J Comput Aided Mol Des (2008) 22:621–627
DOI 10.1007/s10822-008-9182-y

Improving database enrichment through ensemble docking

Shashidhar Rao · Paul C. Sanschagrin · Jeremy R. Greenwood · Matthew P. Repasky · Woody Sherman · Ramy Farid

Received: 6 November 2007 / Accepted: 15 January 2008 / Published online: 6 February 2008

Abstract While it may seem intuitive that using an ensemble of multiple conformations of a receptor in structure-based virtual screening experiments would necessarily yield improved enrichment of actives relative to using just a single receptor, it turns out that at least in the p38 MAP kinase model system studied here, a very large majority of all possible ensembles do not yield improved enrichment of actives. However, there are combinations of receptor structures that do lead to improved enrichment results. We present here a method to select the ensembles that produce the best enrichments that does not rely on knowledge of active compounds or sophisticated analyses of the 3D receptor structures. In the system studied here, the small fraction of ensembles of up to 3 receptors that do yield good enrichments of actives were identified by selecting ensembles that have the best mean GlideScore for the top 1% of the docked ligands in a database screen of actives and drug-like "decoy" ligands. Ensembles of two receptors identified using this mean GlideScore metric generally outperform single receptors, while ensembles of three receptors identified using this metric consistently give optimal enrichment factors in which, for example, 40% of the known actives outrank all the other ligands in the database.

Keywords Enrichment · Ensemble docking · Virtual screening · p38 MAP kinase · Glide

S. Rao · P. C. Sanschagrin · J. R. Greenwood · M. P. Repasky · W. Sherman · R. Farid (✉)
Schrödinger, Inc, 120 West 45th Street, New York, NY 10036, USA
e-mail: ramy@schrodinger.com

Introduction

Docking is an integral part of structure-based drug discovery. Traditional docking and scoring methods employ the rigid receptor paradigm in which the receptor structure, once prepared, is not altered during the process of docking, while the conformational flexibility of the ligand is fully explored [1–6]. However, the available X-ray structures in the Protein Data Bank (PDB) [7] demonstrate significant evidence for ligand induced conformational flexibility in the binding sites of proteins. Such flexibility is manifested in a variety of ways, ranging from subtle movements in the backbone to significant variations in side-chain conformations and extensive loop movements. Lack of consideration of such flexibility in docking generally leads to inaccuracies in pose predictions, reduced recovery of actives, and by implication poorer correlation with binding affinities [8]. In a previous publication, we described the treatment of induced fit effects that involve small to medium range changes in the flexibility of the active site in terms of a workflow combining the docking program (Glide) and protein refinement program (Prime) [9]. It was shown that the methods described in that protocol are very useful in predicting correct poses of bound ligands by accounting for induced fit effecting such as small movements in the backbone and significant changes in side-chain conformations of protein residues. Other groups have also developed protocols for accounting for protein flexibility in docking [10–14]. However, modeling large-scale changes in backbone conformations of proteins, including large loop movements, present significant challenges and therefore may require alternative methods. One such method involves docking into multiple conformations of the receptor, often referred to as ensemble docking [15–19]. We present here an ensemble docking approach that can

lead to improved enrichment of actives in database screening. A simple method is described for choosing the optimal ensemble of protein structures for virtual screening studies using p38 MAP kinase structures from the PDB as a model system.

Methods

Receptor preparation

Thirty-three PDB complexes of p38 MAP kinase structures with bound ligands (1A9U, 1BL6, 1BL7, 1BMK, 1DI9, 1IAN [20] 1KV1, 1KV2, 1M7Q, 1OUK, 1OUY, 1OVE, 1OZ1, 1W7H, 1W82, 1W83, 1W84, 1WBN, 1WBO, 1WBS, 1WBT, 1WBV, 1WBW, 1YQJ, 1YW2, 1YWR, 1ZYJ, 1ZZ2, 1ZZL, 2BAJ, 2BAK, 2BAL and 2BAQ) were prepared using the protein preparation workflow in Maestro v8.0 [21]. Of these, ten had the activation loop in the so-called DFG-out configuration in which ligands are found to bind in an allosteric binding pocket adjacent to the ATP binding site (Fig. 1a) [22]. The other 23 structures had this loop in the so-called DFG-in configuration in which ligands bind to the ATP binding site (Fig. 1b). Hydrogens were added to the protein structures consistent with a pH of 7. Since none of the crystal structures had water-mediated protein–ligand interactions, all the water molecules were deleted. Hydrogens were also added to the ligand molecules followed by the generation of energetically accessible ionization and tautomeric states using Epik v1.5 [23]. The protein assignment program in Maestro was used to automatically set terminal rotameric states for Asn, Gln, and His as well as tautomeric and protonation states of His to optimize the hydrogen-bonding network in the complex. In addition, hydroxyl and thiol torsions in Cys, Ser, and Tyr were optimized. Finally, the protein–ligand complexes were subjected to a restrained minimization using the "impref" tool available in Protein Preparation Wizard in Maestro v8.0. An RMSD cutoff value of 0.3 Å was used.

Ligand preparation

Two sets of p38 MAPK ligands were used in this study: (1) X-ray ligands from the 33 p38 MAP kinase PDB complexes listed above were used to access the ability of each receptor to dock the ligands accurately in cross-docking studies and (2) a collection of 28 known p38 MAP kinase inhibitors from the literature with pIC_{50} values ranging from 4.6 to 10 were used for the database enrichment studies. The 28 p38 MAP kinase inhibitors were chosen in the following manner: a collection of 127 p38 MAP kinase inhibitors with activity values spanning 6 orders of magnitude were binned into

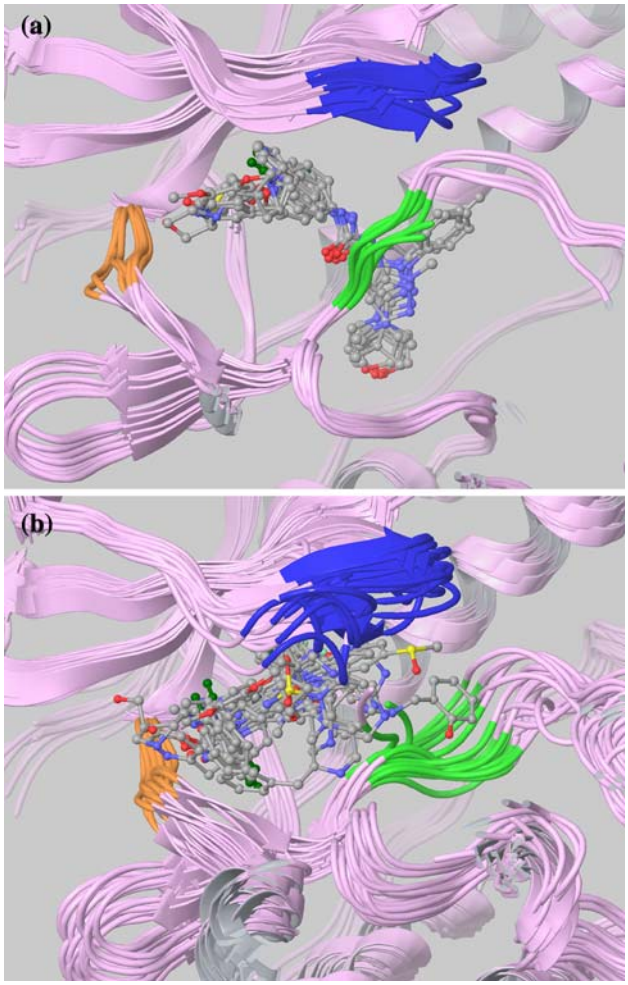

Fig. 1 (**a**) Superposition of 10 DFG-out p38 MAP kinase structures from the PDB showing the backbone structure and bound ligands, characterized by inhibitors bound to the secondary, allosteric binding pocket with hydrogen bonds to the side chain of Glu71 and backbone of Asp168. The hinge region is indicated in orange, the DFG portion of the activation loop in green, and the Gly-rich P-loop in blue. (**b**) Superposition of 23 DFG-in p38 MAP kinase structures from the PDB. Coloring is the same as in (**a**)

pIC_{50} bins of size one starting from 4. Of the 6 occupied bins, the middle 4 had between 20 and 40 ligands each; from each of these well-populated bins, 5 compounds were chosen randomly. All the ligands in the remaining 2 bins were chosen (1 from pIC_{50} bin 4–5 and 5 from pIC_{50} bin 9–10). An additional two compounds were chosen that had reported pIC_{50} values of exactly10.

The Glide decoy set of 1,000 molecules with "drug-like" characteristics and molecular weight less than or equal to 400 was used for the database enrichment studies [24]. All ligands were prepared using LigPrep [25] by generating low energy ionization and tautomeric states within the range of pH 7.0 ± 2.0. In the case of the X-ray ligands and any analogs in set 2, chiralities were retained from the X-ray structure. All molecules subjected to

LigPrep were energy minimized using the OPLS_2005 force field, leading to a initial docking conformation distinct from the crystal structure geometry.

Docking accuracy

Energy grids for docking studies were computed for each of the prepared complexes using default settings in the Glide [26]. The centroid of the bound crystallographic ligand was used to define the center of each grid box. For the purposes of assessing docking accuracy, all possible self and cross-docking studies were carried out with the extra precision (XP) scoring function [5]. The starting PDB structures were all superimposed on the 1A9U structure using the Protein Structure Alignment tool in Maestro to allow for direct RMSD comparisons. Due to uncertainties in the exact alignment of the binding sites resulting from the protein superposition, any cross-docking RMSD values are upper bounds, which makes a 2.0 Å metric for docking accuracy more stringent than in native docking studies. In all results presented here only the top-ranked structure for each compound was retained in the output.

Enrichment

Docking using the standard precision (SP) mode of Glide was performed on each of the prepared PDB structures. For ensembles of multiple structures, the pose and score for a given compound was chosen using the top GlideScore for the compound when docked into any member of the ensemble. Enrichment calculations were then performed following the method for EF′ previously described [4], using the following equation:

$$\mathrm{EF}'(N) = \left(50\%/\mathrm{APR}_{\mathrm{sampled}}\right) \times \left(\mathrm{Hits}_{\mathrm{sampled}}/\mathrm{Hits}_{\mathrm{total}}\right)$$

where $\mathrm{APR}_{\mathrm{sampled}}$ is the average percentile rank of the $\mathrm{Hits}_{\mathrm{sampled}}$ (retrieved known actives) within the subset of the top N percent of the actives. For example, if there were 100 known actives and 1,000 decoy ligands, the value of *EF′(5)* would be 6.25, if 5 (5%) of the known actives were ranked 1, 2, 3, 6, and 10:

$$\mathrm{EF}'(5) = \frac{50\%}{\frac{(1+2+3+6+10)/1100}{5}} \times (5/100) = 6.25$$

This formula adjusts the standard EF calculation to give higher values when the known actives are concentrated at the higher ranks within the chosen recovery slice. The results here are reported in terms of the enrichment for the recovery of 40% of the actives [*EF′(40)*].

Ensembles of one to three receptors were chosen on the basis of the mean GlideScore of the top 1% of all of the docked compounds (decoys plus active compounds). This measure is calculated by first determining the ensemble score for each ligand, defined as the score against the best single receptor of the ensemble set. Then, the ensemble scores for the top 1% of compounds, here 10 compounds out of the 1,028 decoys plus actives, are extracted and averaged. We will refer to this metric as the *MGS-1*. The top ensembles ranked on *EF′(40)* were also extracted; these correspond to the optimal choice of the ensemble members once the set of actives is known and provide a reference to judge the performance when choosing ensembles based on the mean GlideScore metric.

In order to be able to compare GlideScores between ligands bound to the DFG-in and DFG-out receptor structures, it is necessary to account for the higher receptor reorganization energy associated with binding to the DFG-out receptors [27]. The reorganization energy would be applied as a positive offset to the GlideScore for ligands docked into any of the DFG-out structures. The precise value of this reorganization energy is difficult to determine directly, but it can be estimated by assuming that the experimentally measured activities of compounds that bind to DFG-in and DFG-out forms of the p38 are similar. Given this assumption, the offset can be estimated as the difference in the *MGS-1* between DFG-in and DFG-out receptors, since the top 1% of hits are likely to be mostly actives. The *MGS-1* for DFG-in receptors is −8.2 and for DFG-out receptors its −9.3. Therefore, the difference of 1.1 kcal/mol, an estimate of the reorganization energy associated with binding to the DFG-out structures, was added to the GlideScores of all ligands docked into DFG-out structures.

Results

Docking accuracy

Table 1 shows the RMSD values of the docked poses relative to the X-ray poses for the DFG-out PDB ligands. All these ligands dock into the 1KV2 structure with RMSD < 1.7 Å. Nine of the ligands dock into 1KV1, 1WBN and 2BAK with RMSD < 2.0 Å. All the DFG-out receptors dock at least 7 of the ligands with an RMSD of ≤1.9 Å. The ability of all the receptors to dock most of the DFG-out ligands accurately is a reflection of relatively small-induced fit effects observed for this particular set of receptors (see Fig. 1a).

Table 2 shows the 23 × 23 cross-docking matrix with RMSD values for the DFG-in docked poses relative to the X-ray structures. Table 2 also reports the average GlideScore for all ligands docked into each PDB structure. Unlike

Table 1 Root mean square deviations of the top ranked pose for DFG-out ligands in the cross-docking studies of p38 MAP kinase inhibitors

Ligand	Protein									
	1KV1	1KV2	1W82	1W83	1WBN	1WBS	1WBT	1WBV	2BAJ	2BAK
1kv1	**0.3**	*0.7*	*0.6*	8.4	*1.0*	8.3	8.3	*0.8*	*0.9*	*1.2*
1kv2	*1.3*	**0.3**	*0.7*	*1.7*	*1.5*	*1.4*	*1.6*	*2.0*	*1.1*	*1.9*
1w82	*0.6*	*0.5*	**0.3**	*0.9*	*1.5*	*0.6*	*0.9*	*0.5*	*0.6*	*1.2*
1w83	2.3	*1.0*	6.7	**0.4**	*2.0*	*0.9*	*0.5*	*0.7*	3.1	2.2
1wbn	*1.4*	*1.5*	*1.7*	*0.3*	**0.3**	*0.8*	*0.8*	*1.0*	*1.8*	*1.0*
1wbs	*1.5*	*1.1*	11.3	*0.7*	*1.6*	**0.3**	*0.5*	2.6	*1.5*	*1.7*
1wbt	*1.4*	*1.0*	*1.3*	*0.6*	2.1	*0.5*	**0.5**	2.5	*1.4*	*1.1*
1wbv	*1.1*	*1.6*	*1.4*	*0.7*	*0.8*	*1.4*	*0.5*	**1.4**	*1.0*	*1.5*
2baj	*1.8*	*1.7*	*0.3*	4.5	*1.3*	*1.8*	*0.9*	*0.6*	**0.7**	*1.0*
2bak	*1.7*	*1.2*	10.9	3.0	*1.8*	12.0	11.7	*1.9*	11.3	**1.8**
# with RMSD $\leq$ 2 Å	9	10	7	7	9	8	8	8	8	9
Avg. GlideScore	−12.1	−12.2	−11.4	−11.6	−11.2	−12.6	−12.4	−11.2	−12.4	−11.1

Table 2 Root mean square deviations of the top ranked pose for DFG-in ligands in the cross docking studies of p38 MAP kinase inhibitors. Dashes indicate cases where cross-docking produced no ligand pose. The bold numbers represents the values for self-docking

	Protein																						
Ligand	1A9U	1BL6	1BL7	1BMK	1DI9	1IAN	1M7Q	1OUK	1OUY	1OVE	1OZ1	1W7H	1W84	1WBO	1WBW	1YQJ	1YW2	1YWR	1ZYJ	1ZZ2	1ZZL	2BAL	2BAQ
1a9u	**0.7**	2.8	2.2	5.3	2.0	2.8	8.3	2.5	1.8	3.9	1.0	7.0	1.8	2.7	6.9	5.4	8.2	2.4	2.2	5.1	2.7	3.8	9.7
1bl6	1.1	**1.6**	1.0	1.1	6.5	6.9	1.4	6.6	2.0	2.4	1.3	6.2	1.1	1.5	–	1.4	2.3	2.1	1.5	1.1	6.7	2.3	5.0
1bl7	0.9	2.7	**0.7**	0.7	1.8	2.7	2.2	3.0	1.3	2.1	1.8	7.8	1.7	1.4	4.9	1.3	2.4	2.6	2.1	2.0	5.5	2.6	6.6
1bmk	0.4	2.3	0.7	**0.4**	2.2	3.0	2.2	1.8	2.5	2.9	1.7	4.8	1.5	0.8	6.6	2.0	2.5	2.0	2.5	2.4	6.9	2.9	9.0
1di9	1.1	5.6	1.0	8.6	**0.5**	7.5	1.9	7.1	4.6	6.2	0.9	7.2	1.2	4.3	1.6	7.0	2.5	2.1	2.7	2.8	2.4	2.8	2.8
1ian	2.9	2.6	2.4	5.9	3.2	**0.2**	6.6	1.1	6.6	1.9	6.6	6.1	3.3	2.8	6.2	6.8	7.7	2.2	2.6	8.2	4.6	1.6	8.8
1m7q	2.8	2.1	1.7	1.9	2.8	1.6	**0.7**	0.8	1.3	0.6	2.3	11.2	7.3	1.4	–	2.3	–	1.7	1.3	1.4	–	0.9	9.0
1ouk	2.5	2.9	8.4	7.8	2.9	1.4	2.1	**0.4**	2.2	1.9	2.4	6.3	2.5	2.2	9.5	1.6	7.2	1.1	1.4	8.1	–	2.0	8.7
1ouy	5.2	4.8	3.5	5.9	7.5	2.3	1.8	1.5	**1.1**	1.0	3.7	10.2	8.4	5.7	12.6	3.3	8.1	1.9	1.6	1.5	–	2.0	9.3
1ove	2.3	1.2	7.3	1.7	7.0	1.2	0.9	1.0	0.7	**0.4**	5.6	10.8	2.8	1.2	–	3.3	–	1.3	0.9	1.3	–	1.2	9.7
1oz1	0.7	5.7	0.5	4.6	0.9	2.0	1.4	1.9	1.2	2.0	**0.3**	6.0	0.8	0.8	5.0	1.5	6.7	1.8	1.9	4.9	–	5.3	9.3
1w7h	0.6	1.9	1.8	0.8	1.2	3.0	2.2	2.5	1.7	2.1	1.9	**0.4**	1.0	0.7	1.2	0.8	2.4	2.4	2.4	2.6	2.6	2.4	2.2
1w84	0.8	0.3	1.3	0.6	2.5	1.2	1.7	1.1	1.8	1.6	0.7	0.9	**0.6**	0.6	2.3	0.8	1.7	1.5	2.9	6.4	3.1	1.5	1.7
1wbo	2.4	2.4	3.6	2.3	4.8	2.2	2.9	2.5	2.3	3.0	3.4	6.0	2.3	**0.4**	2.3	2.4	2.0	2.8	3.1	2.5	3.5	2.9	1.7
1wbw	0.9	0.4	0.4	0.3	0.8	3.0	0.5	1.2	1.7	2.7	0.6	1.1	0.6	2.3	**0.6**	0.7	6.6	1.4	1.6	6.6	1.6	1.4	1.6
1yqj	4.8	0.6	1.6	1.0	2.5	1.7	1.8	2.4	5.7	4.9	5.0	8.7	4.2	0.6	13.8	**0.2**	3.1	1.3	3.0	3.7	–	2.7	10.1
1yw2	2.3	2.2	1.9	8.7	6.5	5.6	1.8	3.1	1.6	2.1	2.9	7.5	8.0	6.9	7.6	7.7	**0.6**	1.4	7.2	1.2	2.6	1.7	8.3
1ywr	3.1	2.7	2.8	7.3	2.6	1.2	1.2	0.9	1.8	2.0	2.5	7.5	2.2	2.5	8.8	1.5	7.4	**0.3**	2.5	7.9	–	2.8	8.7
1zyj	9.9	10.0	10.3	10.2	2.9	2.0	3.2	3.0	8.5	1.9	3.2	9.0	3.2	2.7	3.4	3.3	8.3	2.1	**0.5**	1.7	8.5	8.1	1.6
1zz2	3.2	7.8	6.8	5.8	4.0	3.7	3.3	2.3	3.6	6.9	3.5	9.4	5.3	7.9	9.5	7.6	6.1	4.0	3.2	**0.2**	10.1	4.0	8.1
1zzl	3.0	3.2	2.8	7.3	2.6	2.0	1.3	1.7	0.8	0.5	2.3	3.7	1.9	1.9	1.8	1.3	1.0	0.6	0.9	0.8	**0.4**	1.1	0.6
2bal	9.1	10.8	9.4	4.4	2.8	4.6	7.1	16.4	10.9	8.6	7.4	3.2	6.2	3.9	3.6	6.0	3.9	8.8	0.7	5.6	9.4	**1.1**	1.6
2baq	10.2	10.7	13.0	9.9	9.7	4.7	4.9	9.8	9.3	2.8	10.0	5.0	7.6	5.7	3.8	9.9	4.3	1.4	4.4	7.3	2.4	1.2	**0.9**
# with RMSD ≤2 Å	9	6	11	9	6	10	12	11	13	10	9	3	10	11	4	11	4	13	10	9	2	11	7
Avg. Glide Score	-8.4	-8.4	-9.7	-8.5	-8.6	-10.8	-9.8	-10.4	-9.7	-11.3	-10.1	-8.0	-9.2	-8.2	-8.1	-10.1	-9.0	-11.2	-9.6	-9.7	-5.6	-10.2	-8.2

in the case of the DFG-out protein structures, none of the DFG-in protein structures is able to dock all the DFG-in ligands to within 2.0 Å of the X-ray poses. This can be attributed at least in part to the additional flexibility observed around the DFG-in binding site relative to the DFG-out binding site (see Fig. 1b). Much of the DFG-in variability can be attributed to the Gly-rich P-loop. The best structures for docking accuracy are 1OUY and 1YWR; each docks 13 ligands to within 2.0 Å of the corresponding X-ray poses.

A plot of the average GlideScore versus the number of ligands with RMSD $\leq$ 2Å (Fig. 2) reveals a surprisingly good correlation ($R^2 = 0.56$). In other words, receptors that tend to give better (more negative) GlideScores also tend to dock more ligands within 2 Å of the X-ray poses. This is the expected result if the GlideScore is a reasonable metric for selecting correctly docked ligands. It also suggests that ensembles of receptor structures that yield the best GlideScores for a set of actives (known or otherwise) would also lead to optimal enrichment factors in virtual database screening (see below).

Enrichment studies

The enrichment results, obtained by docking 28 known active ligands and 1,000 decoy ligands (described in the Methods section) into ensembles of up to 3 structures selected from among the full set of 33 PDB structures, are given in Table 3. Only the top ranked ensembles with

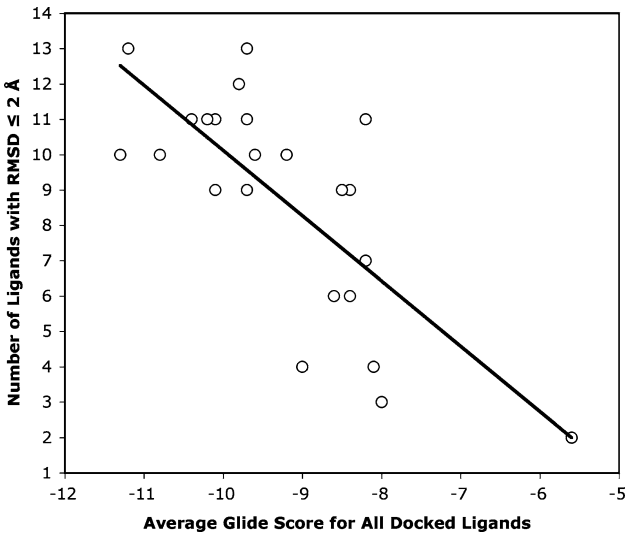

Fig. 2 Plot of the number of ligands from 23 DFG-in X-ray structures that dock with an RMSD of $\leq$2 Å as a function of the average GlideScore for all docked ligands. The correlation (R^2 value) is 0.56

Table 3 Enrichments of p38 MAP kinase actives obtained for ensembles selected from 33 PDB structures using the *MGS-1* metric

MGS-1	*MGS-1* rank	Receptor/Ensemble	*EF′(40)*	*EF′(40)* rank
−10.8	1	1ove	25.5	2
−9.5	2	1yw2	24.4	3
−9.1	3	1ian	12.1	8
−8.9	4	1ywr	20.6	4
−8.8	5	1m7q	27.1	1
−11.1	1	1ian-1ove	33.7	1
−10.9	2	1ove-1yw2	33.7	1
−10.9	2	1ove-1zzl	24.7	21
−10.9	2	**1kv1**-1ove	22.9	27
−10.9	2	1bl7-1ove	32.2	4
−11.1	1	1ian-1ove-1yw2	33.7	1
−11.1	1	1ian-1ove-1zzl	33.7	1
−11.1	1	1ian-**1kv2**-1ove	33.7	1
−11.1	1	1ian-1ove-**1w83**	33.7	1
−11.1	1	1ian-1ove-**1wbn**	33.7	1
−11.1	1	1ian-1ove-**1wbs**	33.7	1
−11.1	1	1ian-1ove-**2baj**	33.7	1
−11.1	1	1ian-1ove-2baq	33.7	1
−11.1	1	1ian-1ove-2bal	33.7	1
−11.1	1	1ian-1ove-**2bak**	33.7	1
−11.1	1	1ian-1ove-1zz2	33.7	1
−11.1	1	1ian-1ove-1zyj	33.7	1
−11.1	1	1ian-1ove-1ywr	33.7	1
−11.1	1	1ian-1ove-1yqj	33.7	1
−11.1	1	1ian-1ove-**1wbv**	33.7	1
−11.1	1	1ian-1ove-**1wbt**	33.7	1
−11.1	1	1ian-1ove-1wbo	33.7	1
−11.1	1	1ian-1ove-1w84	33.7	1
−11.1	1	1ian-1ove-**1w82**	33.7	1
−11.1	1	1ian-1ove-1w7h	33.7	1
−11.1	1	1ian-1ove-1oz1	33.7	1
−11.1	1	1ian-1ouy-1ove	33.7	1
−11.1	1	1ian-1ouk-1ove	33.7	1
−11.1	1	1ian-1m7q-1ove	33.7	1
−11.1	1	1di9-1ian-1ove	33.7	1
−11.1	1	1bmk-1ian-1ove	33.7	1
−11.1	1	1bl7-1ian-1ove	33.7	1
−11.1	1	1bl6-1ian-1ove	33.7	1
−11.1	1	1a9u-1ian-1ove	33.7	1
−11.1	1	1ian-**1kv1**-1ove	32.2	4
−11.1	1	1ian-1ove-1wbw	32.2	4

DFG-out receptors are bolded. The maximum *EF′(40)* for this dataset is 33.7, obtained if no decoy ligands outscore 40% of the known actives in the database. The *EF′(40)* rank is that of the individual ensembles sizes. There are 528 possible 2-receptor ensembles, and 5,456 possible 3-receptor ensembles

respect to the *MGS-1* metric (mean GlideScore for the top 1% of docked compounds; see Methods for more details) are shown. In the case of screening using a single receptor, the receptor with the best *MGS-1* (1ove; −10.8) obtained the second best *EF′(40)*. The best *EF′(40)* of 27.1 is obtained for the receptor with the 5th best *MGS-1* value (1m7q). Only 12% of the receptors receive an *EF′(40)* better than 17 and these are mostly concentrated among the receptors that receive the best *MGS-1* values. For ensembles of two receptors, the results are similar in that ensembles that receive the best *MGS-1* values also correspond to the receptors that give the best *EF′(40)* values. The ensemble 1ian/1ove receives the best *MGS-1* value of −11.1 and also gives the best *EF′(40)* value of 33.7. In fact, 33.7 is the maximum *EF′(40)* value that can be obtained and is the result of having no decoy ligands outrank 40% of the actives. A large majority (85%) of the 528 ensembles of two receptors gives an *EF′(40)* less than 22.9 (the lowest *EF′(40)* value observed for the top ensembles reported in Table 3), so identifying 5 ensembles using the *MGS-1* metric that yield such high *EF′(40)* values is something that clearly cannot happen just by chance. Increasing the ensemble size to 3 does not have much of an effect on the ability of ensembles with good *MGS-1* values to give good enrichments. Twenty-nine of the top 31 ensembles that receive the best *MGS-1* value of −11.1 give *EF′(40)* values of 33.7 and the other two give an *EF′(40)* of 32.2. Only 4.1% of the 3-receptor ensembles gives an *EF′(40)* of 32.2 or better, so once again, having all the top ensembles with respect to *MGS-1* give *EF′(40)* values of 32.2 or greater shows very clearly that *MGS-1* is an excellent metric for selecting ensembles that give high (and in most cases the maximum) enrichment factors. The bar graph in Fig. 3 summarizes these results and shows that the best *EF′(40)* values are generally obtained by ensembles with the best (lowest) *MGS-1* values. The histogram in Fig. 4 highlights how difficult it is to identify ensembles that yield good *EF′(40)* values.

It's interesting to ask whether retrospective examination of the receptor conformations that make up the ensembles with the best *MGS-1* values reveal the reason why these ensembles yield good enrichments while ones with poor *MGS-1* values do not yield good enrichments. Before considering this question, it should be noted that all prospective attempts to use structure only to select ensembles failed to produce ensembles with good enrichments. A crude examination of the 33 receptor conformations suggests that there are approximately 17 distinct clusters of conformations. These arise from 4 distinct conformations for Tyr35 (Gly-rich P-loop), 3 conformations for Met109 (hinge region), 6 conformations for Glu168 (activation loop), and 3 conformations for Phe169 (activation loop). One might expect that the ensembles of 2 and 3 receptor conformations would consist of conformations from different clusters; however, that turns out to not be the case. That is, ensembles that consist of conformations from different clusters are just as likely to give a good (or poor) enrichment as ensembles that consist of conformations from the same cluster. A more careful examination of the receptor conformations reveals that there are in fact nearly as many clusters as receptor conformations. The only two structures that are truly similar in the active site are 1WBS and 1WBT (side chain RMSD for Gly33, Tyr35, Met109, Glu168 and Phe169 = 0.21 Å). Pairs 1BL6/1BMK and 1OUK/1M7Q are quite similar to one another, but in both cases differences in the conformation of Tyr35 in the Gly-rich P-loop are large enough (for both pairs, Tyr35 side

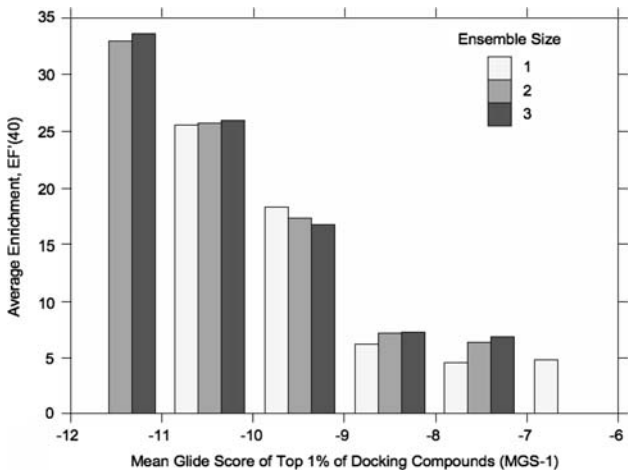

Fig. 3 Bar graph showing the average enrichment of actives obtained for ensembles of 1–3 protein structures as a function of the *MGS-1* (mean GlideScore for the top 1% of docked ligands, known actives and decoys)

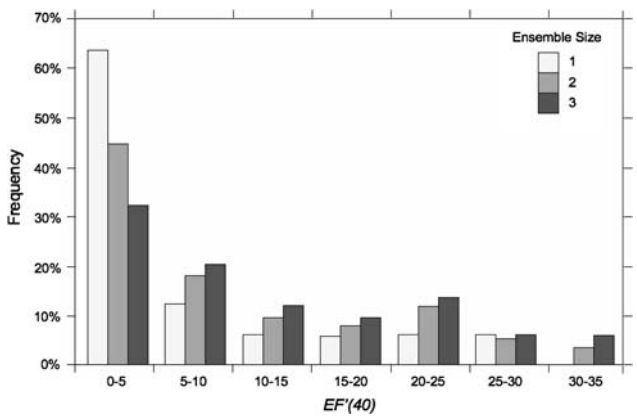

Fig. 4 Histogram showing the percent frequency for ranges of *EF′(40)* values for virtual screens using a single receptor as well as 2- and 3-receptor ensembles. The histogram shows that none of the single receptors and a very small percentage of the ensembles yield the maximum *EF′(40)* value

chain RMSD = 1.0 Å) to have an impact on the GlideScores of docked ligands and thus enrichments. Aside from these two pairs, all the other receptor conformations are different enough from one another (side chain RMSD for Gly33, Tyr35, Met109, Glu168 and Phe169 $\geq$ 0.9 Å) such that it is not possible to explain the enrichment results by prospective or retrospective examination of the structures. These results suggest that combinations of subtly different receptor conformations are what lead to good enrichments.

Conclusions

The results presented here, in which known p38 MAP kinase inhibitors and a set of drug-like decoy ligands were docked to all possible ensembles of up to 3 protein structures, shows that a large majority of the possible ensembles yield poor enrichments. Thus, randomly choosing an ensemble is very likely to yield unsatisfactory enrichments and even more unlikely to yield the best possible enrichment. Extensive efforts to identify ensembles that yield good *EF'(40)* values through use of several structure-based clustering methods have all failed. However, selecting ensembles that give the best mean GlideScore for the top 1% of a database screen appears to be a robust way of identifying ensembles that yield the maximum possible enrichment. This metric works as well as it does because correctly docked active ligands are generally associated with good GlideScores, therefore, ensembles that give good GlideScores for the top 1% of the hits from a virtual screen are more likely to contain a high fraction of actives among the top scoring ligands, thus leading to good enrichment factors. For the particular example of p38 MAP kinase studied here, ensembles of 2 receptors yield better enrichments than any single receptor. Increasing the size of the ensemble to 3 receptors increases the probability of obtaining the optimal enrichment factor. Studies are currently underway to determine whether the mean GlideScore for the top 1% of docked ligands in a database screen (or a smaller percentage to account for the fact that some databases many not contain as many actives as in the current study) is a metric that can be reliably applied to other target classes as a means of identifying ensembles that yield optimal enrichments.

References

1. Leach AR, Shoichet BK (2006) J Med Chem 49:5851
2. Perola E, Walters WP, Charifson PS (2004) Proteins 56:235
3. 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) J Med Chem 47:1739
4. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) J Med Chem 47:1750
5. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) J Med Chem 49:6177
6. Zhou Z, Felts AK, Friesner RA, Levy RM (2007) J Chem Inf Model 47:1599
7. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) J Mol Biol 112:535
8. Sutherland JJ, Nandigam RK, Erickson JA, Vieth M (2007) Lessons in molecular recognition. 2. Assessing and improving cross-docking accuracy. J Chem Inf Model
9. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R (2006) J Med Chem 49:534
10. Bastard K, Prevost C, Zacharias M (2006) Proteins 62:956
11. Cavasotto CN, Kovacs JA, Abagyan RA (2005) J Am Chem Soc 127:9632
12. Schnecke V, Swanson CA, Getzoff ED, Tainer JA, Kuhn LA (1998) Proteins 33:74
13. Zavodszky MI, Lei M, Thorpe MF, Day AR, Kuhn LA (2004) Proteins 57:243
14. Alberts IL, Todorov NP, Dean PM (2005) J Med Chem 48:6585
15. Limongelli V, Marinelli L, Cosconati S, Braun HA, Schmidt B, Novellino E (2007) ChemMedChem 2:667
16. Huang SY, Zou X (2007) Proteins 66:399
17. Claussen H, Buning C, Rarey M, Lengauer T (2001) J Mol Biol 308:377
18. Lorber DM, Shoichet BK (1998) Protein Sci 7:938
19. Polgar T, Keseru GM (2006) J Chem Inf Model 46:1795
20. Only CA atoms are available in the PDB, however, one of the authors of the 1IAN structure (Liang Tong) generously provided us with a refined, all-atom model
21. Maestro v8.0, Schrödinger, Inc.: Portland, OR
22. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, Grob PM, Hickey ER, Moss N, Pav S, Regan J (2002) Nat Struct Biol 9:268
23. Epik v1.5, Schrödinger, Inc.: Portland, OR
24. The 1000 drug-like set of decoy ligands is available for download from the Schrödinger website
25. LigPrep v2.1, Schrödinger, Inc.: Portland, OR
26. Glide v4.5, Schrödinger, Inc.: Portland, OR
27. Frembgen-Kesner T, Elcock AH (2006) J Mol Biol 359:202