J Comput Aided Mol Des (2010) 24:433–447
DOI 10.1007/s10822-010-9356-2

Chemical space sampling by different scoring functions and crystal structures

Natasja Brooijmans · Christine Humblet

Received: 6 October 2009 / Accepted: 5 April 2010 / Published online: 18 April 2010

Abstract Virtual screening has become a popular tool to identify novel leads in the early phases of drug discovery. A variety of docking and scoring methods used in virtual screening have been the subject of active research in an effort to gauge limitations and articulate best practices. However, how to best utilize different scoring functions and various crystal structures, when available, is not yet well understood. In this work we use multiple crystal structures of PI3 K-γ in both prospective and retrospective virtual screening experiments. Both Glide SP scoring and Prime MM-GBSA rescoring are utilized in the prospective and retrospective virtual screens, and consensus scoring is investigated in the retrospective virtual screening experiments. The results show that each of the different crystal structures that was used, samples a different chemical space, i.e. different chemotypes are prioritized by each structure. In addition, the different (re)scoring functions prioritize different chemotypes as well. Somewhat surprisingly, the Prime MM-GBSA scoring function generally gives lower enrichments than Glide SP. Finally we investigate the impact of different ligand preparation protocols on virtual screening enrichment factors. In summary, different crystal structures and different scoring functions are complementary to each other and allow for a wider variety of chemotypes to be considered for experimental follow-up.

Keywords Virtual screening · Docking · Kinases · Protein flexibility · Enrichment · Chemical diversity

Electronic supplementary material The online version of this article (doi:10.1007/s10822-010-9356-2) contains supplementary material, which is available to authorized users.

N. Brooijmans
Structural Biology and Computational Chemistry, Wyeth Research, 401 N. Middletown Road, Pearl River, NY, USA

N. Brooijmans (✉)
Lead Finding Platform, Novartis Institutes for BioMedical Research, 250 Massachusetts Avenue, Cambridge, MA, USA
e-mail: reprints@prodigy.net

C. Humblet
Structural Biology and Computational Chemistry, Wyeth Research, CN 8000, Princeton, NJ, USA

Introduction

With the increase in computational power and availability of structural information, computer-aided drug discovery has become a standard tool in the search for novel leads, which can potentially be optimized to become drugs. Docking algorithms, which predict how a ligand interacts with a (protein) target, are used both during lead discovery and lead optimization. In lead discovery hundreds of thousands, if not millions, of compounds can be screened quickly and prioritized for experimental screening based on an assessment of their complementarity to the target binding site. This process is often called (high-throughput) virtual screening (HTVS or VS), analogously to experimental high-throughput screening (HTS), which has become the standard starting point for many drug discovery projects in the (bio)pharmaceutical industry or academic environments.

Docking algorithms consist of a search algorithm and a scoring function. The search algorithm searches through the rotational and translational degrees of freedom of the ligand in the protein binding site. Dihedral angles of the ligand are searched also, while the receptor is generally considered to be rigid, especially in VS settings. The scoring function assesses the complementarity of the

different ligand poses that are generated and rank orders them. Many different search algorithms and scoring functions have been explored [1].

Validation studies of docking algorithms mostly focus on binding mode predictions or on virtual screening enrichment. In binding mode prediction studies, known protein–ligand complex structures are used to assess the ability of the algorithm to reproduce the experimentally observed binding mode of the ligand. The accuracy of the docked binding poses is assessed by calculating the Root Mean Square Deviation (RMSD) of the pose with respect to the crystallographically-observed binding mode. Usually a cut-off of 2 Ångstrom is used, as it is expected that at RMSD's larger than 2 Å, most of the native contacts observed in the crystal structure between the protein and the ligand are not reproduced. While most docking programs are able to reproduce known binding modes of 70–80% of the test cases in datasets of 100–200 protein–ligand complexes, there are differences in how they perform from one protein target to another [2–9].

Virtual screening validation studies have focused on retrospective experiments in which a large library of "random" compounds is spiked with known ligands for the target of interest. These studies have shown significant enrichment of the known compounds early on in the rank-ordered lists with enrichments significantly exceeding what would be achieved by random selection [2, 10–13]. The successes of prospective virtual screening experiments have also been described in the literature [14–17] and novel kinase inhibitors have been identified through virtual screening as well [18–38]. Docking tools have been shown to perform differently on various target families and relevant to this paper, the Glide algorithm was recently shown to perform well on kinases [9].

While great strides have been made in the development of docking algorithms and their ability to predict binding modes and yield significant enrichment in VS settings, the scoring functions used in docking, especially during HTVS, are only rough estimates of the binding free energy, as evidenced by a lack of correlation between the experimental binding free energy (as measured by the association/dissociation constant K or the IC_{50}) and docking scores [2, 39]. Much research has been spent in the development of better scoring [40, 41] and especially rescoring methods [40, 42–45]. Multi-tiered docking and scoring workflows are especially attractive as a much more limited number of ligands and poses can be considered after an initial screen, and thus more rigorous and time-consuming scoring functions can be applied. The MM-PBSA [46, 47] or MM-GBSA scoring functions [48] are of particular interest in this work, which combine gas phase molecular mechanics interaction terms with polar desolvation energies from the Poisson-Boltzmann (PB) [49, 50] or Generalized Born (GB) [48] equation, and a solvent-accessible surface area term [51, 52] to account for nonpolar desolvation effects upon complex formation. Several recent papers have highlighted the utility of MM-PBSA or MM-GBSA in both pose prediction [42, 43], virtual screening enrichment [42], and in correlation with experimental binding free energies [53, 54].

Another area of interest to potentially improve virtual screening enrichments has been the use of consensus scoring. It is assumed that the combination of multiple scoring functions will result in higher enrichments due to the elimination of false positives [55, 56]. While several authors have described enhanced enrichments in virtual screening when using consensus scoring [55, 57], others reported that consensus scoring never performed better than the best individual scoring function [58, 59].

Often times, crystal structures of the compounds of interest for lead optimization do not become available until later points. The binding mode predictions are thus performed with either the apo structure or an unrelated ligand-bound structure. Docking of a ligand against a non-native structure is often called "cross-docking", and cross-docking experiments only have success rates of $\sim$50% or less [39, 60–65]. These experiments show that small changes in the protein structure are often required to be able to reproduce known binding poses and highlight the deficiencies in the rigid receptor assumption utilized by most docking algorithms. In the HTVS setting it has also been shown that different crystal structures of the same target yield different enrichments [10, 66].

Various approaches have been implemented to take induced fit effects into account. The simplest method is a "soft docking" approach, in which van der Waals overlap between the ligand and the protein is allowed [67]. Different methods that utilize the information from multiple crystal structures have also been developed [66, 68–70]. More explicit methods sample rotameric states of amino acid side chains [71–73]. The various methods that have been developed have reported greater success in reproducing known binding modes when receptor flexibility is taken into account. Retrospective virtual screening studies utilizing multiple crystal structures have appeared in the literature as well. Shoichet's lab published two papers utilizing conformational ensembles of Aldose reductase [67] and thymidylate synthase [74] showing increased enrichments. Similarly to the consensus scoring results using multiple scoring functions, consensus scoring with multiple crystal structures has led to both increases in VS enrichments [75] and more mixed results in terms of enrichments [76].

In this work we utilized the availability of multiple crystal structures of phosphatidyl inositol-3 kinase gamma (PI3 K-γ) in complex with various chemical series to take

induced fit effects into account explicitly. PI3 K-γ is a class IB lipid kinase that phosphorylates PIP-2 to form PIP-3 and is a drug target for diseases with abnormal leukocyte expression and recruitment [77]. Currently there are fifteen PI3 K-γ crystal structures in the public domain and at Wyeth a number of crystal structures in complex with various chemotypes are available. Five crystal structures were chosen for virtual screening by a visual comparison of available crystal structures. Resolution for the chosen structures varies from 2.1 to 2.9 Å. The moderate resolution is mainly due to poor diffraction of several domains other than the kinase domain. Compound densities for the chosen structures are always good. Both prospective and retrospective virtual screening experiments have been performed with the five chosen structures using the GLIDE docking algorithm [7]. Each of the receptors is treated as a separate experiment and the hit lists from each receptor are processed individually. In the prospective VS studies, compounds are visually selected from each of the receptors for experimental follow-up. In the prospective studies chemical overlap of top scoring compounds against each crystal structure is analyzed as well. In the retrospective VS studies utilizing the same crystal structures, rather than using optimized PI3 K-γ inhibitors, an in-house run HTS screen against PI3 K-γ is used to assess VS enrichments. In both the prospective and retrospective studies the chemical space covered by the Glide SP scoring function, the Prime MM-GBSA rescoring function and the consensus score (retrospective study only) will be analyzed.

Methods

Protein structure selection & preparation

Ten public domain and 8 in-house crystal structures were visually analyzed and compared. Five structures were selected that showed the greatest structural variation; the structures were co-crystallized with five unrelated ligands, including Staurosporine (1E8Z.pdb) ("Structure 5") and a pyrazolopyrimidine inhibitor (3IBE.pdb) ("Structure 1"). Each complex structure was minimized using the Embrace minimization protocol in Macromodel 9.5 for 500 steps with the OPLS2005 force field and the GBSA solvation model for water [78]. The protein preparation wizard in the Maestro GUI was used to generate the docking grid. For each structure the co-crystalized ligand was used to define the docking box using default settings [79].

Ligand preparation

For the prospective virtual screening experiments, compounds were selected from both the corporate (CORP), in-house collection, and from ChemNavigator (CNAV) [80]. CNAV contains compounds from a large number of commercial vendors and includes both "real" (sourceable) and "virtual" compounds. To compile the CNAV virtual library for docking we started from the sourceable compounds. For CORP an availability filter was also applied. Subsequently, both CORP and CNAV were filtered for lead-like properties using OEChem-based python scripts [81] for molecular weight ($180 < MW < 400$); rotatable bonds (<7); number of chiral centers (<2); number of rings (>0); ionizable groups (<3); and reactive functionalities (none allowed). ClogP is calculated as well and only compounds with ClogP < 4 are kept. After filtering, which includes an expansion of chiral centers, a single 3D conformation was generated with Omega2 [82]. The Omega2 output was utilized directly by Glide [83].

For the retrospective virtual screening experiments, all the compounds that were tested in the PI3 K-γ HTS campaign were filtered for lead-like properties as for CORP and CNAV. The HTS screen tested $\sim$533,000 compounds, of which $\sim$179 K compounds were lead-like. The OEChem-based expansion of chiral centers resulted in $\sim$213 K compounds. After expansion of chiral centers, LigPrep (LP) was used to generate tautomers, stereoisomers, ionization states, and ring conformations according to several different protocols. Both Ionizer (Ion) and Epik were used to generate tautomers and ionization states. The "LP-Ion" and "LP-Epik" protocols expanded the ligand states at pH = 7.0 and used Ionizer and Epik, respectively, to generate the tautomeric and ionization states. The LP-Ion library contains $\sim$223 K compounds, while LP-Epik has $\sim$215 K compounds. A more expansive annotation was done using the "GoNuts" (GN) protocols by setting the pH range to 7 ± 2 in the "LP-GN-Ion" and "LP-GN-Epik" libraries. The LP-GN-Ion library consists of $\sim$289 K compounds, while the LP-GN-Epik library contains $\sim$292 K compounds. The various LigPrep libraries were used by Glide for virtual screening.

Definition of HTS hits

The PI3 K-γ primary HTS screen was run at a compound concentration of 10 μM utilizing a TR-FRET assay. Hits were defined based on statistics and a %inhibition of $\geq$20.32% was found to be significant. Those compounds with %inhibition of $\geq$20.32% were screened in triplicate in the confirmation screen and a counter screen, again at 10 μM. Compounds with an average % inhibition of $\geq$20.32% in the secondary screen and %inhibition $<$20.32% in the counter screen were defined as the "confirmed hits" for further follow-up by the PI3 K-γ biology team. The "confirmed hits" as defined above have been classified as actives in the retrospective VS analysis.

All other tested compounds were considered inactive and were used as decoys in the VS runs. The HTS collection was filtered for lead-like properties and prepared for docking as described in the previous section. The number of lead-like HTS hits was 558, and the hit rate was 0.3% (558/∼179 K).

Docking and scoring

For the prospective studies, the CORP and CNAV libraries were docked against each of the five chosen structures with the Virtual Screening Workflow (VSW) in Maestro [79]. In the first step Glide was run in high-throughput HTVS mode. 10% of the top-scoring ligands were kept to go onto the next, Glide Single Precision (SP), stage. 10% of the top-scoring ligands were kept. The top-scoring list from Glide SP was processed in two ways to select compounds for experimental testing. First, the top 150 compounds were visualized and those with binding poses that fulfilled a majority of expected molecular interactions were chosen for testing. As the target is a kinase, presence of one or more of the canonical hinge-region hydrogen bonds was a strong determining factor in prioritizing compounds for testing. Second, the top 1,000 compounds from SP were minimized and re-scored using the Prime MM-GBSA function [84]. Again, the top 150 compounds were visually assessed for complementarity to the binding site and prioritized for screening.

For the retrospective studies, the Glide HTVS-SP Virtual Screening Workflow was again used to dock the various HTS libraries against each of the receptors. The top 10,000 compounds were saved and analyzed for retrieval of the confirmed HTS hits. In addition, the top 10,000 compounds were minimized and rescored and re-ranked utilizing the Prime MM-GBSA function. The receptor was kept rigid during the Prime MM-GBSA minimizations. For Prime MM-GBSA rescoring, two methods were used, namely with and without taking the internal energy of the ligand into account.

The Glide SP and Prime MM-GBSA ranked hit lists were used to calculate a consensus score using the sum rank method [85].

VS analysis

For the prospective VS studies, the hit rate (HR) is utilized to analyze the results as the activities for the compounds not selected for testing against PI3 K-γ are unknown. The HR is defined as the number of hits/number of compounds tested. Hits are defined as having an $IC_{50} < 20$ μM.

For the retrospective studies, ROC curves are utilized to show performance compared to randomly selecting compounds for testing, and enrichment factors (EF) are calculated by using the true positive rate (TPR; the y-axis of the ROC plot) and false positive rate (FPR; the x-axis of the ROC plot). The EF is thus defined as $EF = \frac{TPR}{FPR}$, as suggested by Jain and Nicholls [86] and EF's were calculated at various false positive rates.

Diversity analysis

Similarity of the hits retrieved by each receptor in the prospective VS studies to the co-crystallized ligand was assessed with Scitegic Pipeline Pilot using the FCFP_4 fingerprints and a Tanimoto cut-off of 0.5 to define similarity.

To assess the diversity of hit lists, an in-house minimal linkage clustering algorithm was utilized with the Unity fingerprints. The number of clusters and singletons are reported to illustrate the diversity of the tested compounds in the prospective VS studies.

In the retrospective VS studies, Tripos's DataMiner tool [87] was used to analyze the chemical space explored by hits in the Top 500 scoring compounds by either Glide SP docking, Prime MM-GBSA rescoring and consensus scoring.

ECFP_4 similarity analyses were performed using Accelrys Pipeline Pilot V7.5. For each Top 500 hit list of each receptor and each scoring function the average similarity to the co-crystallized ligand was calculated based on the ECFP_4 Tanimoto similarity. The similarity between the top-scoring compounds of the different scoring functions per X-ray structure was assessed using the ECFP_4 fingerprints as well. Diversity within each hit list based on ECFP_4 fingerprints was calculated also.

Results and discussion

Structural diversity of used crystal structures

The observed structural variation in the ATP binding site is due to the co-crystallization of PI3 K-γ with different ligands. As shown in Fig. 1, the five ligands have significantly different shapes and chemical structures in comparison to Staurosporine (Structure 5) and with respect to each other. Chemical similarity of the co-crystallized ligands to each other (all-by-all comparison) was maximally 0.21 based on Scitegic's FCFP_4 fingerprints, which again underscores the diversity of the co-crystallized ligands, resulting in different induced fit effects in the crystal structures. The different co-crystallized ligands do not cause any significant changes in the overall structure of PI3 K-γ (Fig. 2), but for a number of active site residues, significant conformational changes are observed (Fig. 3). Changes in the catalytic lysine (residue 833) and the

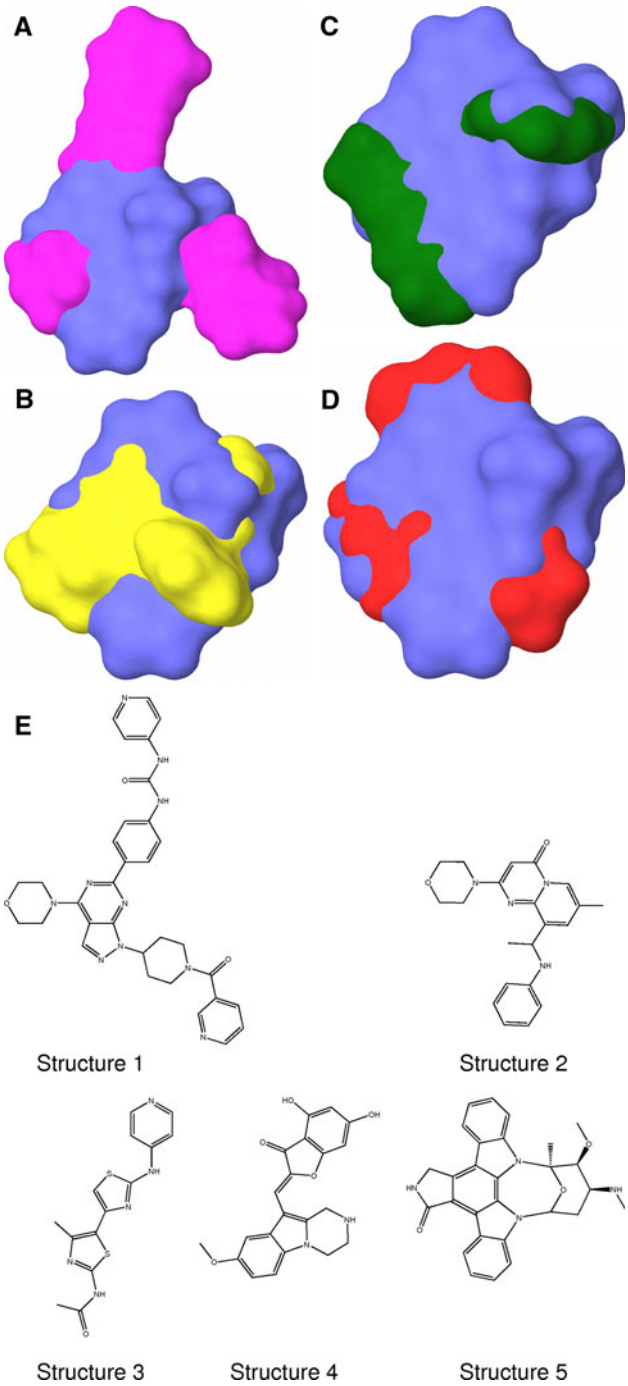

Fig. 1 Shape and chemical representations of co-crystallized ligands. Shape comparisons of co-crystallized ligands. Staurosporine is shown with *blue* surface color, overlays with **a** S1, **b** S2 (*top right*), **c** S3, and **d** S4. **e** Chemical structures of co-crystallized ligands

aspartic acid of the DFG motif (residue 964) are particularly striking as they not only influence how ligands should present hydrogen bonding groups to the enzyme to be able to interact with these residues, but different positioning of these residues also significantly impacts the overall shape of the binding site.

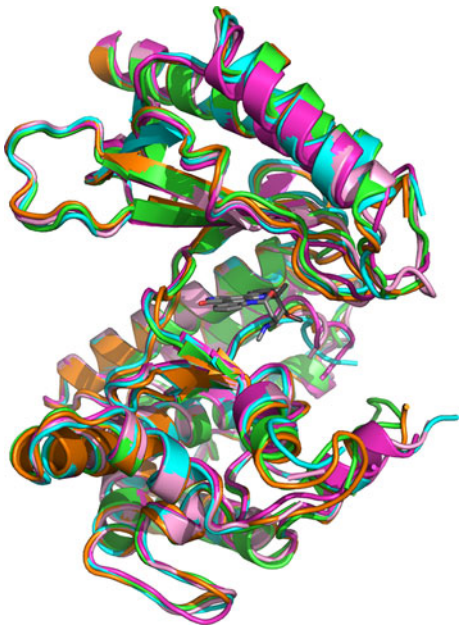

Fig. 2 Ribbon diagram of overlay of five PI3 K-γ X-ray structures used in the study. Overlay of Structure 1 (*cyan*), Structure 2 (*magenta*), Structure 3 (*orange*), Structure 4 (*pink*), and Structure 5 (*green*). Staurosporine is shown in the binding site (*gray carbons*). Structures are oriented such that the N-lobe is at the *top* and the C-lobe at the *bottom*. The glycine-rich loop is the short strand shown on the *left-hand* side of the picture

Prospective VS

As shown in Fig. 4, the VS strategy utilized five distinct PI3 K-γ crystal structures, each co-crystallized with a different ligand. Both the corporate collection and a library of commercially available compounds were screened against each structure. Selection of compounds for follow-up was done after Glide Single Precision (SP) scoring and after Prime MM-GBSA rescoring of the SP hit lists with the top-scoring 1,000 compounds. In this paradigm, a total of 10 hit lists were analyzed visually for each compound collection to select compounds for further follow up, coming from each of the five receptors from two different scoring functions. As part of the VS screening paradigm (and HTS follow-up) common kinase scaffold are excluded. The $\sim$40 scaffolds compiled internally are based on published kinase inhibitors and in-house scaffolds being pursued in other kinase projects. The excluded kinase scaffolds included the co-crystallized ligands for each of the receptors.

For each selected compound, the crystal structure(s) it was prioritized by was analyzed. Figure 5 highlights that the structural differences observed in the crystal structures results in virtually no overlap in selected top-scoring hits from each of the receptors: most of the compounds were selected based on a high score, and thus good fit, to only a single receptor. The figure also shows that most

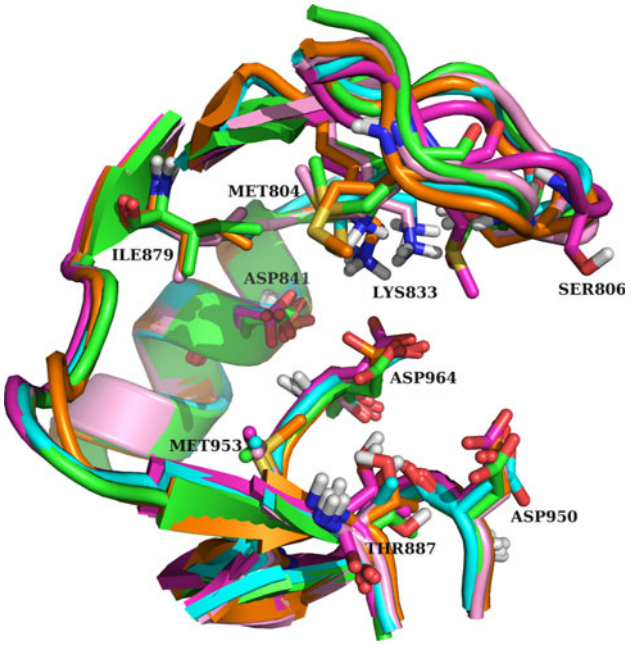

Fig. 3 Close-up of ATP-binding site of the five PI3 K-γ crystal structures used in the study highlighting residue movements. Overlay of Structure 1 (*cyan carbons*), Structure 2 (*magenta carbons*), Structure 3 (*orange carbons*), Structure 4 (*pink carbons*) and Structure 5 (*green carbons*). Residues shown are the gatekeeper residue (Ile 879), Thr887 in the specificity surface, Lys883 (catalytic lysine), glycine-rich loop residues Met804 and Ser 806, Helix C residue Asp841, DFG-motif residue Asp964 and residues Met953 and Asp950

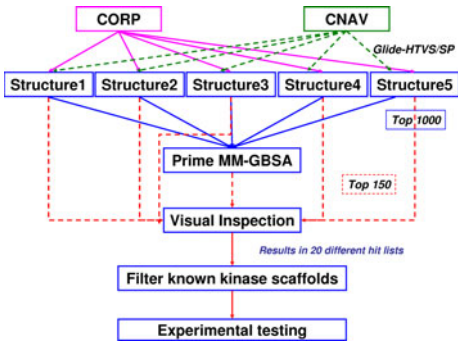

Fig. 4 Schematic of prospective virtual screening strategy. The figure illustrates that the corporate (CORP) and commercially-available collection (CNAV, through ChemNavigator) were each screened against five individual PI3 K-γ crystal structures using Glide HTVS/SP docking. Subsequently, the top 150 scoring compounds from each receptor were inspected visually to select compounds for testing. In addition, the best 1,000 scoring compounds from Glide SP were rescored using Prime MM-GBSA. Finally, all 20 hit lists were filtered using substructure searches to remove known kinase scaffolds, including analogs of the co-crystallized ligands

compounds are being prioritized by only a single scoring function, either SP or the Prime MM-GBSA scoring, although some overlap is observed.

Prospective hit rates per scoring function are shown in Tables 1 and 2 for CORP and CNAV, respectively. On average, more compounds were selected for testing from CORP than CNAV. This was deliberate, as the cost of purchasing compounds is significant and thus the chemistry team leader performed a more rigorous visual assessment focusing on chemically attractive and novel templates. Interestingly, however, the average hit rate for the CNAV compounds for either scoring function is approximately twofold higher on average. This is encouraging as the number of compounds, and thus the number of decoys, in CNAV is much greater, but the Glide HTVS-SP scoring paradigm is able to push the real hits to the top of the list. For both CNAV and CORP the MM-GBSA scoring function appears to give more consistent hit rates (Tables 1B and 2B). This is especially clear for the CNAV results where the average hit rate over the five receptors for MM-GBSA is 15%, while Glide SP yields a hit rate of only 5%. The overall hit rate for compounds selected by both

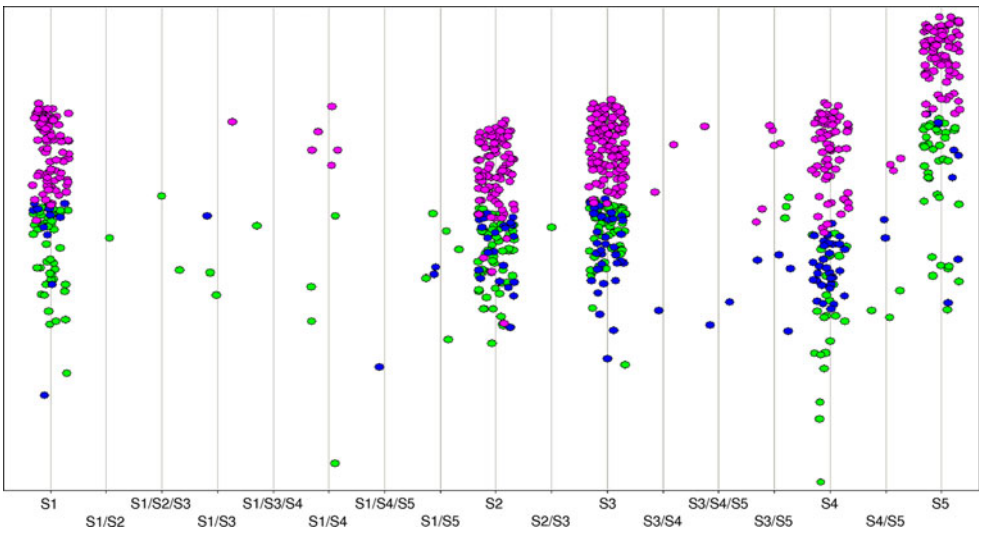

Fig. 5 Analysis of which receptor and scoring function and which crystal structure(s) prioritized compounds for experimental follow-up. Compounds prioritized by Glide SP (*green*), MMGBSA (*magenta*) or by both Glide SP and MMGBSA (*blue*). The *X*-axis shows for each compound which receptor structure(s) it was prioritized by. The *Y*-axis uses the Glide SP score for each compound as a way to separate the compounds, but has no other function than that

Table 1 Prospective hit rates for each structure based on HTVS using the lead like corporate compound collection

Structure	Tested	Selected for IC_{50}	$IC_{50} \leq 20$ μM	Hit Rate
A. Compounds picked from Glide SP docking hit list				
Structure 1	30	2	2	6.7
Structure 2	46	1	0	0.0
Structure 3	27	1	1	3.7
Structure 4	8	0	0	0.0
Structure 5	15	1	0	0.0
Total	126	5	3	2.4
Average				2.1
B. Compounds picked from MMGBSA rescoring after Glide SP docking				
Structure 1	37	2	2	5.4
Structure 2	91	6	4	4.4
Structure 3	48	3	2	4.2
Structure 4	22	0	0	0.0
Structure 5	35	3	3	8.6
Total	233	14	11	4.7
Average				4.5
C. Compounds picked by both Glide SP docking and MMGBSA rescoring				
Structure 1	6	0	0	0.0
Structure 2	21	0	0	0.0
Structure 3	14	1	1	7.1
Structure 4	2	2	1	50.0
Structure 5	6	0	0	0.0
Total	49	3	2	4.1
Average				11.4
D. Overall hit rates by combining all tested compounds				
Structure 1	73	4	4	5.5
Structure 2	158	7	4	2.5
Structure 3	89	5	4	2.5
Structure 4	32	2	1	3.1
Structure 5	56	4	3	5.4
Total	408	22	16	3.9
Average				3.8

scoring functions was even higher at 27% (Tables 1C and 2C), although the average hit rate was not much higher than for MM-GBSA. This is caused by several receptors not yielding any hits for SP-MM-GBSA picked compounds. In general, the number of compounds picked based on a good score by both scoring functions was very low, indicating that Glide SP and MM-GBSA prioritize different chemotypes to the very top despite the fact that only a limited number of compounds (1,000) was rescored by Prime-MM-GBSA. Overall, the VS hit rates are an order of magnitude greater than the HTS lead-like hit rate of 0.3%.

A number of tested compounds scored well against two or more receptors, as shown in Table 3. The CNAV screen did not yield many compounds that scored well against multiple receptors and none were active. As with the Glide SP and MMGBSA results, the different structures appear to favor different types of compounds, presumably as a result of the different residue positions in the binding site.

Not only did the different crystal structures prioritize different chemotypes (Fig. 5), but hits selected by each individual receptor are highly diverse as well based on a clustering analysis. For both the compounds picked from CORP (Table 4A) and from CNAV (Table 4B) approximately half of the tested compounds for each structure are singletons, while the rest of the chosen compounds for each receptor come from a significant number of clusters.

Retrospective VS

In the retrospective VS studies a number of experimental variables were investigated to assess their impact on VS enrichments. Besides separate VS screens against the five

Table 2 Prospective hit rates for each structure based on HTVS using the lead like ChemNavigator compound collection

Structure	Tested	Selected for IC_{50}	$IC_{50} \leq 20$ μM	Hit Rate
A. Compounds picked from Glide SP docking hit list				
Structure 1	13	1	1	7.7
Structure 2	39	1	0	0.0
Structure 3	4	0	0	0.0
Structure 4	16	4	2	12.5
Structure 5	17	2	1	5.9
Total	89	8	4	4.5
Average				5.2
B. Compounds picked from MMGBSA rescoring after Glide SP docking				
Structure 1	25	2	2	8.0
Structure 2	32	8	4	12.5
Structure 3	12	5	4	33.3
Structure 4	5	1	1	20.0
Structure 5	15	2	2	13.3
Total	89	18	13	14.6
Average				17.4
C. Compounds picked by both Glide SP docking and MMGBSA rescoring				
Structure 1	3	0	0	0.0
Structure 2	7	4	3	42.9
Structure 3	3	1	1	33.3
Structure 4	1	1	0	0.0
Structure 5	1	0	0	0.0
Total	15	6	4	26.7
Average				15.2
D. Overall hit rates by combining all tested compounds				
Structure 1	41	3	3	7.3
Structure 2	78	13	7	9.0
Structure 3	19	6	5	26.3
Structure 4	22	6	3	13.6
Structure 5	33	4	3	9.1
Total	193	32	16	8.3
Average				3.8

Table 3 Prospective hit rates for compounds selected by multiple crystal structures

Structures	Library	Tested	Selected for IC_{50}	$IC_{50} \leq 20$ μM	Hit rate
S1/S4	CORP	4	2	1	25.0
S3/S5	CORP	5	1	1	20.0
S1/S4	CNAV	2	0	0	0
S1/S3	CNAV	1	0	0	0
S1/S2	CNAV	1	0	0	0

individual structures utilized in the prospective screens and again the use of Glide SP docking and Prime MM-GBSA rescoring, consensus scoring was used as well. In addition, various ligand preparation protocols were explored. Two different methods to determine ionization and tautomer states, Ionizer and Epik, which are available in LigPrep were assessed. States were generated at two different pH (ranges), namely pH = 7 and pH range 5–9 ("GoNuts").

Tables 5, 6, and 7 summarize the Glide SP, Prime MM-GBSA (with ligand internal energy), and Prime MM-GBSA (without ligand internal energy) respectively, for each of the four different ligand preparation protocols, namely LP-Ion (5A, 6A, 7A), LP-Ion-GoNuts (5B, 6B, 7B), LP-Epik (5C, 6C, 7C), and LP-Epik-GoNuts (5D, 6D, 7D). As in the prospective screens, each receptor and each ligand preparation protocol yields enrichments better than random with EF's ranging from 4 to 40 for Glide SP. Interestingly, for both ligand preparation protocols that utilized ionizer to determine protonation/tautomer states, EF's at a false positive rate of 0.5% are higher than at 0.05% for Glide SP (Tables 5A, 6B). For the Epik prepared libraries, the opposite is true (Table 5C, D). The annotation of a large number of states over a pH range of 5–9

Table 4 Diversity analysis of selected compounds from prospective virtual screening by Glide SP and Prime MMGBSA

Structure	Tested	Singletons	Clusters
A. Glide SP selected compounds			
Structure 1	73	41	10
Structure 2	158	77	20
Structure 3	89	37	5
Structure 4	32	26	2
Structure 5	56	31	9
B. Prime MMGBSA selected compounds			
Structure 1	41	19	7
Structure 2	78	30	12
Structure 3	19	7	5
Structure 4	22	8	8
Structure 5	33	9	9

("GoNuts") can either hurt or increase enrichment in either Ionizer or Epik-prepared libraries, depending on the structure that was used.

For Glide SP (Table 5), the Epik-prepared compound libraries yield significantly better enrichments than the Ionizer prepared ones, increasing the EF at the 0.05% false positive rate by twofold for the "GoNuts" protocol (5B vs. 5D), indicating the Epik methodology is probably a best practice for preparing compound databases, although this should be confirmed with additional validation experiments. Increasing the number of tautomer and ionization states increased enrichment in two structures when Epik is used, but decreases the EFs for the other three structures. It is thus not clear what the best practice for the annotation of tautomer and ionization states is, but a more conservative approach (i.e. over a smaller pH range) might be recommended.

A comparison of the two different sets of Prime MM-GBSA results, with (Table 6) and without internal ligand strain energy (Table 7), shows that including this term consistently results in similar or increased enrichments, making the inclusion of the ligand strain energy a best practice for Prime MM-GBSA scoring. Expansion of tautomers and ionization states at a pH range of 5–9

Table 5 Retrospective ROC enrichment factors for Glide SP docking for each structure at various false positive rates

Structure	EF @ FP = 0.05	EF @ FP = 0.5	EF @ FP = 1.0	EF @ FP = 2.0	EF @ FP = 5.0
A. Rank ordering based on LigPrep (LP) ligand preparation protocol					
Structure 1	17.8	20.8	15.4	11.2	6.9
Structure 2	10.7	12.2	8.6	5.8	3.6
Structure 3	28.5	21.5	16.1	10.3	6.1
Structure 4	3.6	16.8	12.2	9.9	7.1
Structure 5	14.3	20.1	15.4	12.3	7.7
Average	15.0	18.3	13.5	9.9	6.3
B. Rank ordering based on LigPrep Go Nuts (LP-GoNuts) ligand preparation protocol					
Structure 1	7.1	16.1	14	10.6	6.7
Structure 2	14.3	14.3	11.5	7.1	4.4
Structure 3	17.8	21.5	17.4	10.7	6.2
Structure 4	10.7	15	11.8	9.3	7
Structure 5	7.1	17.2	14.7	11.5	7.7
Average	11.4	16.8	13.9	9.8	6.4
C. Rank ordering based on LigPrep Epik (LP-Epik) ligand preparation protocol					
Structure 1	35.7	19	14	11.1	6.9
Structure 2	7.1	12.5	8.2	5.6	3.7
Structure 3	39.2	27.2	18.8	11.5	6.7
Structure 4	17.8	19	14.5	10.7	7.3
Structure 5	17.8	19.3	15.8	12.5	7.4
Average	23.5	19.4	14.3	10.3	6.4
D. Rank ordering based on LigPrep Go Nuts Epik (LP-GoNuts-Epik) ligand preparation protocol					
Structure 1	39.2	22.2	15.9	10.8	6.8
Structure 2	10.7	13.3	9.7	6.1	4
Structure 3	32.1	24.4	18.8	11.6	6.7
Structure 4	28.5	18.3	12.9	9.9	7
Structure 5	10.7	20.4	16.5	12.6	8
Average	24.2	19.7	14.8	10.2	6.5

Table 6 Retrospective ROC enrichment factors for Prime MMGBSA rescoring results (after Glide SP docking) including the internal energy ligand strain term for each structure at various false positive rates

Structure	EF @ FP = 0.05	EF @ FP = 0.5	EF @ FP = 1.0	EF @ FP = 2.0	EF @ FP = 5.0
A. Rank ordering based on LigPrep (LP) ligand preparation protocol					
Structure 1	14.3	14.7	12.7	9.6	7.1
Structure 2	3.6	4.7	3.8	3.5	0
Structure 3	28.5	19	13.6	10.5	6.4
Structure 4	17.8	10.4	9.9	7.8	7.1
Structure 5	28.5	12.5	12.7	11.6	8
Average	18.5	12.3	10.5	8.6	5.7
B. Rank ordering based on LigPrep Go Nuts (LP-GoNuts) ligand preparation protocol					
Structure 1	7.1	10.4	11.5	9.9	7.5
Structure 2	0	2.9	4.7	4.6	5.2
Structure 3	25	17.2	14.3	11.5	8.2
Structure 4	25	13.3	11.3	8.5	7.9
Structure 5	28.5	11.5	12.5	11	8.6
Average	17.1	11.1	10.9	9.1	7.5
C. Rank ordering based on LigPrep Epik (LP-Epik) ligand preparation protocol					
Structure 1	17.8	11.8	9.5	8.3	6.9
Structure 2	0	4.3	4.5	3.7	3.5
Structure 3	21.4	16.1	13.4	10.5	6.6
Structure 4	17.8	19	14	10.6	7.2
Structure 5	28.5	13.3	13.1	12.7	7.4
Average	17.1	12.9	10.9	9.2	6.3
D. Rank ordering based on LigPrep Go Nuts Epik (LP-GoNuts-Epik) ligand preparation protocol					
Structure 1	10.7	12.5	11.3	11	9.4
Structure 2	7.1	14	10.2	6.8	
Structure 3	15	18.6	14.3	12.9	8.9
Structure 4	21.4	12.2	12.7	10.5	8.9
Structure 5	25	15.8	17.4	15.8	10.5
Average	15.8	14.6	13.2	11.4	9.4

("GoNuts" results) again mostly hurts enrichment (except for Structure 4) suggesting that this expansion is probably not necessary for VS experiments. For the Prime MM-GBSA results that include the ligand strain, the use of Epik to prepare the compound library generally doesn't increase the enrichment much, but as this is a best practice for Glide SP docking, use of Epik would be recommended.

Somewhat unexpectedly, Prime MM-GBSA yields significantly lower enrichments compared to Glide SP for the Epik-prepared libaries, generally 10 units lower than the corresponding Glide SP enrichment factors. For the Ionizer-prepared libraries, the Prime MM-GBSA enrichment can be either higher or lower.

Analysis of the consensus-scoring results (Table 8 and Suppl. Table S1) shows that, on average, consensus scoring gives better enrichments than either the Glide SP or Prime MM-GBSA scoring functions individually. EF's range from 4 to 75 at the 0.05% false positive rate. For structures 3, 4, and 5, consensus scoring gives similar or better EF's consistently and for structures 1 and 2 generally one of the individual scoring functions, usually Glide SP, gives superior enrichment over consensus scoring.

Because the prospective VS experiments showed a marked difference in sampled chemical space for the Glide SP and Prime MM-GBSA scoring functions, Dataminer was used to analyze the chemical space covered by the top scoring lists (Top 500 from Glide SP, Prime MM-GBSA, and the consensus results). Two examples of the resulting chemical space plots are shown in Fig. 6 for Structure 1 (LigPrep-Ion protocol with ligand strain for Prime MM-GBSA) and Structure 3 (LigPrep-Epik protocol without ligand strain in Prime MM-GBSA) (complete set in Supplementary Figs. 1–5; detailed analysis in supplementary Table S2). The compounds that are ranked in the Top 500 of the consensus ranked results can be categorized in four different bins, depending on what scoring function scored the compound well. Category 1 and 2 of consensus

Table 7 Retrospective ROC enrichment factors for Prime MMGBSA rescoring results (after Glide SP docking) not including the internal ligand strain energy term for each structure at various false positive rates

Structure	EF @ FP = 0.05	EF @ FP = 0.5	EF @ FP = 1.0	EF @ FP = 2.0	EF @ FP = 5.0
A. Rank ordering based on LigPrep (LP) ligand preparation protocol					
Structure 1	3.6	10.3	11.1	8.3	7.0
Structure 2	0	3.2	3.6	3.5	3.4
Structure 3	21.4	14.7	12.4	10.2	6.3
Structure 4	17.8	8.6	8.1	7.0	7.0
Structure 5	28.5	10.0	10.4	10.3	8.0
Average	14.3	9.4	9.1	7.9	6.3
B. Rank ordering based on LigPrep Go Nuts (LP-GoNuts) ligand preparation protocol					
Structure 1	3.6	9.0	8.4	8.1	7.2
Structure 2	0	2.9	3.4	3.9	4.9
Structure 3	14.3	13.3	12.5	10.8	8.3
Structure 4	14.3	9.0	8.2	7.9	7.7
Structure 5	28.5	9.7	9.9	9.9	8.5
Average	12.1	8.8	8.5	8.1	7.3
C. Rank ordering based on LigPrep Epik (LP-Epik) ligand preparation protocol					
Structure 1	3.6	7.5	7.9	6.8	6.8
Structure 2	0	4.3	3.2	3.0	3.4
Structure 3	14.3	12.9	12.4	10.5	6.6
Structure 4	10.7	7.2	7.0	7.0	7.4
Structure 5	25.0	12.2	11.1	11.1	7.4
Average	10.7	8.8	8.3	7.7	6.3
D. Rank ordering based on LigPrep Go Nuts Epik (LP-GoNuts-Epik) ligand preparation protocol					
Structure 1	3.6	9.3	9.0	8.6	9.2
Structure 2	0	5.0	4.1	4.1	
Structure 3	17.8	14.7	14.7	12.8	8.9
Structure 4	10.7	10.0	10.8	9.7	8.9
Structure 5	25.0	14.3	15.6	14.6	10.4
Average	11.4	10.7	10.9	10.0	7.5

top-scoring compounds are compounds that scored well by either Glide SP or Prime MM-GBSA individually (Top 500), while the other scoring function ranked the same compound outside of the Top 500, but still low enough such that the combined ranks drove them in the consensus top 500 list. The third category consists of compounds that scored in the top 500 of both Glide SP and Prime MM-GBSA. The last category contains compounds that scored moderately well (outside of the top 500) with Glide SP and Prime MM-GBSA, but the combined score was low enough for these compounds to make it into the top 500 of the consensus list.

The plots are color coded for these 4 consensus categories and also for compounds that scored well with only Glide SP or Prime MM-GBSA. Visual assessment of the plots shows that very few compounds scored well with both Glide SP and Prime MM-GBSA, ending up in the top-scoring consensus list. Only between 50 and 100 compounds in the consensus list scored in the top 500 of both Glide SP and Prime MM-GBSA (Suppl. Table S2). Most of the compounds ($\sim$300 from top 500) score well with only Glide SP or Prime MM-GBSA and do not end up in the consensus list. For the various consensus categories, $\sim$40% of the compounds prioritized by the consensus score come from outside the top 500 of both Glide SP and Prime MM-GBSA, and would thus never have been looked at without the calculation of the consensus score. The consensus list further consist of an approximately equal number of compounds that scored in the top 500 of either Glide SP or Prime MM-GBSA, indicating that a number of compounds score very well with one of these scoring functions and moderately well (outside of top 500) with the other one.

To further confirm that each individual structure and each scoring function prioritizes different chemotypes, with each hit list being highly diverse as well, a number of fingerprint similarity analyses were carried out using the ECFP_4 fingerprints in PipelinePilot (http://accelrys.com/products/pipeline-pilot/). Supplementary Table S3 summarizes the findings and the experiments confirm that the three hit lists generated per scoring function are highly

Table 8 Retrospective ROC enrichment factors for consensus scoring combining the Glide-SP docking and Prime MMGBSA (ligand strain included in Prime MM-GBSA results) rescoring results

Structure	EF @ FP = 0.05	EF @ FP = 0.5	EF @ FP = 1.0	EF @ FP = 2.0	EF @ FP = 5.0
A. Rank ordering based on LigPrep (LP) ligand preparation protocol					
Structure 1	32.1	21.5	15.4	11.8	7
Structure 2	3.6	7.9	6.8	4.7	3.5
Structure 3	74.9	26.2	16.3	10.4	6.2
Structure 4	21.4	14.7	13.6	10.8	7.2
Structure 5	28.5	20.1	16.7	13.2	7.8
Average	32.1	18.1	13.8	10.2	6.3
B. Rank ordering based on LigPrep Go Nuts (LP-GoNuts) ligand preparation protocol					
Structure 1	3.6	19.3	14.5	11.4	6.8
Structure 2	3.6	9	8.6	6	4.4
Structure 3	42.8	26.5	15.6	10.5	6.3
Structure 4	17.8	15.8	12.4	10.2	7.1
Structure 5	32.1	21.1	16.1	13.4	7.8
Average	20.0	18.3	13.4	10.3	6.5
C. Rank ordering based on LigPrep Epik (LP-Epik) ligand preparation protocol					
Structure 1	25	15.8	13.3	10.9	7
Structure 2	3.5	7.5	7	5	3.7
Structure 3	74.9	29.4	17.9	11.5	6.6
Structure 4	17.8	19	14.5	10.7	7.3
Structure 5	28.5	23.3	19	13.7	7.4
Average	30.0	19.0	14.3	10.4	6.4
D. Rank ordering based on LigPrep Go Nuts Epik (LP-GoNuts-Epik) ligand preparation protocol					
Structure 1	35.7	19	14.7	11.2	7
Structure 2	10.7	13.3	9	5.7	
Structure 3	42.8	29.4	18.3	11.6	6.9
Structure 4	53.5	16.1	12.5	10.4	7.1
Structure 5	32.1	22.9	19	14.4	8.1
Average	25.0	20.1	14.7	10.7	7.3

diverse with respect to the co-crystallized ligand with the average similarity of top-scoring compounds to the respective ligand being 0.11. Moreover, each hit list is highly diverse with average similarity of each compound to the others in the hit list being 0.11. Finally, a comparison of the hit lists generated by each of the three scoring functions, shows that each scoring function samples different chemical space, as the similarity of compounds between the Top 500 of each scoring function is on average 0.13, despite the fact that Prime MM-GBSA and the Consensus score only sampled from very limited chemical space, namely that prioritized by the Glide SP scoring function.

Conclusions

A number of different variables that have to be considered when setting up large-scale virtual screening experiments have been investigated in this work. The use of different crystal structures of the same target, use of different scoring functions, including consensus scoring, and different ways of preparing and annotating compound libraries were analyzed. We also investigated the chemical space covered by the various scoring functions and crystal structures. Rather than spiking a random collection of compounds with known, optimized inhibitors against the target, in the retrospective VS studies an in-house run HTS campaign was used to investigate recovery of HTS actives.

The prospective VS studies, which utilized a visual analysis step to select compounds for testing, showed good hit rates for all structures used and highlighted that each individual structure that was used prioritizes different chemotypes. Each individual hit list showed good diversity also, with many singletons having been prioritized. In addition, the prospective VS studies highlighted that the scoring functions in Glide SP and Prime MM-GBSA are complementary to each other, even when only a limited

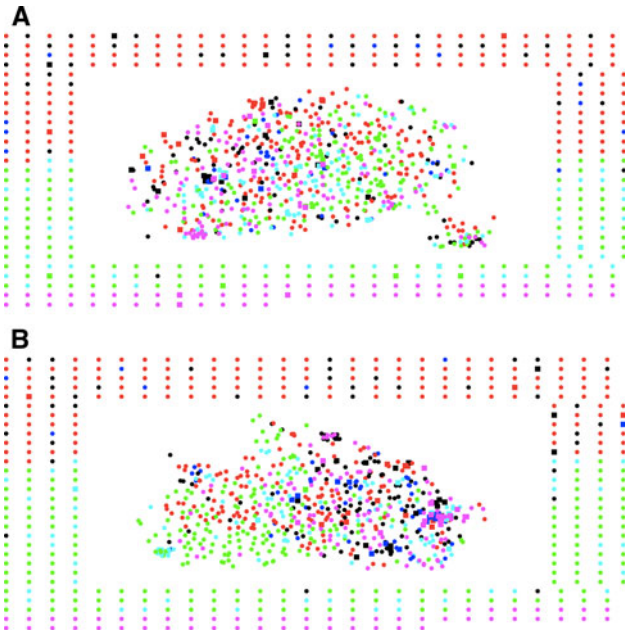

Fig. 6 Chemical space plots for compounds scoring in the top 500 of Glide SP, Prime MM-GBSA and consensus. **a** Plot for Structure 1, LP-Ion protocol including ligand strain term for Prime-MMGBSA. **b** Plot for Structure 3 (LP-Epik protocol, including ligand strain energy term. Compounds at the *edges* are chemically unrelated to other compounds in the dataset (singletons). Compounds in the *center* are chemically related to each other, and the distance between points is a measure of chemical similarity. Actives are indicated by *squares*, inactives by *circles*. Compounds that only scored in the Top 500 by Glide SP are shown in *red*; compounds that only scored in the Top 500 by Prime MM-GBSA are shown in *green*; compounds that scored in the Top 500 of both the Consensus score and Glide SP are shown in *black*; compounds that scored in the Top 500 of both the Consensus score and Prime MM-GBSA are shown in *cyan*; compounds that scored in the Top 500 of all three functions, Glide SP, Prime MM-GBSA, and Consensus are shown in *dark blue*; compounds that scored in the Top 500 of the Consensus score but not in the Top 500 of either Glide SP or Prime MM-GBSA are shown in *dark blue*

number of compounds are rescored, in that they prioritize different compounds for follow-up. Finally, Prime MM-GBSA gave more consistent hit rates than Glide SP.

The retrospective studies showed that the use of Epik to annotate compound tautomer and ionization states is recommended, while the pH range at which this expansion is performed should probably be more limited than the 5–9 range used in the "GoNuts" protocol. Somewhat unexpectedly, Prime MM-GBSA gave lower EF's on average than Glide SP, although a recent paper looking at enrichment in fragment-based VS showed a similar result [88]. For Prime MM-GBSA, the best EF's were achieved when the ligand internal strain energy was taken into account.

Analysis of chemical space covered by Glide SP, Prime MM-GBSA and the consensus score clearly showed that these scoring functions are complementary to each other and cover different chemical space. The consensus score was able to recover additional chemotypes that scored moderately well with either Glide SP or Prime MM-GBSA, or both, but were not present in the top 500 of either score individually. These three scoring functions thus aid in the recovery of a wider variety of chemotypes and potential inhibitors. Chemical space coverage by different scoring functions is not generally analyzed in VS papers, but a paper analyzing VS enrichments in PDE4B showed different recovery rates for different chemical series by consensus scoring [89]. Although we focused on PI3 K-γ in this work, we believe that in general, different scoring functions will recover different chemotypes and is thus a way to increase the number of chemical series in VS.

Chemical space covered by each of the individual crystal structures also showed little overlap in compounds prioritized by each receptor. Based on past cross-docking results this is not surprising, but to our knowledge no VS papers have utilized multiple crystal structures to increase the chemical diversity of hits. Analysis of consensus scoring across multiple receptors will be discussed in a future paper.

Acknowledgments NB thanks Derek Cole and Mike Bowman for support of the VS efforts, Jason Jussif for experimental testing of VS hits, Joel Bard and Kris Svenson for PI3 K-γ crystal structures, and Yongbo Hu for preparation of CORP and CNAV VS libraries.

References

1. Brooijmans N, Kuntz ID (2003) Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct 32(1): 335–373
2. 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(20):5912–5931
3. Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WT, Mortenson PN, Murray CW (2007) Diverse, high-quality test set for the validation of protein–ligand docking performance. J Med Chem 50(4):726–741
4. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein–ligand docking using GOLD. Proteins 52(4):609–623
5. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748
6. Kramer B, Rarey M, Lengauer T (1999) Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. Proteins 37(2):228–241
7. 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(7):1739–1749
8. Perola E, Walters WP, Charifson PS (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins 56(2):235–249

9. Cross JB, Thompson DC, Rai BK, Baber JC, Fan KY, Hu Y, Humblet C (2009) Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. J Chem Inf Model 49(6):1455–1474
10. Sheridan RP, McGaughey GB, Cornell WD (2008) Multiple protein structures and multiple ligands: effects on the apparent goodness of virtual screening results. J Comput Aided Mol Des 22(3–4):257–265
11. Kontoyianni M, Sokol GS, McClellan LM (2005) Evaluation of library ranking efficacy in virtual screening. J Comput Chem 26(1):11–22
12. Muegge I, Enyedy IJ (2004) Virtual screening for kinase targets. Curr Med Chem 11(6):693–707
13. Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP (2005) Comparison of automated docking programs as virtual screening tools. J Med Chem 48(4):962–976
14. Alvarez JC (2004) High-throughput docking as a source of novel drug leads. Curr Opin Chem Biol 8(4):365–370
15. Babaoglu K, Simeonov A, Irwin JJ, Nelson ME, Feng B, Thomas CJ, Cancian L, Costi MP, Maltby DA, Jadhav A, Inglese J, Austin CP, Shoichet BK (2008) Comprehensive mechanistic analysis of hits from high-throughput and docking screens against beta-lactamase. J Med Chem 51(8):2502–2511
16. Klebe G (2006) Virtual ligand screening: strategies, perspectives and limitations. Drug Discov Today 11(13–14):580–594
17. Kraemer O, Hazemann I, Podjarny AD, Klebe G (2004) Virtual screening for inhibitors of human aldose reductase. Proteins 55(4):814–823
18. Nagarajan S, Doddareddy M, Choo H, Cho YS, Oh KS, Lee BH, Pae AN (2009) IKKbeta inhibitors identification part I: homology model assisted structure based virtual screening. Bioorg Med Chem 17(7):2759–2766
19. Park H, Bahn YJ, Jeong DG, Woo EJ, Kwon JS, Ryu SE (2008) Identification of novel inhibitors of extracellular signal-regulated kinase 2 based on the structure-based virtual screening. Bioorg Med Chem Lett 18(20):5372–5376
20. Luo C, Xie P, Marmorstein R (2008) Identification of BRAF inhibitors through in silico screening. J Med Chem 51(19): 6121–6127
21. Fu DH, Jiang W, Zheng JT, Zhao GY, Li Y, Yi H, Li ZR, Jiang JD, Yang KQ, Wang Y, Si SY, Jadomycin B (2008) An Aurora-B kinase inhibitor discovered through virtual screening. Mol Cancer Ther 7(8):2386–2393
22. Kolb P, Huang D, Dey F, Caflisch A (2008) Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. J Med Chem 51(5):1179–1188
23. Cavasotto CN, Ortiz MA, Abagyan RA, Piedrafita FJ (2006) In silico identification of novel EGFR inhibitors with antiproliferative activity against cancer cells. Bioorg Med Chem Lett 16(7):1969–1974
24. Cozza G, Bonvini P, Zorzi E, Poletto G, Pagano MA, Sarno S, Donella-Deana A, Zagotto G, Rosolen A, Pinna LA, Meggio F, Moro S (2006) Identification of ellagic acid as potent inhibitor of protein kinase CK2: a successful example of a virtual screening application. J Med Chem 49(8):2363–2366
25. Cozza G, Gianoncelli A, Montopoli M, Caparrotta L, Venerando A, Meggio F, Pinna LA, Zagotto G, Moro S (2008) Identification of novel protein kinase CK1 delta (CK1delta) inhibitors through structure-based virtual screening. Bioorg Med Chem Lett 18(20):5672–5675
26. Foloppe N, Fisher LM, Howes R, Potter A, Robertson AGS, Surgenor AE (2006) Identification of chemically diverse Chk1 inhibitors by receptor-based virtual screening. Bioorg Med Chem 14(14):4792–4802
27. Hancock CN, Macias A, Lee EK, Yu SY, MacKerell AD Jr, Shapiro P (2005) Identification of novel extracellular signal-regulated kinase docking domain inhibitors. J Med Chem 48(14):4586–4595
28. Hu X, Prehna G, Stebbins CE (2007) Targeting plague virulence factors: a combined machine learning method and multiple conformational virtual screening for the discovery of yersinia protein kinase A inhibitors. J Med Chem 50(17):3980–3983
29. Li J, Tan J-z, Chen L-l, Zhang J, Shen X, Mei C-l, Fu L-l, Lin L-p, Ding J, Xiong B, Xiong X-s, Liu H, Luo X-m, Jiang H-l (2006) Design, synthesis and antitumor evaluation of a new series of N-substituted-thiourea derivatives. Acta Pharmacol Sin 27(9):1259–1271
30. Park H, Bahn YJ, Jeong DG, Woo EJ, Kwon JS, Ryu SE (2008) Identification of novel inhibitors of extracellular signal-regulated kinase 2 based on the structure-based virtual screening. Bioorganic & Medicinal Chemistry Letters 18(20):5372–5376
31. Peach ML, Tan N, Choyke SJ, Giubellino A, Athauda G, Burke TR Jr, Nicklaus MC, Bottaro DP (2009) Directed discovery of agents targeting the met tyrosine kinase domain by virtual screening. J Med Chem 52(4):943–951
32. Peng H, Huang N, Qi J, Xie P, Xu C, Wang J, Yang C (2003) Identification of novel inhibitors of BCR-ABL tyrosine kinase via virtual screening. Bioorg Med Chem Lett 13(21):3693–3699
33. Pierce AC, Jacobs M, Stuver-Moody C (2008) Docking study yields four novel inhibitors of the protooncogene pim-1 kinase. J Med Chem 51(6):1972–1975
34. Qin Z, Zhang J, Xu B, Chen L, Wu Y, Yang X, Shen X, Molin S, Danchin A, Jiang H, Qu D (2006) Structure-based discovery of inhibitors of the YycG histidine kinase: new chemical leads to combat *Staphylococcus epidermidis* infections. BMC Microbiol 6:96
35. Richardson CM, Nunns CL, Williamson DS, Parratt MJ, Dokurno P, Howes R, Borgognoni J, Drysdale MJ, Finch H, Hubbard RE, Jackson PS, Kierstan P, Lentzen G, Moore JD, Murray JB, Simmonite H, Surgenor AE, Torrance CJ (2007) Discovery of a potent CDK2 inhibitor with a novel binding mode, using virtual screening and initial, structure-guided lead scoping. Bioorg Med Chem Lett 17(14):3880–3885
36. Segura-Cabrera A, Rodriguez-Perez MA (2008) Structure-based prediction of *Mycobacterium tuberculosis* shikimate kinase inhibitors by high-throughput virtual screening. Bioorg Med Chem Lett 18(11):3152–3157
37. Toledo-Sherman L, Deretey E, Slon-Usakiewicz JJ, Ng W, Dai J-R, Foster JE, Redden PR, Uger MD, Liao LC, Pasternak A, Reid N (2005) Frontal affinity chromatography with MS detection of EphB2 tyrosine kinase receptor. 2. Identification of small-molecule inhibitors via coupling with virtual screening. J Med Chem 48(9):3221–3230
38. Warner SL, Bashyam S, Vankayalapati H, Bearss DJ, Han H, Von Hoff DD, Hurley LH (2006) Identification of a lead small-molecule inhibitor of the Aurora kinases using a structure-assisted, fragment-based approach. Mol Cancer Ther 5(7):1764–1773
39. Duca JS, Madison VS, Voigt JH (2008) Cross-docking of inhibitors into CDK. 2 structures 1. J Chem Inf Model 48(3):659–668
40. O'Boyle NM, Brewerton SC, Taylor R (2008) Using buriedness to improve discrimination between actives and inactives in docking. J Chem Inf Model 48(6):1269–1278
41. 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(21): 6177–6196
42. Kuhn B, Gerber P, Schulz-Gasch T, Stahl M (2005) Validation and use of the MM-PBSA approach for drug discovery. J Med Chem 48(12):4040–4048

43. Thompson DC, Humblet C, Joseph-McCarthy D (2008) Investigation of MM-PBSA rescoring of docking poses. J Chem Inf Model 48(5):1081–1091
44. Ruvinsky AM (2007) Role of binding entropy in the refinement of protein–ligand docking predictions: analysis based on the use of 11 scoring functions. J Comput Chem 28(8):1364–1372
45. Ruvinsky AM, Kozintsev AV (2006) Novel statistical-thermodynamic methods to predict protein–ligand binding positions using probability distribution functions. Proteins 62(1):202–208
46. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. J Am Chem Soc 120(37): 9401–9409
47. Massova I, Kollman PA (1999) Computational alanine scanning to probe protein–protein interactions: a novel approach to evaluate binding free energies. J Am Chem Soc 121(36):8133–8143
48. Bashford D, Case DA (2000) Generalized born models of macromolecular solvation effects. Annu Rev Phys Chem 51(1): 129–152
49. Gilson MK, Rashin A, Fine R, Honig B (1985) On the calculation of electrostatic interactions in proteins. J Mol Biol 184(3): 503–516
50. Warwicker J, Watson HC (1982) Calculation of the electric potential in the active site cleft due to [alpha]-helix dipoles. J Mol Biol 157(4):671–679
51. Sharp KA, Nicholls A, Fine RF, Honig B (1991) Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. Science 252(5002):106–109
52. Sitkoff D, Ben-Tal N, Honig B (1996) Calculation of alkane to water solvation free energies using continuum solvent models. J Phys Chem 100(7):2744–2752
53. Guimarães CR, Cardozo M (2008) MM-GB/SA rescoring of docking poses in structure-based lead optimization. J Chem Inf Model 48(5):958–970
54. Huo S, Wang J, Cieplak P, Kollman PA, Kuntz ID (2002) Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. J Med Chem 45(7):1412–1419
55. Charifson PS, Corkery JJ, Murcko MA, Walters WP (1999) Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. J Med Chem 42(25):5100–5109
56. O'Boyle NM, Liebeschuetz JW, Cole JC (2009) Testing assumptions and hypotheses for rescoring success in protein–ligand docking. J Chem Inf Model 49:1871–1878
57. Krovat EM, Langer T (2004) Impact of scoring functions on enrichment in docking-based virtual screening: an application study on renin inhibitors. J Chem Inf Comput Sci 44(3):1123–1129
58. Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW, Taylor RD, Watson P (2004) Virtual screening using protein–ligand docking: avoiding artificial enrichment. J Chem Inf Comput Sci 44(3):793–806
59. Stahl M, Rarey M (2001) Detailed analysis of scoring functions for virtual screening. J Med Chem 44:1035–1042
60. Sutherland JJ, Nandigam RK, Erickson JA, Vieth M (2007) Lessons in molecular recognition. 2. Assessing and improving cross-docking accuracy. J Chem Inf Model 47(6):2293–2302
61. Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M (2004) Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. J Med Chem 47(1):45–55
62. Birch L, Murray CW, Hartshorn MJ, Tickle IJ, Verdonk ML (2002) Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. J Comput Aided Mol Des 16(12): 855–869
63. Murray CW, Baxter CA, Frenkel AD (1999) The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. J Comput Aided Mol Des 13(6):547–562
64. Ragno R, Frasca S, Manetti F, Brizzi A, Massa S (2005) HIV-reverse transcriptase inhibition: inclusion of ligand-induced fit by cross-docking studies. J Med Chem 48(1):200–212
65. May A, Zacharias M (2008) Protein–ligand docking accounting for receptor side chain and global flexibility in normal modes: evaluation on kinase inhibitor cross docking. J Med Chem 51(12):3499–3506
66. Jain AN (2009) Effects of protein conformation in docking: improved pose prediction through protein pocket adaptation. J Comput Aided Mol Des 23(6):355–374
67. Ferrari AM, Wei BQ, Costantino L, Shoichet BK (2004) Soft docking and multiple receptor conformations in virtual screening. J Med Chem 47(21):5076–5084
68. Claussen H, Buning C, Rarey M, Lengauer T (2001) FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 308(2):377–395
69. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS (2002) Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins 46(1):34–40
70. Knegtel RM, Kuntz ID, Oshiro CM (1997) Molecular docking to ensembles of protein structures. J Mol Biol 266(2):424–440
71. Leach AR (1994) Ligand docking to proteins with discrete side-chain flexibility. J Mol Biol 235(1):345–356
72. Zavodszky MI, Kuhn LA (2005) Side-chain flexibility in protein–ligand binding: the minimal rotation hypothesis. Protein Sci 14(4):1104–1114
73. Leach AR, Lemon AP (1998) Exploring the conformational space of protein side chains using dead-end elimination and the A* algorithm. Proteins 33(2):227–239
74. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK (2004) Testing a flexible-receptor docking algorithm in a model binding site. J Mol Biol 337(5):1161–1182
75. Cavasotto CN, Abagyan RA (2004) Protein flexibility in ligand docking and virtual screening to protein kinases. J Mol Biol 337(1):209–225
76. Barril X, Morley SD (2005) Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. J Med Chem 48(13):4432–4443
77. Ruckle T, Schwarz MK, Rommel C (2006) PI3 K[gamma] inhibition: towards an 'aspirin of the 21st century'? Nat Rev Drug Discov 5(11):903–918
78. Macromodel. Schrodinger LLC, New York (2005)
79. Schrodinger (2007) Maestro 8.0. Manual, p 294
80. In: iResearchLibrary, ChemNavigator, San Diego, CA
81. OEChemTK (2008) OpenEye Scientific Software, Inc, Santa Fe, NM
82. OMEGA (2007) OpenEye Scientific Software: Santa Fe, NM
83. Glide. In: Manual. Schrodinger, LLC, Portland, OR, USA, p 112 (2007
84. Prime (2007) Schrodinger LLC, New York
85. Feher M (2006) Consensus scoring for protein–ligand interactions. Drug Discov Today 11(9–10):421–428
86. Jain AN, Nicholls A (2008) Recommendations for evaluation of computational methods. J Comput Aided Mol Des 22(3–4):133–139
87. Benchware DataMiner (2007) Tripos, L.P. Saint Louis, MO
88. 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
89. Mpamhanga CP, Chen B, McLay IM, Ormsby DL, Lindvall MK (2005) Retrospective docking study of PDE4B ligands and an analysis of the behavior of selected scoring functions. J Chem Inf Model 45(4):1061–1074