J Comput Aided Mol Des (2012) 26:921–934
DOI 10.1007/s10822-012-9590-x

Virtual fragment screening: exploration of MM-PBSA re-scoring

Sameer Kawatkar · Demetri Moustakas · Matthew Miller · Diane Joseph-McCarthy

Received: 4 April 2012 / Accepted: 19 July 2012 / Published online: 7 August 2012

Abstract An NMR fragment screening dataset with known binders and decoys was used to evaluate the ability of docking and re-scoring methods to identify fragment binders. Re-scoring docked poses using the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) implicit solvent model identifies additional active fragments relative to either docking or random fragment screening alone. Early enrichment, which is clearly most important in practice for selecting relatively small sets of compounds for experimental testing, is improved by MM-PBSA re-scoring. In addition, the value in MM-PBSA re-scoring of docked poses for virtual screening may be in lessening the effect of the variation in the protein complex structure used.

Keywords Virtual screening · Fragment screening · Prostaglandin D2 synthase · MM-PBSA · Docking

Introduction

Fragment based drug discovery (FBDD) has become popular as a lead generation approach, with fragment screening often being performed alongside, or in place of, more conventional high-throughput screening (HTS) to identify promising chemical leads [1–6]. An often-cited advantage of fragment screening is that due to their low molecular weights (typically <300 daltons), fragments are less chemically complex than larger compounds, thereby reducing the chemical space that must be searched to identify bioactive chemotypes [7]. In principle, this enables diverse, drug-like chemotypes to be explored by sampling a relatively small number of compounds. Related to their low molecular weight (MW), fragment leads generally have greater ligand efficiency [7–10], and more drug-like physical properties [11–14]. The principal disadvantage of fragment screening is that fragments typically bind with much lower potency than higher MW compounds, such as those found by HTS. In this case, the lower MW of fragments results in fewer protein–ligand interactions (van der Waals interactions, polar interactions) to contribute to the affinity. As a result, screening fragments using biochemical assays requires a large amount of compound. Biophysical assays [7] such as Biacore, NMR, ITC, or Thermofluor, are more sensitive and able to detect fragments that bind with lower affinity, however, these methods have lower throughput than automated biochemical assays. These practical considerations somewhat limit the number of fragment samples that can be tested.

Additionally, since leads resulting from fragment screening tend to have lower potency than leads originating from high-throughput screening of larger compounds, drug discovery teams are often reluctant to pursue an FBDD strategy in the absence of biostructural data revealing the binding mode of the fragment lead, to facilitate hopefully more rapid optimization of the fragment potency through structure-based drug design [15]. Without this guiding information, fragment leads appear to require more effort to optimize than high-throughput screening leads (since they are weaker starting points), and leads derived from HTS are often prioritized over leads derived from fragment screening.

S. Kawatkar · D. Moustakas · M. Miller · D. Joseph-McCarthy (✉)
Infection Innovative Medicines Unit, Chemistry Department, AstraZeneca, R&D Boston, 35 Gatehouse Drive, Waltham, MA 02451, USA
e-mail: djosephmc@verizon.net

Virtual screening is often used to select a set of fragments for screening that should be enriched in bioactivity, just as it is often used to enhance more conventional screening of lead-like compounds. Virtual fragment screening can enrich the number of active fragments in the screening library, helping to mitigate the risk of screening smaller numbers of compounds. Additionally, docking provides binding mode hypotheses for the fragments, which can be used in the absence of three-dimensional structure of the target-protein complex to optimize the biochemical potency of an initial fragment hit. Virtual fragment screening therefore offers potential solutions to many of the issues inherent in fragment screening.

Since fragments are less chemically and structurally complex than lead-like compounds, there tend to be fewer conformational degrees of freedom for a docking algorithm to search. This should translate to more accurate dockings (relative to docking larger compounds), as it is likely that docking will successfully sample the correct fragment conformation. However, since scoring fragment poses can be difficult, obtaining the correct pose (or orientation of the fragment in the binding site) may also be difficult [16, 17]. Since fragment hits tend to have greater solubility and polarity than traditional HTS hits [18], along with lower binding affinities relative to larger, more lipophilic, HTS-like compounds, ligand desolvation may dominate the binding energies of fragments. Principally, the desolvation free energies of the protein and ligand, and the solvent screening of electrostatic interactions are thought to be important [19–22]. Docking scoring functions, which derive much of their predictive power by summing the van der Waals and polar intermolecular interactions, tend to assign fragment poses lower scores with an overall smaller dynamic range [16]. Docking scoring functions often either neglect solvation effects entirely, or approximate them with relatively simplistic models (e.g. empirically-derived group additive desolvation parameters, or distance-dependant dielectric Coulombic terms) [23, 24]. Including solvation effects would be expected to have a more positive effect on the scoring of docking poses for fragments than for larger molecules, since solvation effects likely constitute a greater proportion of the binding free energy than in systems with more lead-like molecules.

There are many reported approaches that use explicit water molecules in the system to calculate the solvent–solute energies more accurately [25–31]. However, these approaches—thermodynamic integration (TI), free-energy perturbation (FEP), grand canonical Monte Carlo (GCMC), to name a few—are very computationally intensive (requiring a great deal of sampling of water configurations), and are not compatible with the demands of a typical docking algorithm, which often evaluates the scores of thousands of poses of each molecule.

The class of scoring functions based on continuum solvent models offers a promising compromise between using expensive explicit solvent models, and outright neglecting the contributions of solvent–solute interactions. Continuum solvent models, principally the Poisson-Boltzmann and Generalized Born approximations, model the solute with explicit atoms, and treat the solvent as a high dielectric continuum field. This enables approximation of the energies of desolvation (driven by the energy required to move charged atoms into and out of a high dielectric medium) of the protein and the ligand in a computationally efficient manner (without having to sample a huge number of water molecule configurations) [32]. The MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) re-scoring function is

$$\begin{aligned}\Delta G_{bind} &\approx E_{bind} \\ &= E_{vdw,P-L} + E_{Coul,P-L} \\ &\quad + \left(E_{PB,P-L} - E_{PB,P} - E_{PB,L}\right) \\ &\quad + \left(E_{SA,P-L} - E_{SA,P} - E_{SA,L}\right) \end{aligned} \quad (1)$$

where $E_{vdw,P-L}$ is the protein–ligand van der Waals interaction energy, $E_{Coul,P-L}$ the protein–ligand Coulombic electrostatic interaction energy, E_{PB} is the Poisson-Boltzmann electrostatic energy, and E_{SA} is the non-polar energy. The latter two terms (E_{PB} and E_{SA}) are intended to account for the major effects of water upon protein–ligand binding [33]. The Poisson-Boltzmann (PB) continuum solvent model [34–36] treats water as a high dielectric continuum, within which it computes the electrostatic potential of the solute. This incorporation of the high dielectric solvent continuum has the effect of screening Coulombic interactions within the solute molecules, and it has been shown to accurately reproduce aqueous hydration energies for small organic molecules. The surface area (SA) term is empirically parameterized to account for the non-polar "hydrophobic effect" in protein–ligand binding. The decrease in solvent-exposed solute SA upon binding is correlated with more entropically favorable configurations of the solvent [37]. Existing implementations of the MM-PBSA scoring function are sufficiently fast to allow its use in re-scoring large numbers of docked fragment poses. Due to the very large number of score evaluations made for each ligand during docking (perhaps hundreds of thousands), simpler scoring functions are still often used as the primary docking scoring function; a smaller number of docked poses of each ligand may then be re-scored with a scoring function based on a continuum solvent model [38, 39]. This hybrid approach is thought to allow the faster scoring function to filter out obviously incorrect ligand poses, prior to re-scoring with the more accurate solvation scoring function [40–42].

In this work, we investigate re-scoring docked fragment poses with an MM-PBSA scoring function to improve

docking accuracy. Self-docking accuracy and enrichment rates after MM-PBSA re-scoring relative to experimental hit rates are explored for a prostaglandin D_2 synthase (PGDS) fragment screening dataset. We have previously shown for this dataset that Glide docking performs well and yields some enrichment over random sampling [43]. The results presented herein demonstrate that a virtual fragment screening workflow of docking followed by MM-PBSA re-scoring identifies more active fragments than either docking or experimental fragment screening alone would, and that the early enrichment (the number of actives enriched in the top 1 % of the ranked compound set) is particularly improved by MM-PBSA re-scoring.

Methods

PGDS test set

PGDS carries out the isomerization of PGH_2 to PGD_2, an allergic and inflammatory mediator (Fig. 1), in the presence of the cofactor glutathione (GSH) [1, 44]. A ~2,000 fragment library was screened by NMR spectroscopy for binders to PGDS and 28 fragment hits with binding affinities of $K_d \leq 500$ μM were identified [1] (see Fig. 2). A set of X-ray structures of PGDS in complex with seven ligands, **L1**–**L7**, respectively, also exists (PDB ids: 2CVD, 2VCQ, 2VCW, 2VCX, 2VCZ, 2VD0, and 2VD1) [1, 45]. The protein structures corresponding to each complex structure will be referred to as **P1**–**P7**. The resolution of the complex structures ranges from 1.45 to 2.25 Å. In the present work, the X-ray complex structures are used as a test set for self-docking, and the full screening dataset (**L2**–**L29** and the 1,823 non-binders) is used as an enrichment test set. **L1** was not part of the fragment screen and is somewhat larger than the other compounds tested.

Protein structures and preparation

In the **P1**–**L1** complex structure, the protein conformation in the active site is significantly different than for the other six protein structures (**P2**–**P7**), which are very similar to one another (Fig. 3). In the **P1** protein structure, Trp104 undergoes a significant conformational change compared to its position in other structures to accommodate the larger **L1** ligand. The minimum pair-wise α-carbon root-mean-square deviation (RMSD) for **P1** from the other six protein structures is 1.46 Å, while the maximum pair-wise α-carbon RMSD among the **P2**–**P7** proteins is only 0.33 Å. All seven X-ray structures display the co-factor GSH bound in a similar position. A structural Mg^{2+} ion is also observed bound in three of the protein structures (**P1**, **P4** and **P7**). For the docking study, the GSH and Mg^{2+} were retained in the structures as observed.

All protein structures were prepared using the Maestro 8.5 (Schrodinger, LLC, 2008, New York, NY) Protein Preparation Wizard workflow; waters were deleted, bond orders were assigned, hydrogens were added, and metals were treated appropriately when present. Next, the orientations of hydroxyl groups, amide groups of Asn and Gln, and the charge states of His residues were optimized followed by a restrained minimization of the protein structure using the default constraint of 0.30 Å RMSD overall and the OPLS 2001 force field [46].

Virtual screening library preparation

Virtual libraries of ligands **L1**–**L7** and of the 1,851 fragments (which include ligands **L2**–**L7**) screened by NMR for binding to PGDS [1] were generated as follows. SMILES representations [47] of the ligands were input to Leatherface [48], an in-house molecular editor based on the OpenEye (OpenEye Scientific Software, Inc., 2008, Santa Fe, NM) OEChem toolkit which was used to generate reasonable protonation and tautomeric states. The ligprep utility in Maestro 8.5 was used to generate three dimensional coordinates for all ligands. Leatherface generated protonation and tautomeric states; any specified chiralities were retained, and one low energy ring conformation was generated per ligand. Average physical properties were calculated for the full library of 1851 compounds and shown to be similar to those for the 28 actives (see Table 1).

Docking and scoring protocol

The Glide program with the SP scoring function [49, 50] was used to flexibly dock ligands for both the self-docking and enrichment studies. Glide scoring grids were generated for each of the seven prepared PGDS structures. A grid box

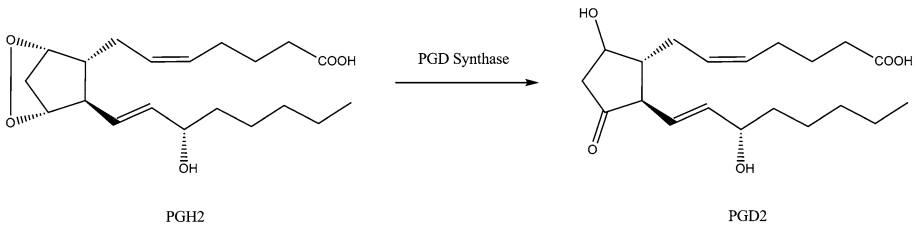


Fig. 1 The reaction catalyzed by PGDS

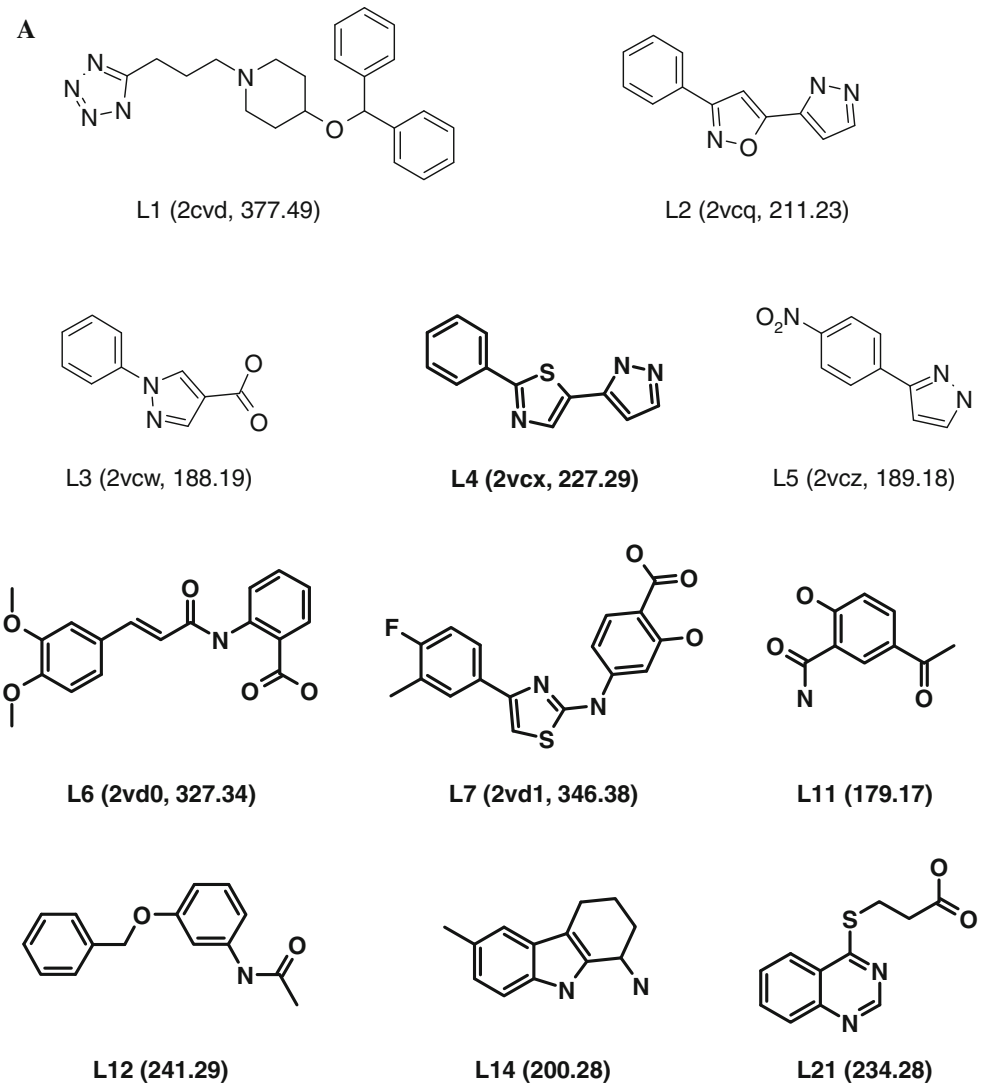


Fig. 2 PGDS binders. In **a** are the ligands used for the self-docking studies (**L1**–**L7**) as well as all true positives identified in the enrichment studies in the top 1 % of the ranked dataset by any protocol (shown in bold and listed in Table 4). In **b** are the additional true positives that are not identified in the top 1 % in the enrichment studies. **L1** is not in the library docked for the enrichment studies. PDB codes for the corresponding protein–ligand complexes if available and molecular weights of ligands are given in *parentheses*

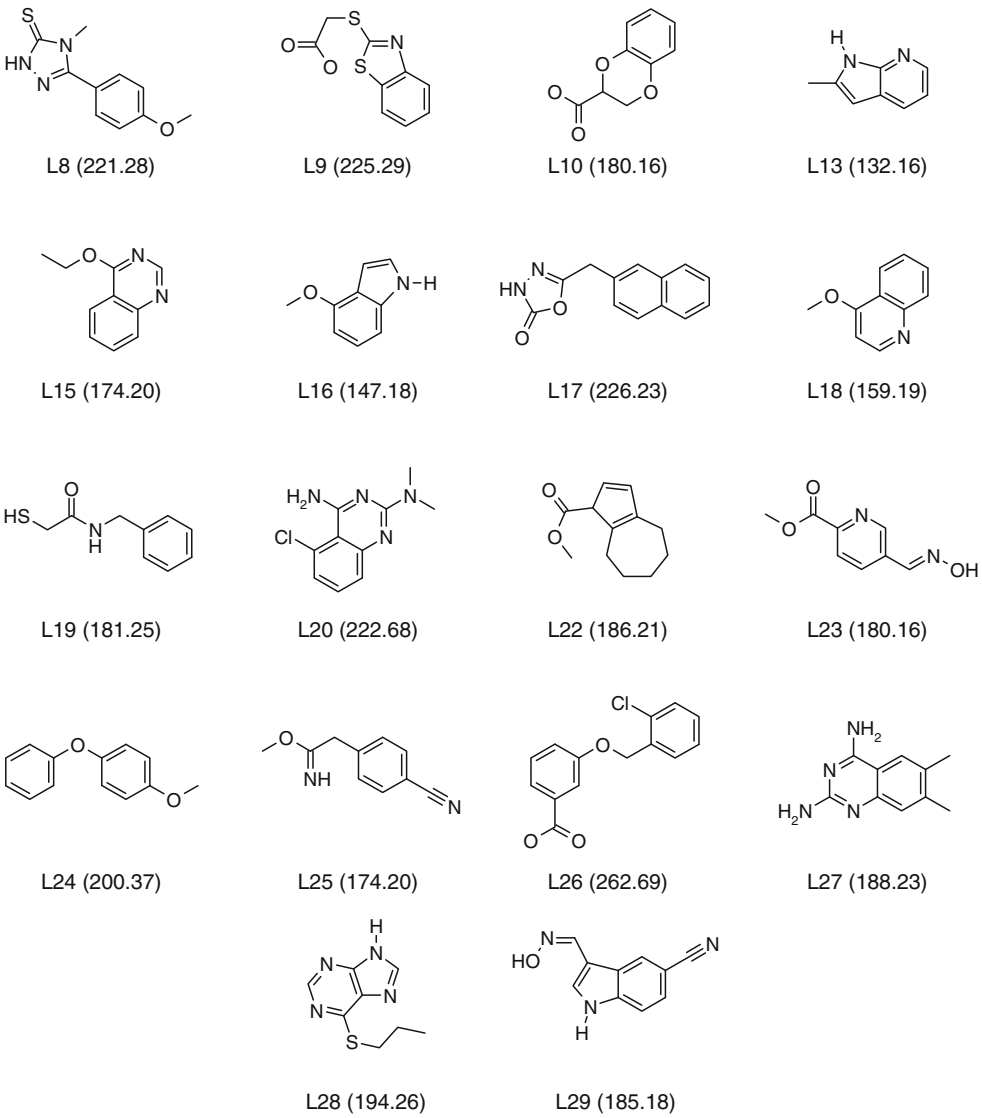
of default size ($20 \times 20 \times 20$ Å^3) was centered on the corresponding ligand position and no constraints were included. Default parameters were used for the docking runs, and the top nine poses based on the GlideSP docking score were saved for each ligand (saving more than nine poses turns off the degeneracy check within Glide and can therefore result in a set of nearly identical poses).

MM-PBSA re-scoring of docked poses

MM-PBSA re-scoring involves calculating the specified molecular mechanics, PB electrostatic, and SA terms (see Eq. 1) for the complex, the ligand, and the protein, respectively, and then subtracting those for the ligand and the protein from those for the complex to approximate the free energy of binding. As part of the re-scoring step, the ligand is often minimized in the complex first. In this paper, eight different docking pose re-scoring protocols are evaluated: MM-PB, MM-PBSA, Cart-MM-PBSA, RB-MM-PBSA, Tor-MM-PBSA, MM-PB-a, MM-PBSA-a, and TorV-MM-PBSA-a (see Table 2). This was done in part to tease apart the contribution of the various terms to the calculated free energy of binding. MM-PB refers to MM-PBSA re-scoring without the SA term, the prefix (Cart-, RB-, Tor-, or TorV-) refers to the type of minimization that was used to optimize the complex, if any, and the suffix "−a" indicates that the DOCK-AMBER 6-9 softened Lennard-Jones vdW potential was employed. The MM-PB-a, MM-PBSA-a, and TorV-MM-PBSA-a protocols utilize the DOCK-AMBER vdW potential [51, 52] to calculate the MM terms and Bondi radii [53], OPLS 2001 [46] partial charges for the protein, and MMFF94 partial charges [54, 55] for the ligand to compute the PB term. The MM-PB, MM-PBSA, Cart-MM-PBSA, RB-MM-PBSA and Tor-MM-PBSA protocols employ the MMFF94 vdW potential [53, 54] for the MM terms and Bondi radii and MMFF94 partial charges for the ligand and protein to compute the PB term. With the MM-PB, MM-PBSA, MM-PB-a, and MM-

Fig. 2 continued

B



PBSA-a protocols, the ligand is not re-minimized. The Cart-MM-PBSA, RB-MM-PBSA and Tor-MM-PBSA protocols perform minimization of the ligand in the complex using Cartesian minimization of individual atomic positions (cart_opt), rigid-body minimization of the entire ligand (rigid_opt), and torsional minimization of the ligand torsions (tor_opt), respectively, using the OpenEye Szybki minimizer with default settings. With the TorV-MM-PBSA-a protocol, a "Simplex" minimizer is used to perform torsional minimization of the vdW term only, using the Nelder-Mead Simplex algorithm to eliminate any steric clashes between the protein and the ligand. More specifically, a program was written with the OEChem C++ toolkit and the NLopt C++ library [http://abinitio.mit.edu/wiki/index.php/NLopt_Algorithms] to apply the NLopt Simplex method to the torsional angle and rigid-body degrees of freedom of a molecule. Up to 1,000 steps of minimization were performed, with an initial step size of 0.3 Å for translation, 0.3 radians for rotation, 0.3 radians for torsion angle changes, and a threshold of 0.1 kcal · mol^{-1} to determine convergence.

All of the re-scoring protocols used the OpenEye Zap toolkit to calculate the Poisson-Boltzmann (PB) and Coulombic electrostatics terms, and the OEArea component of OEChem library to calculate the SA terms used to derive the final score. The change in the surface area upon complex formation ($\Delta SA = SA\ complex - SA\ protein - SA\ ligand$) is multiplied by a factor of 30 cal · $Å^{-2}$ to account for the non-polar desolvation energy of binding (that is, $E_{SA} = \gamma \Delta SA$, where $\gamma = 30$ cal · $Å^{-2}$) [56].

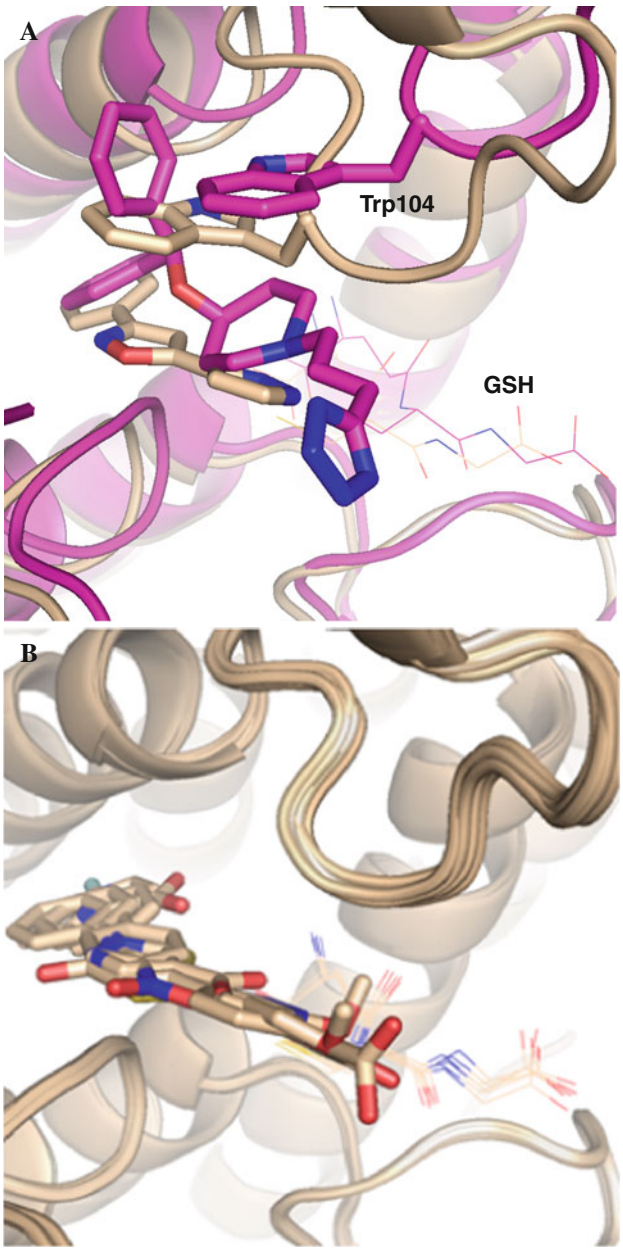


Fig. 3 Superposition of the PGDS X-ray complex structures. **a** The **P1**–**L1** (in magenta carbons) and **P2**–**L2** (in tan carbons) complex structures are shown superimposed by protein backbone atoms; the protein is shown in *ribbon* representation with Trp 104 and the ligands are in *thick lines*, and GSH in *thin lines* colored by element. **b** The complex structures corresponding to proteins **P2**–**P7** are shown superimposed and *colored* similarly

Analysis of self-docking accuracy and enrichment

For the self-docking results, a pose is considered "correct" if it has an RMSD of $\leq$2 Å from the X-ray position for that ligand bound to the target. We calculated this value using the RMSD python script in the OEChem toolkit. To evaluate the enrichment of PGDS binders from the NMR dataset by the virtual screening, a Receiver Operating Characteristic (ROC) curve [57, 58] is generated, plotting the fraction of binders (true positives) against the fraction of non-binders (true negatives), for the database as prioritized by the various re-scoring protocols. The enrichment factor at any point on the ROC plot is then calculated as the magnitude of the deviation of the ROC curve from the diagonal line, where the diagonal represents the expected hit rate for random selection of compounds from the data set.

$$\text{Enrichment Factor} = (\text{Hits}_{\text{sel}}/\text{Hits}_{\text{tot}})/(\text{NB}_{\text{sel}}/\text{NB}_{\text{tot}})$$

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, NB_{tot} is the total number of non-binders in data set, and NB_{sel} is the number of non-binders found in X% of the ranked database. The area under the curve (AUC) value is computed for each protocol for the first 1 and 5 % of the ranked data set, by integrating the area under the ROC curve generated for each protocol between the origin of the plot and the 1 and 5 % x-axis intercepts.

Results and discussion

Self-docking using Glide

Six (**L2**–**L7**) of the seven PGDS ligands (Fig. 2) dock into their native structures correctly with an RMSD <2 Å (Table 3) using GlideSP. In the case of **L1**, the two phenyl rings are correctly positioned relative to the X-ray pose of the ligand (Fig. 4a). The piperidine ring is in a twist-boat conformation in the X-ray structure. However, in the top-scoring docked pose it is in a chair conformation. In the X-ray structure, the tetrazole points out into bulk solvent, while in the top scoring pose, the tetrazole bends back toward the protein and forms a hydrogen bond with the guanidinium moiety in Arg14 resulting in an overall RMSD of 5.47 Å. The second best scoring pose for **L1** with GlideSP, however, is correct with an RMSD of 1.63 Å to the X-ray structure position, with the tetrazole moiety pointing out into solvent as it is in the X-ray complex structure (Fig. 4b). Examples of ligands correctly self-docked by GlideSP are shown in Fig. 4c, d.

For the re-scoring protocols involving minimization of the ligand, the orientation and conformation of the top scoring GlideSP pose does not significantly change, as evidenced by the negligible effect on the RMSD (Fig. 5a). In fact, with the TorV-MM-PBSA-a protocol, the average RMSD over all nine saved poses for all seven protein structures was 0.36 Å with a standard deviation of 0.28 Å, indicating that, in general, the re-scoring minimization did not significantly alter the GlideSP poses. The ranking of the

Table 1 Average physical properties of fragment data set

	MW	HAC	sLogP[a]	Rotatable bonds
Full library	196	13.7	1.7	2.3
	(85–443)	(6–31)	(−1.9–5.4)	(0–10)
Actives	206.5	14.8	2.2	2.4
	(132–344)	(10–24)	(0.7–4.4)	(0–7)

Ranges of values are given in parentheses

[a] Calculated using the approach of Wildman and Crippen [59]

poses following re-scoring, however, does change with or without minimization (Fig. 5b). For example, in the case of ligands **L5** and **L7**, the top-ranked pose by GlideSP is not ranked highest by any of the protocols evaluated. Three protocols, MM-PB, MM-PBSA and Cart-MM-PBSA, are able to select a correct pose as the top pose for all seven ligands, including **L1** (Fig. 5c). RB-MM-PBSA, MM-PB-a, MM-PBSA-a, and TorV-MM-PBSA-a select a correct pose for 6 of the 7 ligands, and MM-PB-a and MM-PBSA-a rank a near-native pose as the top of the set (RMSD = 2.1 Å). TorV-MM-PBSA-a fails to select a correct pose for **L1**, while RB-MM-PBSA and Tor-MM-PBSA fail to for **L3**. Tor-MM-PBSA re-scoring selects a correct pose for only 5 of the 7 complexes, respectively; it does not identify a correct pose for **L1** or **L3**. As mentioned above, in its native pose the tetrazole in **L1** points out into bulk solvent and possibly is more flexible as it does not form any interactions with protein residues. Accounting for solvation effects by including a PB electrostatics term in re-scoring on balance helps to identify a correct pose for **L1**. For **L3**, it appears, based on the sensitivity to the minimizer, that selection of the correct pose is strongly influenced by sterics. With the TorV-MM-PBSA-a protocol only the vdW and not the electrostatic term is minimized and a correct pose is selected. The minimizer likely helps to select the right pose by improving the vdW score, as defined by the corresponding re-scoring potential. The Tor-MM-PBSA and RB-MM-PBSA minimizers failed to improve sufficiently the vdW score within the pocket. In contrast, the Cartesian minimizer utilized in the Cart-MM-PBSA protocol is more effective and therefore able to select a correct pose. Overall these results demonstrate that 5 of the 8 re-scoring protocols explored perform somewhat better than GlideSP in identifying a correct pose for all seven ligands.

Enrichment of PGDS actives

For the 6 protocols that performed the same as or better than GlideSP, the enrichment of 28 binders (including ligands **L2–L7**) from the set of 1,851 fragments screened by NMR for binding to PGDS was investigated. The

Table 2 MM-PBSA re-scoring protocols

Force field	Minimization			Re-scoring		Protocol name
	Functional form	Tool	Protocol	Functional form	SA factor (cal/Å^2)[a]	
vdW: MMFF94	vdW + Coul + PB	Szybki	None	vdW + Coul + PB + SA	0	MM-PB
PB: Bondi radii; MMFF94 partial charges protein and ligand			None		30	MM-PBSA
			Cartesian		30	Cart-MM-PBSA
			Rigid body		30	RB-MM-PBSA
			Torsional		30	Tor-MM-PBSA
vdW: AMBER	Reduced vdW	Simplex torsional minimizer	None	Reduced vdW + Coul + PB + SA	0	MM-PB-a
PB: Bondi radii; OPLS partial charges protein, MMFF94 partial charges ligand			None		30	MM-PBSA-a
			Torsional		30	TorV-MM-PBSA-a

MMFF94 [54, 55], AMBER [51, 52]

Lennard-Jones vdW functional form, using softened values of 6 and 9 for the attractive and repulsive exponents, rather than the typically used values of 6 and 12

$E = \sum_{i=1}^{lig}\sum_{j=1}^{rec}\left[\frac{A_{ij}}{r_{ij}^{9}} - \frac{B_{ij}}{r_{ij}^{6}}\right]$ where Aij and Bij are the DOCK/AMBER vdW attractive and repulsive parameters for the interaction between the *i*th and *j*th ligand and receptor atoms, r_{ij} is the distance between the *i*th and *j*th atom of the ligand and receptor, respectively

[a] The surface area factor, γ, where hydrophobic term $E_{SA} = \gamma\Delta SA$

Table 3 Self-docking for PGDS dataset using Glide

Ligand	Top pose		Lowest RMSD pose			First 'correct' pose[a]		
	RMSD	Docking score	RMSD	Rank	Docking score	RMSD	Rank	Docking score
L1	5.47	−8.28	1.60	6	−8.05	1.63	2	−8.19
L2	0.30	−6.36	0.30	1	−6.36	0.30	1	−6.36
L3	0.20	−7.07	0.20	1	−7.07	0.20	1	−7.07
L4	0.09	−7.25	0.09	1	−7.25	0.09	1	−7.25
L5	1.42	−7.08	0.11	5	−6.62	1.42	1	−7.08
L6	0.5	−8.23	0.5	1	−8.23	0.5	1	−8.23
L7	1.55	−7.83	0.69	4	−7.56	1.55	1	−7.83

Units are (Å) for RMSD and (kcal/mol) for Docking score

[a] Lowest ranking pose with RMSD $\leq$ 2 Å

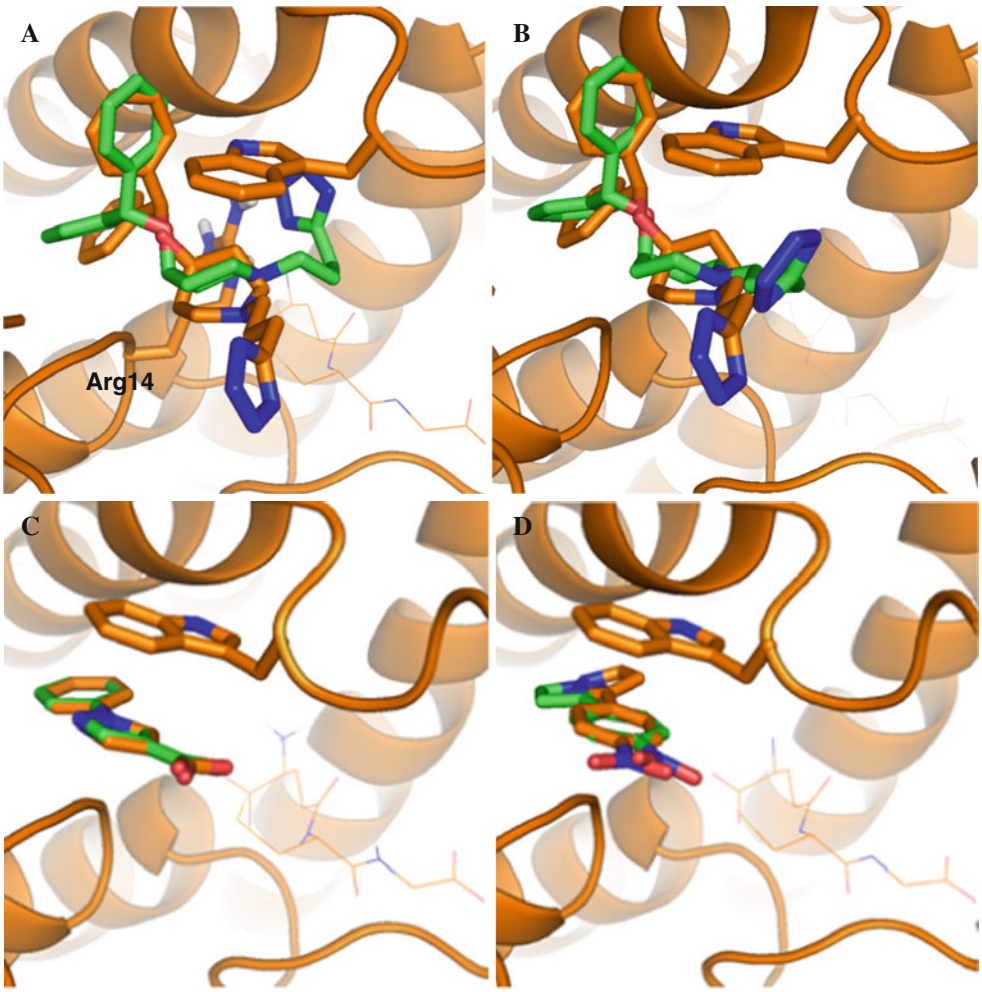


Fig. 4 Top-ranked poses identified in self-docking using GlideSP (*colored* by element with carbon in *green*) compared to the corresponding X-ray complex structure (colored by element with carbons in *orange*). The top ranked GlideSP pose for **P1** is shown in **a**, while the second best GlideSP pose for **P1** is shown in **b**. The top ranked GlideSP poses for **P3** and **P5** are shown in **c** and **d,** respectively. Select residues that interact with the ligand are shown in *thick lines*

protocols MM-PB, MM-PBSA, Cart-MM-PBSA, MM-PB-a, MM-PBSA-a and TorV-MM-PBSA-a are evaluated for their ability to improve virtual screening performance.

As revealed by the enrichment factors (EFs) at 1 % (Fig. 6) and 5 % of the ranked database, GlideSP can yield reasonable enrichment rates; however, the performance of

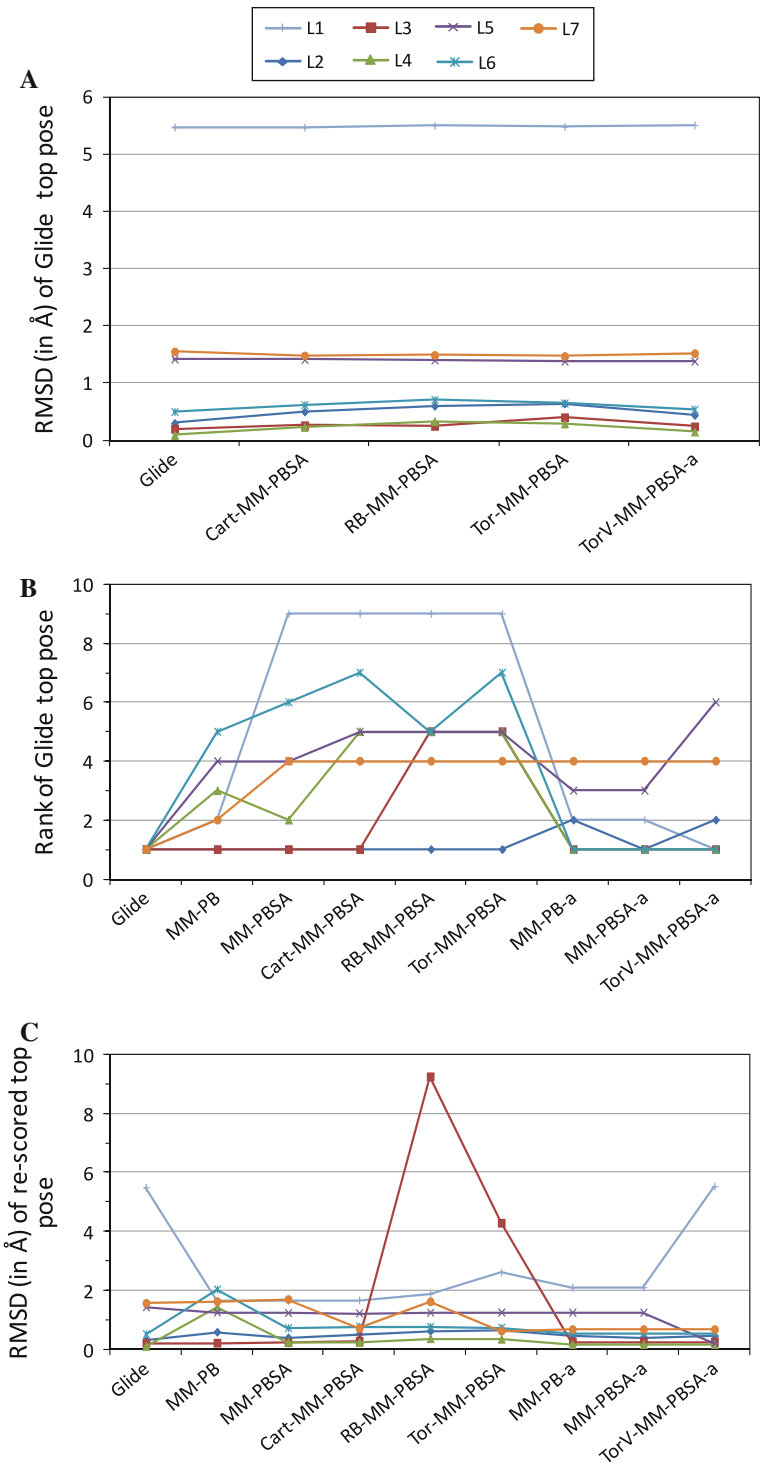


Fig. 5 Analysis of self-docking results. In **a**, the RMSD of the top ranking Glide pose versus the X-ray position before (GlideSP) and after re-scoring involving ligand minimization (Cart-MM-PBSA, RB-MM-PBSA, Tor-MM-PBSA, TorV-MM-PBSA-a) for **L1** to **L7**. In **b**, the rank of the top Glide pose before and after re-scoring is plotted for each ligand versus the re-scoring protocol used. In **c**, the RMSD of the resulting top pose after re-scoring is plotted for each ligand versus the re-scoring protocol used. The legend for the ligands for all three plots is given in **a**

GlideSP in this respect is not consistent. The success of the virtual screening is highly dependent on the choice of protein structure used in docking. The **P3** protein structure yields the best enrichment (EF = 10.44) corresponding to identifying 3 active compounds (**L6**, **L11**, and **L14**) in the top 1 % of the ranked dataset, while the **P7** structure produces the worst enrichment with no active compounds in the top 1 % of the ranked dataset (Table 4). Using the **P3**, **P4**, **P5** and **P6** protein structures the EF decreases as more compounds are screened as expected. For the **P3** protein structure, for example, screening top-ranked 5 % of the dataset yields an EF of 2.84. For the docking to the **P1** protein structure, which is the most dissimilar of the seven PGDS protein structures, GlideSP docking results in an EF

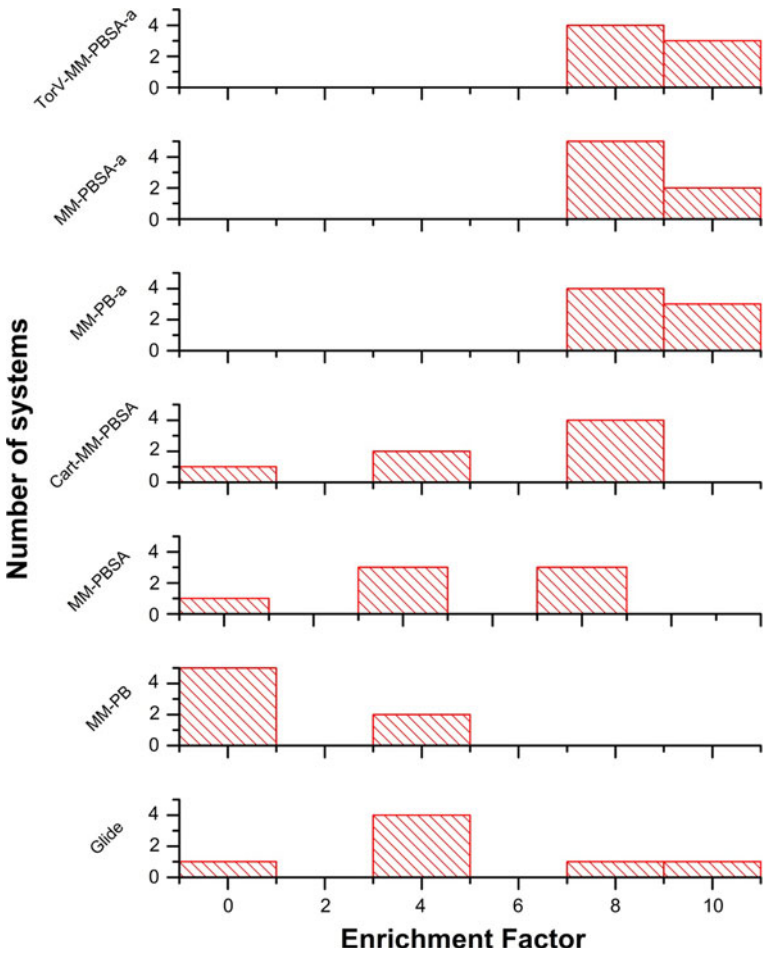


Fig. 6 Enrichment factors (EFs) for the top 1 % of the ranked dataset plotted versus re-scoring protocol utilized. For each protocol, the number of docking runs (corresponding to the seven protein structures) yielding a given enrichment factor is plotted in the histogram. Using GlideSP, for example, one docking run yielded an EF of 0, four runs yielded EFs of 4, one yielded an EF of 8, and one yielded an EF of 10

of 3.6, corresponding to only 1 active compound (**L7**) ranked in the top 1 % of the dataset.

One re-scoring protocol, MM-PB, actually consistently decreases the enrichment relative to GlideSP (Figs. 6, 7; Table 4). In fact, the early enrichment (top 1 %) decreases with the **P2**, **P4**–**P7** protein structures, as the MM-PB protocol fails to rank any active compounds in top 1 % of the dataset. The approach identifies one active compound in the top 1 % of the dataset for **P1** and **P3** structures (**L6** and **L4**, respectively). The MM-PBSA and Cart-MM-PBSA protocols perform better than MM-PB, improving the early enrichment relative to GlideSP with the **P1**, **P5** and **P7** protein structures. The performance of Cart-MM-PBSA compared to MM-PBSA for the early enrichment is improved for **P3** (**L6** and **L7** identified vs. **L4**) and **P4** (**L7** vs. none) but decreased for **P2** (none vs. **L21**). Cart-MM-PBSA re-scoring outperforms GlideSP for **P1**, **P5**, and **P7**, similarly to GlideSP for **P6** (identifying **L6** in the top 1 % of the dataset), and worse than GlideSP for **P2**, **P3**, and **P4**. These three protocols all use the MMFF94 force field for the MM terms and PB charges, and while the SA term and minimization improves the results with these re-scoring protocols none is able to outperform GlideSP. Ligands **L6** and **L7** are often identified in the top 1 %, but they are larger "fragments" with molecular weight of 327 and 346, respectively. Also, both of these ligands contain an acid functionality that forms a hydrogen bond with the protein, suggesting either that these interactions are adequately accounted for by the re-scoring functions or that explicit charge–charge interactions are over counted.

The MM-PB-a, MM-PBSA-a and TorV-MM-PBSA-a protocols do fairly consistently outperform GlideSP in terms of the early enrichment (Figs. 6, 7; Table 4). In fact, these protocols improve the results relative to GlideSP for **P1**, **P2**, **P5**, **P6** and **P7**. While these early enrichments results are based on a small number of actives, the trends are the same for multiple re-scoring protocols. With **P4** the results are similar for the re-scoring and GlideSP; two binders are identified but with the re-scoring protocols they are **L6** and **L7** while with GlideSP they are **L6** and **L14**. With **P3**, GlideSP identifies one additional binder in the top 1 % relative to the re-scoring protocols. These three protocols all use the AMBER softened vdW potential, OPLS partial charges for the protein, and MMFF94 partial charges for the ligand. Utilizing a softened vdW potential and similar protein partial charges to those used by Glide

Table 4 True positives identified in the top 1 % fraction screened by docking

Method	Protein structures used for docking							Unique hits across all grids
	P1	P2	P3	P4	P5	P6	P7	
GlideSP	L7	L6	L14	L14	L14	L6	–	L6
			L11	L6				L7
			L6					L11
								L14
MM-PB	L6	–	L4	–	–	–	–	*L4*
								L6
MM-PBSA	L6	L21	L4	–	L6	L6	L7	*L4*
	L7				L7		L6	L6
								L7
								L21
Cart-MM-PBSA	L6	–	L6	L7	L7	L6	L7	L6
	L7		L7		L6		L6	L7
MM-PB-a	L6	L7	L6	L6	L6	L6	L6	L7
	L7	L6	L7	L7	L7	L7	L7	L6
		L12			L12		L12	*L12*
MM-PBSA-a	L6	L7	L6	L6	L6	L6	L6	L7
	L7	L6	L7	L7	L7	L7	L7	L6
		L12			L12			*L12*
TorV-MM-PBSA-a	L6	L7	L6	L6	L6	L6	L6	L7
	L7	L6	L7	L7	L7	L7	L7	L6
		L12			L12	L12		*L12*

Ligands italicized are not identified in the top 1 % with GlideSP alone. All ligands listed in this table are shown in Fig. 2a

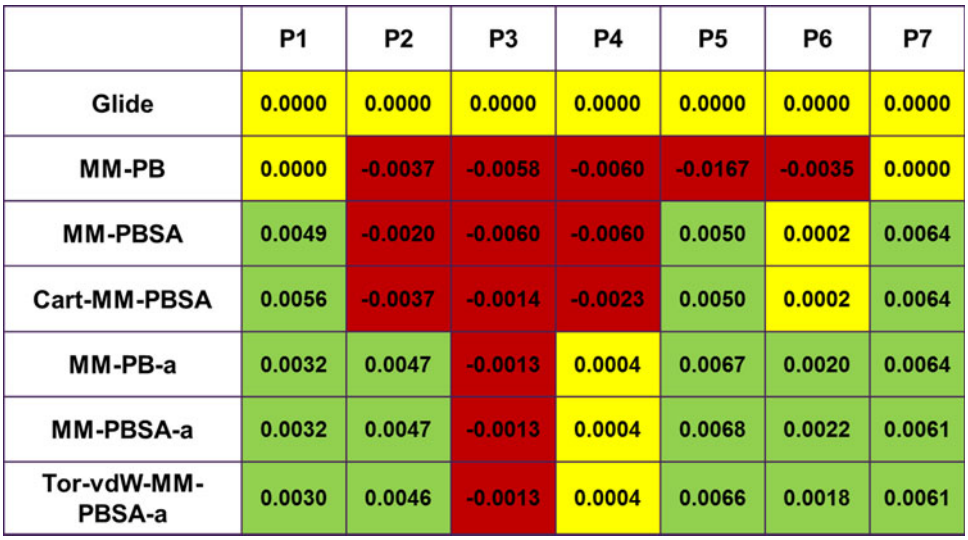


Fig. 7 A heat map of the relative AUC from enrichment plots for 1 % fraction screened, indicating the performance of each re-scoring protocol for early enrichment compared to GlideSP. *Green* highlights improved enrichment over GlideSP (AUC values that were at least 0.001 units greater than the reference GlideSP AUC), while *yellow* indicates the same enrichment as GlideSP, and *red* worse enrichment than GlideSP (AUC values that were at least 0.001 units less than the reference GlideSP AUC). The relative AUC values are given in each bin

(OPLS version 2001 [55] with Glide version 5.02) may contribute to these improved results. With these three protocols the force field is more similar to that used to generate the starting pose (GlideSP) in terms of the electrostatics. While different for the vdW term, a softened potential is employed, potentially lessening any detrimental effect of changing force fields. When comparing these three protocols, adding the SA term does not improve the

results, and the effect of minimization is negligible, again potentially in both cases due to the use of a softened vdW potential. In addition, to determine whether the re-scoring protocols were simply selecting for compounds with high MW, we computed the 1 % AUC value for ranking the screening library by MW. The average 1 % AUC value produced by the TorV-MM-PBSA-a protocol was 17 % larger than that produced when ranking by MW (0.0064 vs. 0.0053).

Both the self-docking (binding mode prediction) and virtual screening results reflect the balance between steric and electrostatic terms in the MM-PBSA scoring function. The principal difference between the MM-PB and the MM-PB-a classes of scoring functions is the functional form of the vdW term; the MM-PB functions used the MMFF vdW term, while the MM-PB-a functions used the DOCK/AMBER Lennard-Jones vdW term, with a softened repulsive potential. The MMFF vdW scores tended to be less favorable than the DOCK/AMBER vdW scores, and also tended to vary more greatly from one pose to the next. To determine whether the use of a softened repulsive potential with the DOCK/AMBER vdW model with in the MM-PB-a, MM-PBSA-a, and TorV-MM-PBSA-a protocols is responsible for the enhanced early enrichment demonstrated by those protocols, we repeated the TorV-MM-PBSA-a calculations using the standard 6-12 Lennard-Jones functional form. The results for these calculations on all seven protein structures reveal that both the 6-9 and 6-12 Lennard-Jones vdW terms produce nearly identical 1 % AUC values, with average values that differ by 2.7 % (0.0064 vs. 0.0062). By comparison, the average 1 % AUC values for the 7 GlideSP dockings are 47 % lower than the average 1 % AUC values produced by the TorV-MM-PBSA-a re-scoring protocol (0.0064 vs. 0.0034).

It is possible that the DOCK/AMBER vdW functional form and parameters are more highly correlated with the vdW terms used by the GlideSP score, so that the docked poses generated by Glide are already reasonably sterically optimized for the DOCK/AMBER MM force field. This may explain why the MM-PB-a class of scoring functions performs well for both pose prediction and virtual screening, without the need for minimization or the addition of the non-polar SA term. The MM-PB class of scoring functions, which use the MMFF force field, do benefit somewhat from the inclusion of the SA term and the use of a minimizer. Perhaps this is because the poses generated by the Glide docking are less intrinsically well-matched to the steric/vdW energy landscape of the MMFF force field, resulting in noisy or unfavorable vdW scores. Therefore, adding the SA term, which is much less sensitive to small changes in the ligand pose, and using the Cartesian minimizer to better match the pose geometry to the MMFF force field, improves the performance of the MM-PB scoring functions.

In summation, GlideSP performs well for this test set in terms of self-docking while for enrichment the performance varies depending on the protein structure used. With MM-PBSA re-scoring, the resulting early enrichment can be improved, but the choice of the force field used is important and changing the force field on going from docking to re-scoring seems to impact negatively the results. Minimization can help correct for this but the use of a reduced vdW term may be preferable and computationally more efficient. Of the eight re-scoring protocols evaluated in this study MM-PB-a, MM-PBSA-a, and TorV-MM-PBSA-a yielded the greatest enrichment. The addition of the non-polar SA term did not favorably impact the results, suggesting that simple models of the hydrophobic effect do not adequately describe protein–ligand binding. Torsional minimization of the softened vdW force field did improve the early enrichment in some cases. As such, TorV-MM-PBSA-a was overall the preferred re-scoring approach, at least for this system.

In practice, one may not have multiple complex structures with fragments bound to use for a virtual screen, and even if multiple structures are available there is not likely to be compelling a priori knowledge about which one would be best to use for docking. If only one of the protein structures was available, then the MM-PBSA re-scoring schemes may increase the odds of finding more actives, particularly in the first 1 %. Overall the value in MM-PBSA re-scoring of docking poses for virtual screening may be in hedging against the variation in the co-cyrstal structure that is used.

Acknowledgments We thank Steve Wesolowski for useful discussions and comments on the manuscript.

References

1. 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
2. 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
3. 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
4. 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
5. Albert JS, Blomberg N, Breeze AL, Brown AJ, Burrows JN, Edwards PD, Folmer RH, Geschwindner S, Griffen EJ, Kenny PW, Nowak T, Olsson LL, Sanganee 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
6. Erlanson DA, Wells JA, Braisted AC (2004) Tethering: fragment-based drug discovery. Ann Rev Biophys Biomol Struct 33:199–223
7. Carr RAE, Congreve M, Murray CW, Rees DC (2005) Fragment-based lead discovery: leads by design. Drug Discov Today 10:987–992
8. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9:430–431
9. Kuntz ID, Chen K, Sharp KA, Kollman PA (1999) The maximal affinity of ligands. Proc Natl Acad Sci USA 96:9997–10002
10. Abad-Zapatero C, Metz JT (2005) Ligand efficiency indices as guideposts for drug discovery. Drug Discov Today 10:464–469
11. 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
12. Oprea TI (2000) Property distribution of drug-related chemical databases. J Comput Aided Mol Des 14:251–264
13. 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
14. 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
15. Joseph-McCarthy D (2009) Challenges of fragment screening. J Comput Aided Mol Des 23:449–451
16. Marcou G, Rognan D (2006) Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. J Chem Inf Model 47:195–207
17. Teotico DG, Babaoglu K, Rocklin GJ, Ferreira RS, Giannetti AM, Shoichet BK (2009) Docking for fragment inhibitors of AmpC beta-lactamase. Proc Natl Acad Sci USA 106:7455–7460
18. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent developments in fragment-based drug discovery. J Med Chem 51:3661–3680
19. Southall NT, Dill KA, Haymet ADJ (2001) A view of the hydrophobic effect. J Phys Chem B 106:521–533
20. Ball P (2007) Water as an active constituent in cell biology. Chem Rev 108:74–108
21. Chandler D (2005) Interfaces and the driving force of hydrophobic assembly. Nature 437:640–647
22. Blokzijl W, Engberts JBFN (1993) Hydrophobic effects. Opinions and facts. Angew Chem Int Ed 32:1545–1579
23. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL (2004) Assessing scoring functions for protein − ligand interactions. J Med Chem 47:3032–3047
24. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3:935–949
25. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL (2004) Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 25:265–284
26. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33:889–897
27. Beuming T, Farid R, Sherman W (2009) High-energy water sites determine peptide binding affinity and specificity of PDZ domains. Prot Sci 18:1609–1619
28. Clausen RP, Naur P, Kristensen AS, Greenwood JR, Strange M, Brauner-Osborne H, Jensen AA, Nielsen AST, Geneser U, Ringgaard LM, Nielsen B, Pickering DS, Brehm L, Gajhede M, Krogsgaard-Larsen P, Kastrup JS (2009) The glutamate receptor GluR5 agonist (S)-2-Amino-3-(3-hydroxy-7,8-dihydro- 6H-cyclohepta[d]isoxazol-4-yl)propionic acid and the 8-methyl analogue: synthesis, molecular pharmacology, and biostructural characterization. J Med Chem 52:4911–4922
29. Abel R, Young T, Farid R, Berne BJ, Friesner RA (2008) Role of the active-site solvent in the thermodynamics of factor XA ligand binding. J Am Chem Soc 130:2817–2831
30. Shivakumar D, Deng Y, Roux B (2009) Computations of absolute solvation free energies of small molecules using explicit and implicit solvent model. J Chem Theory Comput 5:919–930
31. Barillari C, Taylor J, Viner R, Essex JW (2007) Classification of water molecules in protein binding sites. J Am Chem Soc 129:2577–2587
32. Sitkoff D, Sharp KA, Honig B (1994) Accurate calculation of hydration free energies using macroscopic solvent models. J Phys Chem 98:1978–1988
33. Massova I, Kollman PA (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspect Drug Discov Des 18:113–135
34. Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. Science 268:1144–1149
35. Gilson MK, Honig BH (1987) Calculation of electrostatic potentials in an enzyme active site. Nature 330:84–86
36. Edinger SR, Cortis C, Shenkin PS, Friesner RA (1997) Solvation free energies of peptides: comparison of approximate continuum solvation models with accurate solution of the poisson—boltzmann equation. J Phys Chem B 101:1190–1197
37. Levy RM, Zhang LY, Gallicchio E, Felts AK (2003) On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute−solvent interaction energy. J Am Chem Soc 125:9523–9530
38. Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 47:409–443
39. Jain AN (2006) Scoring functions for protein-ligand docking. Curr Protein Pept Sci 7:407–420
40. Thompson DC, Humblet C, Joseph-McCarthy D (2008) Investigation of MM-PBSA rescoring of docking poses. J Chem Inf Model 48:1081–1091
41. 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
42. Huang N, Kalyanaraman C, Irwin JJ, Jacobson MP (2005) Physics-based scoring of protein − ligand complexes: enrichment of known inhibitors in large-scale virtual screening. J Chem Inf Model 46:243–253
43. Kawatkar S, Wang H, Czerminski R, Joseph-McCarthy D (2009) Virtual fragment screening: an exploration of various docking and scoring protocols for fragments using Glide. J Comput Aided Mol Des 23:527–539
44. Ikai K, Ujihara M, Fujii K, Urade Y (1989) Inhibitory effect of tranilast on prostaglandin D synthetase. Biochem Pharmacol 38:2673–2676
45. 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
46. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105:6474–6487
47. Weininger D (1988) SMILES 1. Introduction and encoding rules. J Chem Inf Comput Sci 28:31–36
48. Kenny PW, Sadowski J (2004) Structure modification in chemical databases. In: Opera TI (ed) Chemoinformatics in drug discovery. Wiley, Weinheim, pp 271–285
49. 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
50. 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
51. Meng EC, Shoichet BK, Kuntz ID (1992) Automated docking with grid-based energy evaluation. J Comput Chem 13:505–524
52. Meng EC, Gschwend DA, Blaney JM, Kuntz ID (1993) Orientational sampling and rigid-body minimization in molecular docking. Proteins: Struct, Funct, Bioinf 17:266–278
53. Bondi A (1964) van der Waals volumes and radii. J Phys Chem 68:441–451
54. Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem 17:490–519
55. Halgren TA (1996) Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. J Comput Chem 17:520–552
56. Tanford C (1979) Interfacial free energy and the hydrophobic effect. Proc Natl Acad Sci USA 76:4175–4176
57. 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
58. Sing T, Sander O, Beerenwinkel N, Lengauer T (2005) ROCR: visualizing classifier performance in R. Bioinformatics 21:3940–3941
59. Wildman SA, Crippen GM (1999) Prediction of physicochemical parameters by atomic contributions. J Chem Inf Comput Sci 39:868–873