried out by screening an AstraZeneca proprietary 20 K fragment library developed by Blomberg and coworkers (and described in a separate article in this issue of JCAMD, 2009) against the adenosine-binding site of the enzyme. Virtual screening data was compared to in-house experimental screening data.

Protein structure preparation

Of the seven PGDS complex structures considered in this study, six of the protein structures are very similar (maximum pairwise α -carbon root-mean-squared deviation (RMSD) among the set is 0.33 Å), while the seventh, 2CVD, with the larger ligand bound is significantly different in the active site (minimum pairwise α -carbon RMSD from the other six protein structure is 1.46 Å; see Fig. 2). In the 2CVD structure, Trp 104, for e.g., undergoes a significant conformational change relative to its position in the other structures to accommodate the larger ligand. All seven of the structures have the GSH cofactor bound in a similar position near the active site. The Mg²⁺ ion cofactor is observed in three of the complex structures (2CVD, 2VCX, and 2VD1). All were crystallized in the presence of high concentrations of MgCl₂ (2 mM for

Fig. 1 Inhibitors bound to PGDS in the X-ray complex structures used in the present study. Molecular weight in daltons is given in *parenthesis*

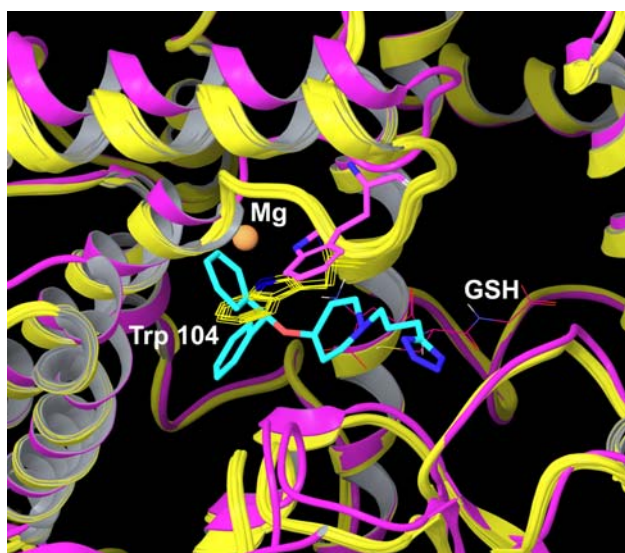
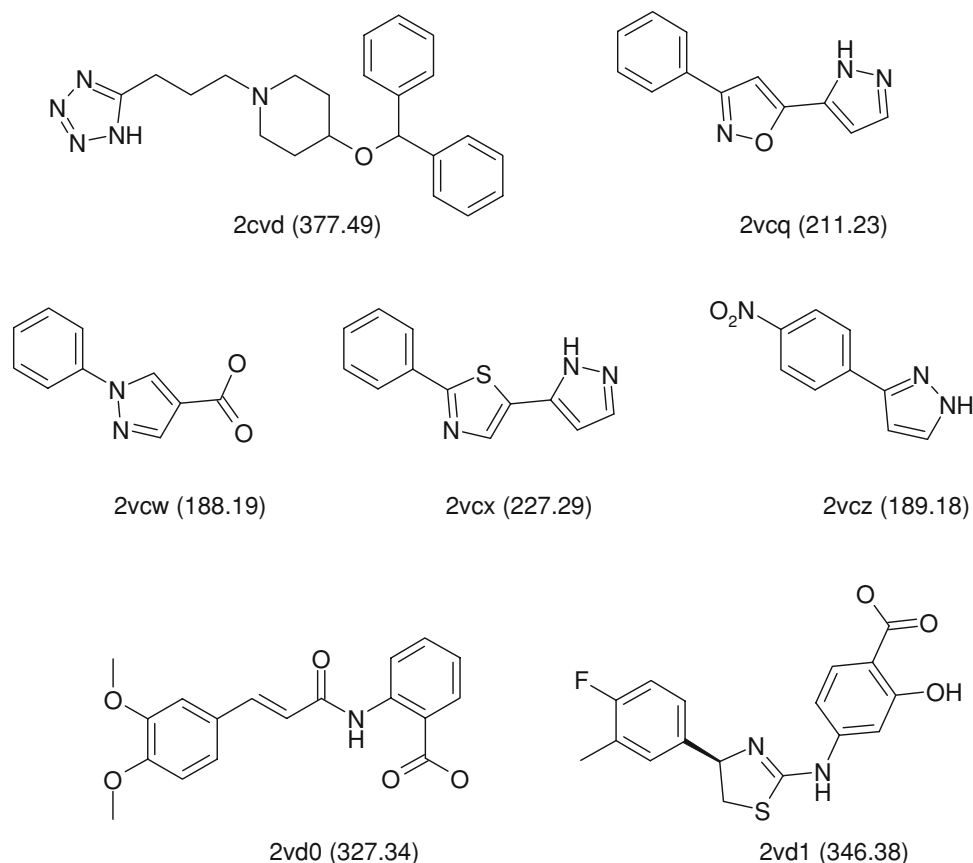


Fig. 2 Superposition of seven PGDS X-ray complex structures. Ribbon representations for 2VCQ, 2VCW, 2VCX, 2VCZ, 2VD0, and 2VD1 are shown in yellow and 2CVD is in magenta. GSH and Mg²⁺ from 2VD1 structure are shown colored by element. The 2CVD ligand is shown in *thick lines* colored by element with carbons in cyan

2CVD and 5 mM for the other six). For the docking study, the Mg²⁺ was retained in the three structures in which it was observed.

All structures were prepared using the Maestro 8.5 protein preparation wizard (Schrodinger, LLC, 2008, New York, NY); waters were deleted, bond orders assigned, hydrogens added, and metals treated appropriately when present. Next, the orientation of hydroxyl groups, amide groups of Asn and Gln, and charge state of His residues were optimized. Finally, a restrained minimization of the protein structure was performed using the default constraint of 0.30 Å RMSD and the OPLS 2001 force field.

The in-house determined X-ray structure of *S. pneumoniae* ligase was similarly prepared.

Virtual screening library preparation

Virtual libraries of (1) the seven ligands in the PGDS complex structures, (2) 1,847 of the fragments screened for binding to PDGS [21], and (3) the 20 K fragment library [24] screened for ligase (with 19,299 compounds actually screened) were generated following the same procedure. Smiles representation [25] of the ligands were input to Leatherface [26], an in-house molecular editor based on the OEChem toolkit (OpenEye Scientific Software, Inc., 2008, Santa Fe, NM) that was used to generate protonation and tautomeric states. Three dimensional (3D) coordinates were then generated for all ligands using the Ligprep utility in

Maestro8.5 (Schrodinger, LLC, 2008, New York, NY). Leatherface protonation and tautomeric states and any specified chiralities were retained. One low energy ring conformation was generated per ligand.

Docking scoring grid preparation

Prepared protein structures were used to generate Glide scoring grids for the subsequent docking calculations. For each PGDS protein structure, a grid box of default size ($20 \times 20 \times 20 \text{ \AA}^3$) was centered on the corresponding ligand position. Default parameters were used and no constraints were included during grid generation.

For the ligase protein structure, the scoring grid was generated using a box size of $24 \times 24 \times 24 \text{ \AA}^3$ and the ligand range was defined using a box of size of $12 \times 12 \times 12 \text{ \AA}^3$. Hydrogen bond constraints were also included during the grid generation (see Fig. 3).

Docking and scoring protocols

Four scoring protocols were used for the docking: GlideSP, GlideXP (SP and XP options with default settings), and GlideSP-EF and GlideXP-EF (SP and XP options with the “expanded funnel” described below). In general, default parameters were used for the docking runs unless otherwise specified. For each PGDS docking run, the top nine poses based on the Glide docking score were saved for each ligand; this limit was chosen since increasing the number

of saved poses to ten or more turns off degeneracy checking (RMSD within 0.5 \AA) and may result in a saved set of nearly identical poses. For the ligase enrichment run, one pose was saved per ligand.

The Glide method is described as docking funnel that uses a series of filters to sample the protein binding site and search for acceptable poses [18]. In the flexible docking mode, Glide generates a set of conformers for each input ligand and then performs an exhaustive search for possible positions and orientations of ligand over the active site. The ligand poses that Glide generates pass through a series of hierarchical filters that evaluate the interaction of the ligand with the receptor. Poses that pass these initial screens are subjected to energy minimization on precomputed van der Waals and electrostatic grids for the receptor. Final scoring is then carried out on the energy-minimized poses. If GlideScore is selected as the final scoring function (the default), the composite scoring function, Emodel, is used to rank the ligand poses and to select which pose for a given ligand will be output to the user. Emodel combines GlideScore with the nonbonded interaction energy between the ligand and the protein and, for flexible ligand docking, the ligand strain energy. Typically the GlideScore is used to rank docked poses of different ligands and the top ranked pose based on the GlideScore is also the lowest Emodel pose. However, for certain ligands, the lowest GlideScore pose does not correspond to the lowest Emodel pose. For those ligands, if only one pose is saved per ligand for the docking run, the lowest Emodel pose is retained.

The way poses pass through the filters for the initial geometric and complementarity fit between the ligand and receptor molecules can be modified in the Settings tab of the Ligand Docking module using the Advanced Settings option. This section has three settings that control the selection of initial poses that pass through the initial Glide filters: (1) the number n of poses per ligand that are kept for the initial phase of docking (the grid refinement calculation); the default is 5,000 for flexible docking. (2) The “scoring window” for retaining initial poses which sets the rough-score cutoff relative to the best rough score found; the default requires that a pose be within 100.0 kcal/mol of the best rough-score pose to survive. (3) The top-scoring m poses per ligand retained for energy minimization on the receptor grid; the default is 400 for SP and 800 for XP. It is possible that the initial screening approach described above may miss key conformations of the ligand such that an acceptable pose would be rejected before the energy minimization stage in the docking funnel. With the “expanded funnel” protocol mentioned above, the sampling is increased by setting n to 50,000, the scoring window to 500 kcal/mol, and m to 1,000.

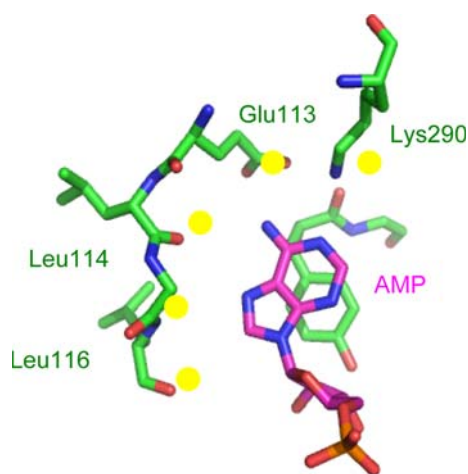


Fig. 3 Structure (PDB id: 1OWO) of the adenosine monophosphate (AMP) binding site in ligase. Residues are colored by element with carbons in *green* for Ligase and *magenta* for AMP. Atoms used as hydrogen bond docking constraints are indicated by *yellow dots*. The three “hinge-like” hydrogen bond constraints by analogy to kinase structures are to Leu114 (backbone carbonyl oxygen) and Leu116 (backbone carbonyl oxygen and backbone nitrogen)

MM-GBSA re-scoring of docked poses

Molecular mechanics generalized born surface area (MM-GBSA) approaches can be applied as a way of including implicit solvation into the estimation of the free energy of ligand binding. In this study, MM-GBSA calculations were carried out using the `prime_mmgbsa` utility (Schrodinger, LLC, 2008, New York, NY) [27]. As a post-docking step, docked ligand poses generated with GlideSP were re-scored using the MM-GBSA script in two modes, respectively, (1) with a conformationally rigid protein structure and (2) with a partially flexible protein structure. The flexible region was defined as any residue with an atom within 12 Å of the ligand in the 2VD1 structure, and during the relaxation of the protein–ligand complex this portion of the protein was allowed to move along with the ligand.

Analysis of docking runs

For the PGDS self-docking and cross-docking runs, the RMSD of each pose from the X-ray structure position of the corresponding ligand bound to PGDS was calculated using the RMSD python script in the OECHEM toolkit (OpenEye Scientific Software, Inc., 2008, Santa Fe, NM). Self-docking refers to docking a ligand back into its native protein structure, while cross-docking involves docking the ligand back into one of the other six protein structures examined. A docked pose was considered correct if it was within 2 Å of the X-ray complex structure position. For the PGDS and ligase enrichment studies, enrichment plots (% actives identified versus % ranked database virtually screened) were generated and for ligase enrichment factors at 1% of the ranked database were calculated as in Chen et al. [28] as:

$$\text{Enrichment Factor (EF)} = \text{Hits}_{\text{sel}} / \text{Hits}_{\text{tot}} \times \text{NC}_{\text{tot}} / \text{NC}$$

where Hits_{sel} is the number of actives selected by the docking at the specified X% of the ranked database, Hits_{tot} is the total number of actives, NC_{tot} is the total number of molecules in the screened database, and NC is the number of compounds in X% of the ranked database. For the PGDS and ligase studies, receiver operating characteristic (ROC) plots [29, 30] were also generated; the y-axis of each ROC plot is the number of actives identified divided by the total number of actives (% actives) and the x-axis is the number of inactives virtually screened divided by the total number of inactives (the library size minus number of actives; % inactive database).

For ligase, the virtual screening hit rate was calculated using the number of confirmed actives (with measurable IC50s) present in the top 1,000 compounds of the ranked database. The experimental hit rate was 4%, which was calculated by dividing the 794 actives with an IC50 < 1 mM by the 19,299 library fragments screened.

Results and discussion

Fragment docking results were compared to experimental screening data for two test cases, PGDS and ligase. For the PGDS system, self-docking and cross-docking with the seven X-ray complex structures described above were performed and the enrichment of the 24 NMR binders over the set of ~2,000 fragments in the NMR screening library was investigated. For the self-docking (Fig. 4; Table 1), with GlideSP five of the seven ligands are correctly docked back into their native protein structures. More specifically, the top pose correctly predicted the ligand position for 2VCQ, 2VCW, 2VCX, 2VCZ, and 2VD0 with RMSDs less than 2 Å (Table 1, for 2VD0, see Fig. 6a), while for the 2CVD and 2VD1 ligands, the top pose had an RMSD of 2.9 and 7.64 Å, respectively. For 2CVD, the largest portion of the molecule which contains the two phenyl rings is correctly positioned (Fig. 6b). The piperidine ring is in a twist boat conformation in the X-ray structure and in a lower energy chair conformation in the top pose; however, if multiple ring conformations are generated with LigPrep and the sampling of ring conformations is turned off during the docking an overall similar docked pose is obtained with the ring in a twist boat. In the X-ray position the tetrazole points out into bulk solvent, while in the top docked pose, the tetrazole bends back in toward the protein and forms a hydrogen bond with side chain of Gln 36, resulting in the overall RMSD of 2.9 Å (Fig. 6b). The ligands (as well as Trp 104) are unambiguously positioned in electron density in 2Fo-Fc maps for all of the structures. In the 2CVD structure, the tetrazole position is likely due to crystal packing effects. For the 2VD1 case (Fig. 6c), the docked pose for the ligand is largely out of the pocket. In the X-ray structure, the fluorophenyl of the ligand is deep in the hydrophobic pocket defined by Met 99, Trp 104, Ile 155, and the side chain of Arg 14, and the benzoic acid interacts with the side chain of Gln 36. In the top GlideSP pose, the acid moiety interacts with the side chains of Lys 198 and Gln 109; it is possible that the scoring scheme is over-weighting these electrostatic interactions.

Overall, GlideSP performs better than GlideXP. The expanded funnel option increases the compute time by 65–95% (over GlideSP) but does not improve the docking accuracy. In fact, GlideSP-EF, with the increased ligand sampling, selected a flipped pose as top pose for 2VD0 with an RMSD of 8.61 Å (Fig. 6d; Table 2). The position of 2VD0 was correctly docked by docked by GlideSP. The 2VD0 ligand, however, is the most symmetric of the seven ligands (Fig. 1) and the second ranked pose generated with the GlideSP-EF protocol was correct and was nearly iso-energetic with the top ranked pose (GlideScores of –8.21 vs. –8.1 kcal/mol, but Emodel scores of –67.47 vs. –67.72 kcal/mol). If, for GlideSP, the best RMSD pose

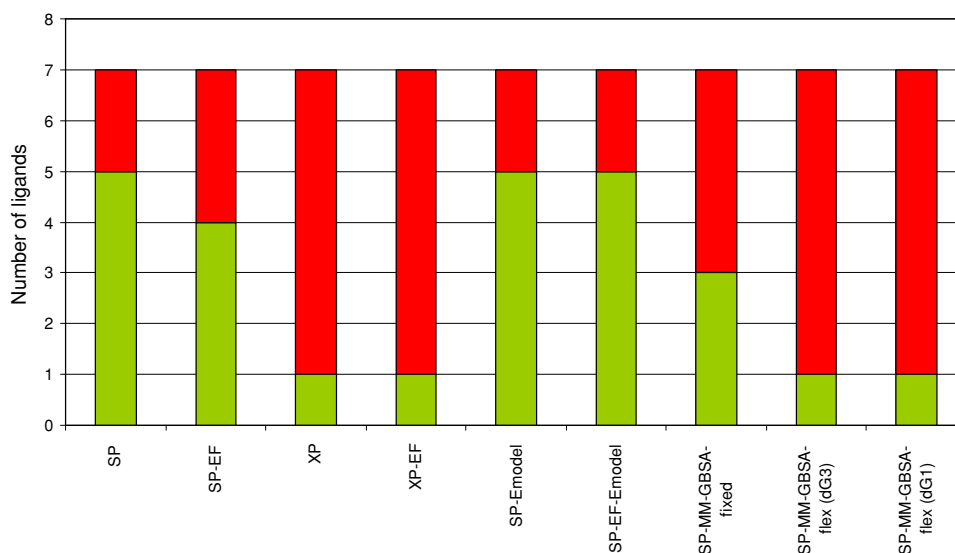


Fig. 4 Histogram of the number of ligands with RMSD value for the top pose ≤ 2 Å (green) and ligands with RMSD > 2 Å (red) for each self-docking run. Results for GlideSP, GlideSP-EF, GlideXP, GlideXP-EF, GlideSP with Emodel ranking, MM-GBSA re-scoring with the protein fixed, and MM-GBSA re-scoring with a partially flexible protein are shown. For the MM-GBSA with the flexible protein two

different energy terms were used, respectively, for the ranking, $\Delta G3$ which includes the ligand strain ($\Delta G3 = E_{\text{complex}}(\text{minimized}) - (E_{\text{ligand}}(\text{minimized}) + E_{\text{receptor}}(\text{from minimized complex}))$), and $\Delta G1$ which includes the ligand and the protein strain ($\Delta G1 = E_{\text{complex}}(\text{minimized}) - (E_{\text{ligand}}(\text{minimized}) + E_{\text{receptor}}(\text{minimized}))$)

Table 1 Self-docking for PGDS dataset using GlideSP

Ligand	Top pose		First 'correct' pose ^a			Lowest RMSD pose		
	RMSD	Docking score	RMSD	Rank	Docking score	RMSD	Rank	Docking score
2CVD	2.9	-7.96	-	-	-	2.06	9	-6.63
2VCQ	0.57	-6.29	0.57	1	-6.29	0.57	1	-6.29
2VCW	0.20	-7.07	0.20	1	-7.07	0.20	1	-7.07
2VCX	0.13	-7.00	0.13	1	-7.00	0.13	1	-7.00
2VCZ	1.17	-6.53	1.17	1	-6.53	0.87	6	-6.25
2VD0	0.64	-8.15	0.64	1	-8.15	0.45	3	-7.97
2VD1	7.64	-5.87	-	-	-	5.49	4	-5.74

^a Lowest ranking pose with RMSD ≤ 2 Å

Table 2 Self-docking for PGDS dataset using GlideSP-EF

Ligand	Top pose		First 'correct' pose ^a			Lowest RMSD pose		
	RMSD	Docking score	RMSD	Rank	Docking score	RMSD	Rank	Docking score
2CVD	2.91	-7.94	-	-	-	2.74	2	-7.69
2VCQ	0.57	-6.24	0.57	1	-6.24	0.57	1	-6.24
2VCW	0.21	-7.07	0.21	1	-7.07	0.21	1	-7.07
2VCX	0.96	-7.01	0.96	1	-7.01	0.12	2	-7.00
2VCZ	1.16	-6.56	1.16	1	-6.56	0.79	6	-6.26
2VD0	8.61	-8.21	0.62	2	-8.10	0.43	3	-7.97
2VD1	7.64	-5.85	-	-	-	5.27	7	-5.70

^a Lowest ranking pose with RMSD ≤ 2 Å

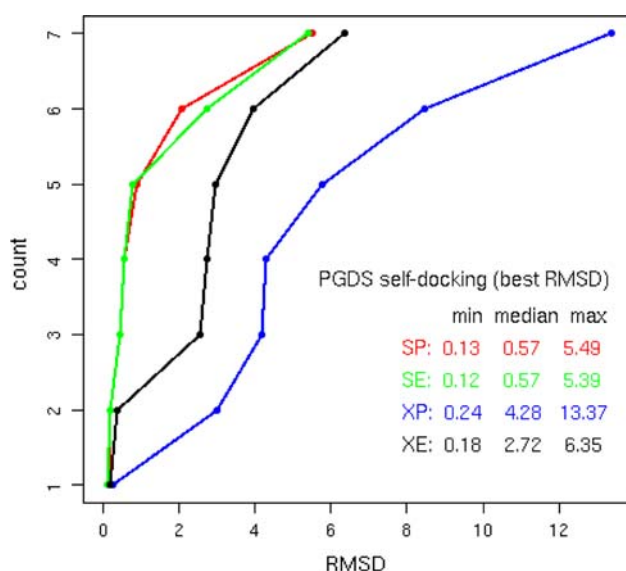
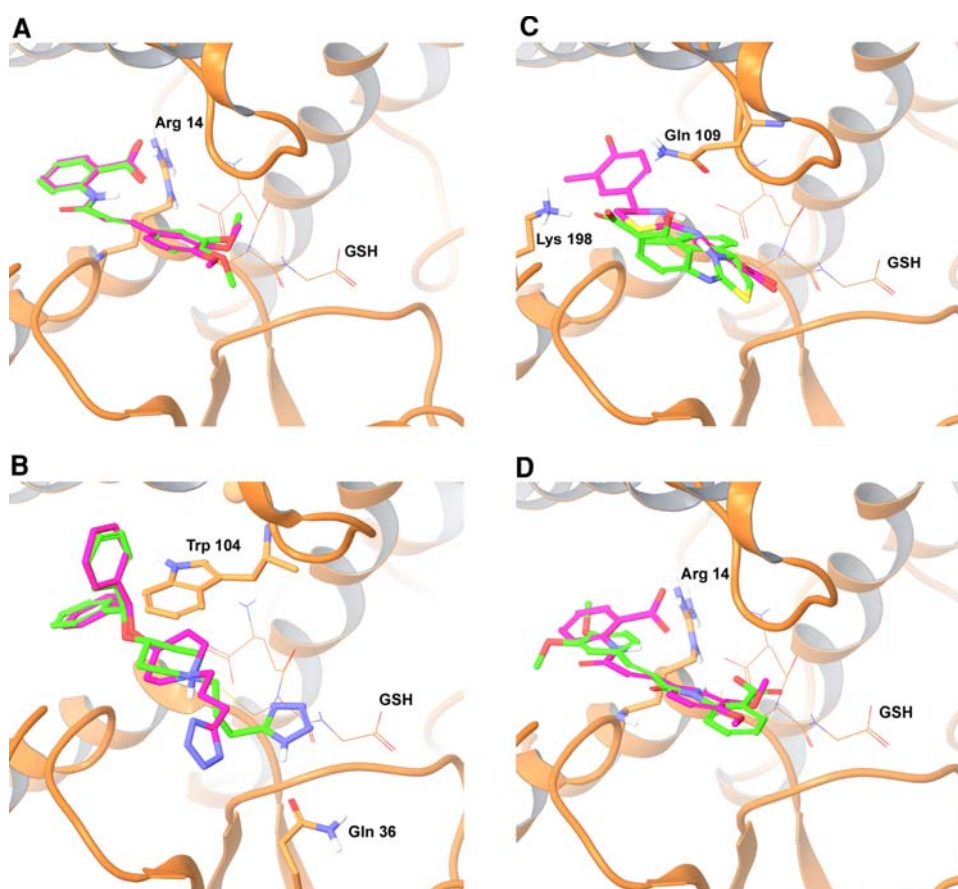


Fig. 5 A plot for each self-docking protocol tested of the best RMSD per ligand versus number of ligands. The RMSD is for the best pose from the up to nine saved per ligand from each docking run. Results are plotted for the GlideSP, GlideSP-EF, GlideXP, and GlideXP-EF protocols

(out of the nine saved per ligand) is examined versus the top ranked pose, the results improve somewhat but the general trends are the same (Fig. 5; Table 1); SP scoring

Fig. 6 Examples of top poses identified in self-docking runs. In each case the X-ray complex structure is shown colored by element with protein carbons in orange and ligand carbons in magenta, while the docked ligand pose is colored by element with carbon in green. Residues that interact with the ligand are shown in thick lines while the GSH cofactor is in thin lines. In **a**, the top GlideSP pose for 2VD0 ligand (RMSD = 0.64 Å) is shown. In **b**, the top GlideSP pose for 2CVD ligand (RMSD = 2.9 Å) with a portion of the ligand incorrectly docked is shown. In **c**, the top GlideSP pose for 2VD1 (RMSD = 7.64 Å) with the ligand largely popped out of the pocket is shown. In **d**, the top GlideSP-EF pose for 2VD0 ligand (RMSD = 8.61 Å) with the ligand flipped is shown



performs better than XP and the EF sampling does not significantly improve the results.

A comparison of GlideSP with GlideScore ranking of poses and GlideSP with Emodel ranking shows no significant differences (Fig. 4). What is surprising is that the MM-GBSA re-scoring of the nine docked poses per ligand markedly decreases the docking accuracy with the fixed protein and with the partially flexible protein, respectively. In fact with the flexible protein, the decrease in docking accuracy is even more pronounced (Fig. 4).

For the cross-dockings of the seven ligands into each of the other protein structures, the overall docking accuracy is somewhat reduced over self docking as expected, but the same trends with the scoring schemes are observed. Using the 2VD1 protein structure and GlideSP, four out of seven ligands are correctly docked (Fig. 7). Specifically, using GlideSP, the top poses for 2VCQ, 2VCW, 2VCX, and 2VCZ are correct, whereas the top poses for 2CVD and 2VD0 have RMSD of 9.57 and 2.67 Å, respectively (Table 3). For 2CVD, the ligand is flipped because the 2VD1 protein structure cannot accommodate the larger ligand in the correct position. Attempts to use the Induced Fit Docking protocol [31] within Maestro 8.5 also failed to generate a correct pose for the 2CVD ligand in the other protein structures. For 2VD0, the acid moiety in the docked

Table 3 Cross-docking with 2VD1 grid using GlideSP

Ligand	Top pose		First 'correct' pose ^a			Lowest RMSD pose		
	RMSD	Docking score	RMSD	Rank	Docking score	RMSD	Rank	Docking score
2CVD	9.57	−5.20	–	–	–	6.19	7	−3.48
2VCQ	0.77	−6.67	0.77	1	−6.67	0.77	1	−6.67
2VCW	0.43	−7.01	0.43	1	−7.01	0.43	1	−7.01
2VCX	1.05	−6.93	1.05	1	0.48	0.48	2	−6.82
2VCZ	1.51	−6.41	1.51	1	1.14	1.14	3	−6.31
2VD0	2.67	−7.42	0.88	2	−7.37	0.66	4	−7.24

^a Lowest ranking pose with RMSD ≤ 2 Å

Table 4 Cross-docking runs in 2VD1 grid using GlideSP-EF

Ligand	Top pose		First 'correct' pose ^a			Lowest RMSD pose		
	RMSD	Docking score	RMSD	Rank	Docking score	RMSD	Rank	Docking score
2CVD	9.62	−5.52	–	–	–	9.62	1	−5.52
2VCQ	0.77	−6.67	0.77	1	−6.67	0.77	1	−6.67
2VCW	0.43	−7.01	0.43	1	−7.01	0.43	1	−7.01
2VCX	1.05	−6.93	1.05	1	−6.93	0.48	2	−6.83
2VCZ	1.51	−6.41	1.51	1	−6.41	1.14	3	−6.32
2VD0	0.90	−7.47	0.90	1	−7.47	0.71	4	−7.28

^a Lowest ranking pose with RMSD ≤ 2 Å

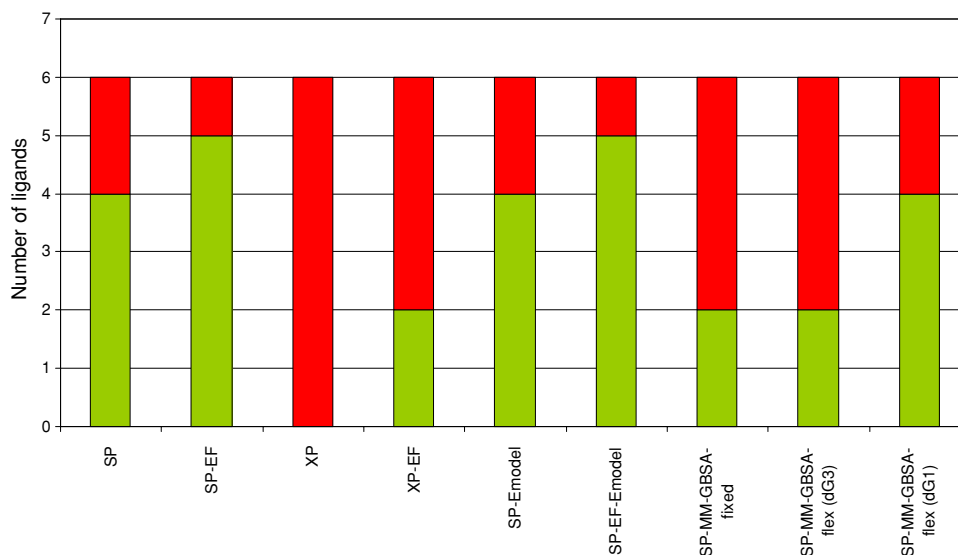


Fig. 7 Histogram of the number of ligands with RMSD value for the top pose ≤ 2 Å (green) and ligands with RMSD > 2 Å (red) for each cross-docking run. Results for GlideSP, GlideSP-EF, GlideXP, GlideXP-EF, GlideSP with Emodel ranking, MM-GBSA re-scoring with the protein fixed, and MM-GBSA re-scoring with a partially flexible protein are shown. For the MM-GBSA with the flexible

protein two different energy terms were used, respectively, for the ranking, $\Delta G3$ which includes the ligand strain ($\Delta G3 = E_{\text{complex}}(\text{minimized}) - (E_{\text{ligand}}(\text{minimized}) + E_{\text{receptor}}(\text{from minimized complex}))$), and $\Delta G1$ which includes the ligand and the protein strain ($\Delta G1 = E_{\text{complex}}(\text{minimized}) - (E_{\text{ligand}}(\text{minimized}) + E_{\text{receptor}}(\text{minimized}))$)

pose is positioned incorrectly. In this case, for cross-docking with the 2VD1 protein structure, GlideSP-EF did correctly dock the 2VD0 ligand with an RMSD of 0.9 Å (Table 4), resulting in slightly improved performance over

GlideSP. Overall SP performs better than XP and the expanded funnel does not dramatically improve the results. Again, for the cross docking GlideSP with GlideScore versus Emodel ranking did not significantly change the

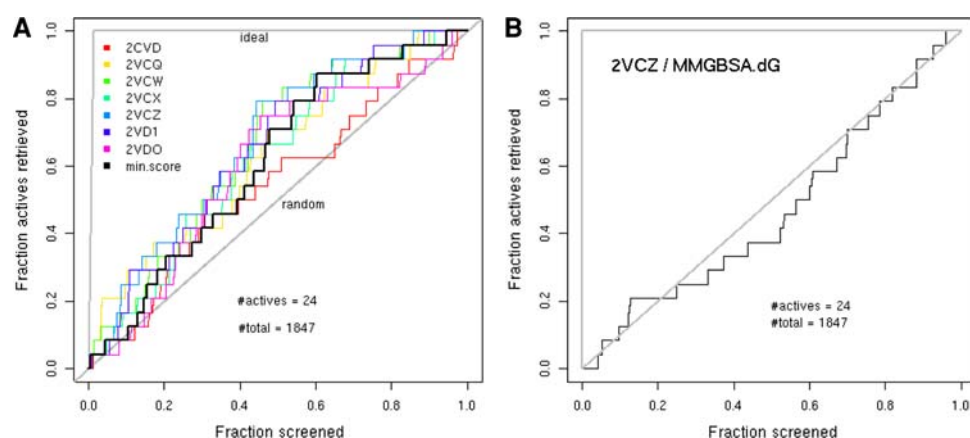


Fig. 8 In **a** enrichment of the 24 PGDS actives out of the ~2,000 fragment generic NMR screening library plotted for each of the docking runs corresponding to the seven PGDS protein structures (from 2CVD, 2VCQ, 2VCW, 2VCX, 2VCZ, 2VD1, 2VDO), respectively. The *heavy black curve* corresponds to enrichment obtained by using the minimum score for each ligand in any one of the seven

protein structures for the ranking. GlideSP was used for the docking. In **b** enrichment after MM-GBSA re-scoring of the 2VCZ (*dark blue curve* in **a**) poses (up to nine per ligand) with the protein held fixed. For both **a** and **b**, % ranked database is plotted versus % actives identified and the *diagonal line* represents the enrichment due to random screening

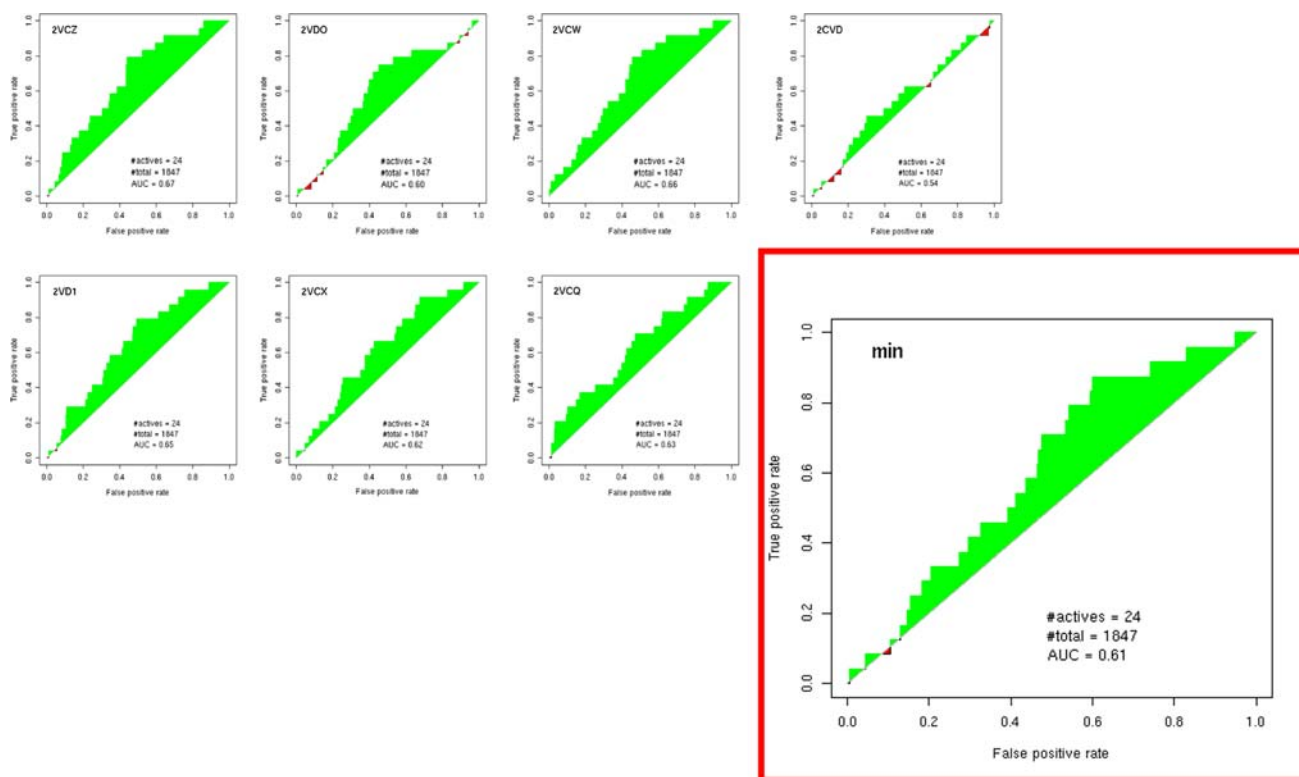


Fig. 9 ROC plots of % inactives in the ranked database versus % actives identified for the PGDS dataset of 24 actives out of the ~2,000 fragment generic NMR screening library. These plots (**a–g**) represent the enrichment from each docking run, respectively,

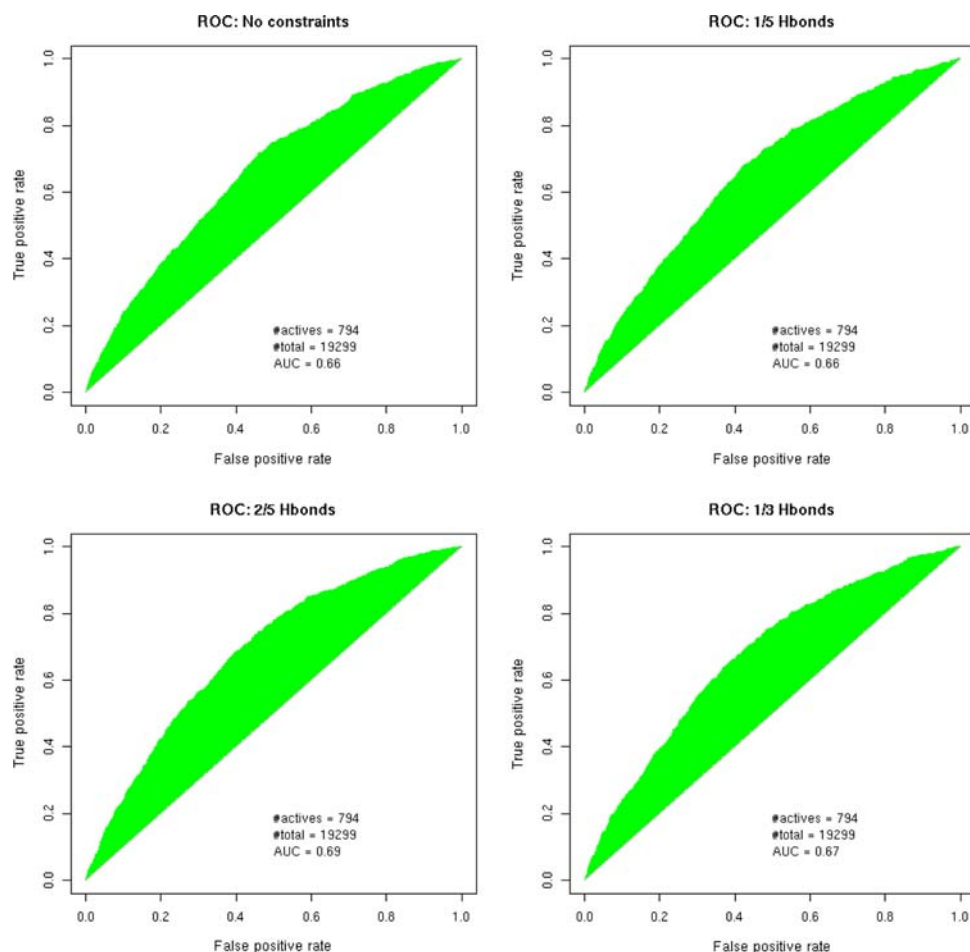
screening with one of the seven PGDS protein structures (from 2CVD, 2VCQ, 2VCW, 2VCX, 2VCZ, 2VD1, 2VDO). In **h**, the ROC plot calculated using the minimum score for each ligand in any one of the seven protein structures for the ranking is shown

results and the MM-GBSA re-scoring of the docked poses either decreased or did not improve the docking accuracy. Cross-docking results using the other six protein structures are similar (see Supplementary material) except that as anticipated using the 2CVD protein structure (the most

different of the seven protein structures) yielded the poorest results with none of the other six ligands correctly docked using GlideSP.

The enrichment of the 24 NMR binders (including the seven ligands bound in the X-ray complex structures and

Fig. 10 ROC plots of % inactives in the ranked database versus % actives for the ligase test case for actives with $IC_{50}s \leq 1$ mM. The enrichment is represented in **a** for the no constraint run, in **b** for the run requiring that 1/5 hydrogen bond constraints (depicted in Fig. 3) be satisfied, in **c** for the run requiring that 2/5 hydrogen bond constraints be satisfied, and in **d** for the run requiring that 1/3 “hinge-like” hydrogen bonds be satisfied



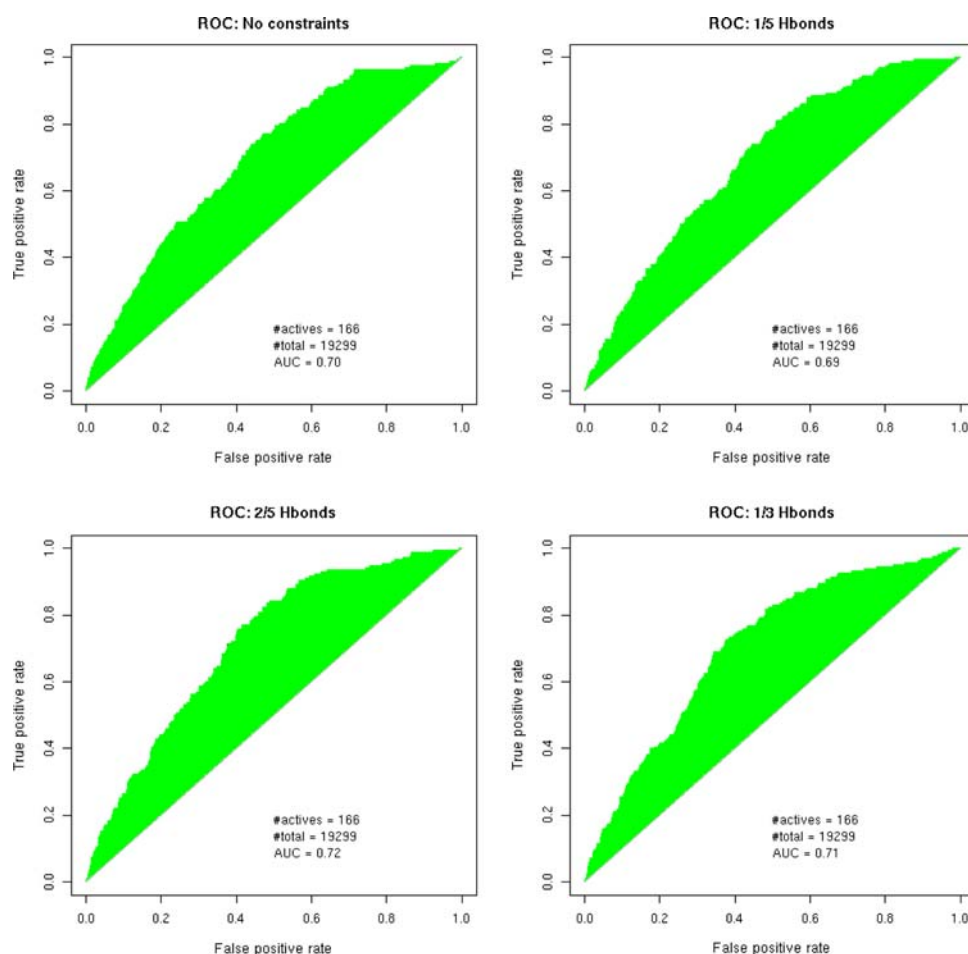
shown in Fig. 1) was investigated for virtual screening with the different PGDS protein structures. In Fig. 8, the enrichment of the PGDS actives as a fraction of the ranked database of screened fragments is plotted for the docking runs using each of the seven protein structures, respectively. In these standard enrichment plots, the diagonal represents the expected hit-rate based on a random ranking of compounds and points that fall above the line display the enrichment provided by the docking calculation or the ability of the docking run to identify more active compounds than random screening would. As might have been expected, the virtual screen using the 2CVD protein structure, the one that was the most different of the seven, gave the poorest enrichment. Somewhat surprisingly again, for the 2VCZ structure which produced one of the best enrichments based on the GlideScore (Fig. 8a), the MM-GBSA re-scoring of the three saved poses per fragment eliminated that enrichment (Fig. 8b).

ROC curves plotting the percent of inactives in the ranked database against the percent of actives are shown in Fig. 9 for the seven docking runs; ROC plots are normalized enrichment plots and in this case the two types of plots are very similar. Also shown in Fig. 9h, is a ROC plot

calculated using the minimum score (the best score) for each fragment in any of the seven protein structures to rank the database. The area under the curve (AUC) for the minimum score plot (0.61) is the average of the worst AUC (0.54 for the 2CVD run) and best AUC (0.67 for the 2VCZ run) over the whole set of structures. Thus, the minimum score plot suggests that for a given target, if a set of structures were available and there was no pre-existing data to indicate that one structure would be preferred to the others for docking, ranking the database by the minimum score in any one of the structures would be a reasonable approach to getting enrichment.

The second fragment docking test case presented herein is for ligase. The 20 K generic fragment library described above in the “Methods” Section was screened using a high concentration biochemical assay and 794 actives with $IC_{50}s < 1$ mM were identified (Adam Shapiro, unpublished data). Docked poses were generated using Glide SP and various sets of hydrogen bond constraints in the AMP binding site. The effect of the docking constraints on the enrichment was investigated. Five hydrogen bonds are made between AMP and the protein in the adenosine binding pocket of ligase (Fig. 3). The effect of requiring

Fig. 11 ROC plots of % inactives in the ranked database versus % actives for the ligase test case for actives with $IC_{50}s \leq 100 \mu M$. The enrichment is represented in **a** for the no constraint run, in **b** for the run requiring that 1/5 hydrogen bond constraints (depicted in Fig. 3) be satisfied, in **c** for the run requiring that 2/5 hydrogen bond constraints be satisfied, and in **d** for the run requiring that 1/3 “hinge-like” hydrogen bonds be satisfied



that 1/5, 2/5, or 1/3 “hinge-like” hydrogen bonds (by analogy to kinase structures) be satisfied by the docking poses was studied. ROC plots for the no constraints, 1/5 hydrogen bonds, 2/5, and 1/3 kinase-like hydrogen bonds constraint runs are shown in Fig. 10. Each of the four docking runs show some enrichment over random and the AUCs are very similar for each plot. Somewhat surprisingly the effect of the constraints in this case is minimal. This could either be because the docking without constraints already correctly positions the ligands to make those hydrogen bonds or because the decoys are equally likely to be able to satisfy the constraints. X-ray complex structures with ligase exist for three of the ligands with $IC_{50}s$ less than $100 \mu M$. For one of the ligands, the docking with and without the constraints positions the ligand in a similar way in the binding pocket. For the other two ligands, the 1/3 hydrogen bonds constraint runs do not position the ligand correctly because in the X-ray position the ligand does not make any hydrogen bonds to the three hinge-like backbone atoms so applying the constraint in this case forces the ligand to adopt an incorrect pose.

For the enrichment ROC plots in Fig. 10 all fragments with $IC_{50}s \leq 1 \text{ mM}$, the highest IC_{50} possible based on the assay conditions, were considered as actives. When only the better actives, those with $IC_{50}s \leq 100 \mu M$, are considered, in general the enrichment is increased (Fig. 11). Overall the experimental hit rate from the high concentration fragment screen was 4.1% (794/19,299). If based on the no constraint virtual screen only 1% of the library, or 193 fragments, were assayed, the hit rate would have been 13.5%. Thus, the enrichment factor (EF) at 1% of the ranked database was ~ 3.3 for the no constraint run. This EF increased to 4.2 when only the better actives were considered. While these enrichment factors are not outstanding (see Table 5 for EFs for each run), they are

Table 5 Enrichment factor at 1% of ranked library for Ligase

Constraints	Actives $\leq 1 \text{ mM}$	Actives $\leq 100 \mu M$
None	3.3	4.2
1/5 Hbonds	2.7	3.6
2/5 Hbonds	3.4	4.2
1/3 “hinge” Hbonds	3.4	5.4

significantly better than random and the ligase test case is particularly difficult. For this test case, a library of 19,299 property-matched fragments was screened and the enrichment of very weak actives was examined.

Conclusions

Fragment-based drug discovery approaches allow for a greater coverage of chemical space and generally produce high quality, albeit weak, leads. Virtual and experimental fragment screening are increasingly being used to create target-focussed libraries for experimental fragment screening. This paper represents one of the first published examples of fragment docking test cases with experimentally validated non-actives as well as actives.

In this paper, the use of nine different docking and scoring protocols for virtual fragment screening are explored. GlideSP with either GlideScore or Emodel ranking performs the best of the schemes tested for PGDS fragment docking. Self and cross-docking accuracy is similar to what has generally been reported for lead-like molecules (e.g., [32]). Most surprisingly MM-GBSA re-scoring of docked poses for PGDS does not improve the self-docking, cross-docking, or enrichment. Other groups have found that MM-GBSA re-scoring with binding site minimization can improve docking accuracy for congeneric series of drug-like molecules (e.g., [27]) and for distinguishing known binders from known decoys for simple, engineered model binding sites (e.g., [33]). The results presented in this paper and elsewhere (e.g., [34]), however, suggest that success with MM-GBSA re-scoring may be system dependent.

For the ligase test case, the use of various hydrogen bond constraints does not significantly improve the docking performance. For both test cases, GlideSP with default settings is able to produce enrichment of actives over random sampling. The enrichment rates obtained for the ligase test case, especially for the better actives, are within the ranges reported for virtual screening of drug-like molecules (e.g., [28, 35, 36]). Attempts to improve upon this enrichment through the use of more computationally intensive procedures that are often now routinely applied either decreased or did not improve the enrichment.

Fragment screening is an emerging area with great potential for drug discovery. The results presented here show that virtual fragment screening also has potential and that even in very difficult test cases it yields results that are significantly better than random. It is probably fair to say, however, that fragment-docking protocols have yet to be fully optimized. Enhancements to the technology specifically aimed at increasing the accuracy of fragment docking are needed and may require improved enthalpy and entropy predictions.

Acknowledgments We thank Joann Prescott-Roy, Rutger Folmer, Loredana Spadola, and Peter Kenny for their help in locating and curating data sets and Adam Shapiro for providing experimental data on ligase in advance of publication.

References

- Howard S, Berdini V, Boulstridge JA, Carr MG, Cross DM, Curry J, Devine LA, Early TR, Fazal L, Gill AL, Heathcote M, Maman S, Matthews JE, McMenamin RL, Navarro EF, O'Brien MA, O'Reilly M, Rees DC, Reule M, Tisi D, Williams G, Vinkovic M, Wyatt PG (2009) Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multitargeted kinase inhibitor with potent aurora kinase activity. *J Med Chem* 52:379–388. doi:10.1021/jm800984v
- Edwards PD, Albert JS, Sylvester M, Aharony D, Andisik D, Callaghan O, Campbell JB, Carr RA, Chessari G, Congreve M, Frederickson M, Folmer RH, Geschwindner S, Koether G, Kolmodin K, Krumrine J, Mauger RC, Murray CW, Olsson LL, Patel S, Spear N, Tian G (2007) Application of fragment-based lead generation to the discovery of novel, cyclic amidine beta-secretase inhibitors with nanomolar potency, cellular activity, and high ligand efficiency. *J Med Chem* 50:5912–5925. doi:10.1021/jm070829p
- Geschwindner S, Olsson LL, Albert JS, Deinum J, Edwards PD, de Beer T, Folmer RH (2007) Discovery of a novel warhead against beta-secretase through fragment-based lead generation. *J Med Chem* 50:5903–5911. doi:10.1021/jm070825k
- Albert JS, Blomberg N, Breeze AL, Brown AJ, Burrows JN, Edwards PD, Folmer RH, Geschwindner S, Griffen EJ, Kenny PW, Nowak T, Olsson LL, Sangane H, Shapiro AB (2007) An integrated approach to fragment-based lead generation: philosophy, strategy and case studies from AstraZeneca's drug discovery programmes. *Curr Top Med Chem* 7:1600–1629. doi:10.2174/156802607782341091
- Erlanson DA, Wells JA, Braisted AC (2004) Tethering: fragment-based drug discovery. *Annu Rev Biophys Biomol Struct* 33:199–223. doi:10.1146/annurev.biophys.33.110502.140409
- Hohwy M, Spadola L, Lundquist B, Hawtin P, Dahmen J, Groth-Clausen I, Nilsson E, Persdotter S, von Wachenfeldt K, Folmer RH, Edman K (2008) Novel prostaglandin D synthase inhibitors generated by fragment-based drug design. *J Med Chem* 51:2178–2186. doi:10.1021/jm701509k
- Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* 6:29–40. doi:10.1038/nrd2201
- Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* 9:430–431. doi:10.1016/S1359-6446(04)03069-7
- Kuntz ID, Chen K, Sharp KA, Kollman PA (1999) The maximal affinity of ligands. *Proc Natl Acad Sci USA* 96:9997–10002. doi:10.1073/pnas.96.18.9997
- Abad-Zapatero C, Metz JT (2005) Ligand efficiency indices as guideposts for drug discovery. *Drug Discov Today* 10:464–469. doi:10.1016/S1359-6446(05)03386-6
- Carr RAE, Congreve M, Murray CW, Rees DC (2005) Fragment-based lead discovery: leads by design. *Drug Discov Today* 10:987–992. doi:10.1016/S1359-6446(05)03511-7
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23:3–25. doi:10.1016/S0169-409X(96)00423-1

13. Oprea TI (2000) Property distribution of drug-related chemical databases. *J Comput Aided Mol Des* 14:251–264. doi:[10.1023/A:1008130001697](https://doi.org/10.1023/A:1008130001697)
14. Oprea TI, Davis AM, Teague SJ, Leeson PD (2001) Is there a difference between leads and drugs? A historical perspective. *J Chem Inf Comput Sci* 41:1308–1315. doi:[10.1021/ci010366a](https://doi.org/10.1021/ci010366a)
15. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 45:2615–2623. doi:[10.1021/jm020017n](https://doi.org/10.1021/jm020017n)
16. Congreve M, Carr R, Murray C, Jhoti H (2003) A ‘rule of three’ for fragment-based lead discovery? *Drug Discov Today* 8:876–877. doi:[10.1016/S1359-6446\(03\)02831-9](https://doi.org/10.1016/S1359-6446(03)02831-9)
17. Joseph-McCarthy D, Baber JC, Feyfant E, Thompson DC, Humblet C (2007) Lead optimization via high-throughput molecular docking. *Curr Opin Drug Discov Dev* 10:264–274
18. 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 PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47:1739–1749. doi:[10.1021/jm0306430](https://doi.org/10.1021/jm0306430)
19. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* 47:1750–1759. doi:[10.1021/jm030644s](https://doi.org/10.1021/jm030644s)
20. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J Med Chem* 49:6177–6196. doi:[10.1021/jm051256o](https://doi.org/10.1021/jm051256o)
21. Hohwy M, Spadola L, Lundquist B, Hawtin P, Dahmén J, Groth-Clausen I, Nilsson E, Persdotter S, von Wachenfeldt K, Folmer R, Edman K (2008) Novel prostaglandin D synthase inhibitors generated by fragment-based drug design. *J Med Chem* 51:2178–2186. doi:[10.1021/jm701509k](https://doi.org/10.1021/jm701509k)
22. Aritake K, Kado Y, Inoue T, Miyano M, Urade Y (2006) Structural and functional characterization of HQL-79, an orally selective inhibitor of human hematopoietic prostaglandin D synthase. *J Biol Chem* 281:15277–15286. doi:[10.1074/jbc.M506431200](https://doi.org/10.1074/jbc.M506431200)
23. Engler MJ, Richardson CC (1982) DNA ligases. In: Boyer PD (ed) *The enzymes*. Academic Press, Inc., New York, NY, pp 3–29
24. Kenny PW (2009) *J Comput Aided Mol Des* In (this issue)
25. Weininger D (1988) SMILES 1. Introduction and encoding rules. *J Chem Inf Comput* 28:31–36
26. Kenny PW, Sadowski J (2004) Structure modification in chemical databases. In: Oprea TI (ed) *Chemoinformatics in drug discovery*. Wiley, Weinheim, pp 271–285
27. Lyne PD, Lamb ML, Saeh JC (2006) Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J Med Chem* 49:4805–4808. doi:[10.1021/jm060522a](https://doi.org/10.1021/jm060522a)
28. Chen HM, Lyne PD, Giordanetto F, Lovell T, Li J (2006) Evaluating molecular-docking methods for pose prediction and enrichment factors. *J Chem Inf Model* 46:401–415. doi:[10.1021/ci0503255](https://doi.org/10.1021/ci0503255)
29. Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO (2005) Virtual screening workflow development guided by the “receiver operating characteristic” curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem* 48:2534–2547. doi:[10.1021/jm049092j](https://doi.org/10.1021/jm049092j)
30. Sing T, Sander O, Beerenwinkel N, Lengauer T (2005) ROCr: visualizing classifier performance in R. *Bioinformatics* 21:3940–3941. doi:[10.1093/bioinformatics/bti623](https://doi.org/10.1093/bioinformatics/bti623)
31. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R (2006) Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem* 49:534–553. doi:[10.1021/jm050540c](https://doi.org/10.1021/jm050540c)
32. Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW (2008) Protein-ligand docking against non-native protein conformers. *J Chem Inf Model* 48:2214–2225. doi:[10.1021/ci8002254](https://doi.org/10.1021/ci8002254)
33. Graves AP, Shivakumar DM, Boyce SE, Jacobson MP, Case DA, Shoichet BK (2008) Rescoring docking hit lists for model cavity sites: predictions and experimental testing. *J Mol Biol* 377:914–934. doi:[10.1016/j.jmb.2008.01.049](https://doi.org/10.1016/j.jmb.2008.01.049)
34. Thompson DC, Humblet C, Joseph-McCarthy D (2008) Investigation of MM-PBSA rescoring of docking poses. *J Chem Inf Model* 48:1081–1091. doi:[10.1021/ci700470c](https://doi.org/10.1021/ci700470c)
35. Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP (2005) Comparison of automated docking programs as virtual screening tools. *J Med Chem* 48:962–976. doi:[10.1021/jm049798d](https://doi.org/10.1021/jm049798d)
36. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) A critical assessment of docking programs and scoring functions. *J Med Chem* 49:5912–5931. doi:[10.1021/jm050362n](https://doi.org/10.1021/jm050362n)