J Comput Aided Mol Des (2010) 24:1053–1062
DOI 10.1007/s10822-010-9394-9

Biased retrieval of chemical series in receptor-based virtual screening

Natasja Brooijmans · Jason B. Cross · Christine Humblet

Received: 28 April 2010 / Accepted: 19 October 2010 / Published online: 30 October 2010

Abstract Using the kinases in the DUD dataset and an in-house HTS dataset from PI3K-γ, receptor-based virtual screening experiments were performed using Glide SP docking. While significant enrichments were observed for eight of the nine targets in the set, more detailed analyses highlighted that much of the early enrichment (10–80%) is the result of retrieval of a single cluster of active compounds. This biased retrieval was not necessarily due to early enrichment of the cluster containing the co-crystallized ligand. Virtual screening validation studies could thus benefit from including cluster-based analyses to assess enrichment of diverse chemotypes.

Keywords Virtual screening · Protein kinases · Binding sites · Models: molecular · Crystallography: X-ray · Docking · Protein kinase inhibitors

Natasja Brooijmans and Jason B. Cross contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-010-9394-9) contains supplementary material, which is available to authorized users.

N. Brooijmans (✉)
Wyeth Research, Chemical Sciences, 401 N. Middletown Road, Pearl River, NY 10965, USA
e-mail: reprints@prodigy.net

J. B. Cross
Wyeth Research, Chemical Sciences, 500 Arcola Road, Collegeville, PA 19426, USA

C. Humblet
Wyeth Research, Chemical Sciences, 865 Ridge Road, Princeton, NJ 08543, USA

Present Address:
N. Brooijmans
Novartis Institutes for Biomedical Research, 250 Massachusetts Avenue, Cambridge, MA 02139, USA

J. B. Cross
Cubist Pharmaceuticals, Inc., 65 Hayden Avenue, Lexington, MA 02421, USA

Introduction

The rapid increase in the number of protein structures populating the Protein Data Bank (PDB) [1], as well as proprietary structures solved by pharmaceutical companies, has made high throughput molecular docking an important tool in drug discovery research [2]. A common application of these techniques is structure-based virtual screening (SBVS), in which subsets of compounds from large collections are prioritized for experimental testing. These prioritized compound sets are often enriched with potential leads [3]. Several recent studies have evaluated the virtual screening performance of different molecular docking programs [4–13], and while the success rates greatly vary from one docking program to another depending on the target, some trends for success rates with specific protein families are beginning to emerge [4]. The focus of this work is on SBVS against kinase targets.

The discovery of kinase inhibitors is of great interest to the pharmaceutical industry and SBVS is regularly used to identify novel leads for these targets [14, 15]. The kinase catalytic domain consists of an N-terminal and C-terminal lobe with the ATP binding site sandwiched in between the lobes (Fig. 1). The two lobes are connected through the hinge region, which forms critical hydrogen bonding interactions to the substrate ATP and virtually all ATP-competitive kinase inhibitors (Fig. 2). SBVS has successfully identified

lead compounds for many kinase targets in oncology (BCR-Abl [16], ERK2 [17, 18], EphB2 [19], CDK2 [20], EGFR [21, 22], Aurora A [23] and B [24], Chk1 [25], Pim-1 [26], Met [27]) and anti-infectives (YycG [28], YpkA [29], MtSk [30]), as well as for targets with a broader range of potential applications, such as CK2 [31, 32] (oncology, anti-infective, inflammation), CK1δ [33] (oncology, anti-infective, neuroscience), and GSK-3β [34] (diabetes, osteoporosis, neuroscience). SBVS has also been used to develop validated kinase-targeted compound libraries [35, 36].

Docking programs essentially consist of two main components, a sampling algorithm that generates various putative protein–ligand complexes and a scoring function that evaluates each inhibitor-bound complex in terms of steric and physical–chemical fit to the binding site. Numerous docking programs have been developed since the first small molecule docking algorithm was published in the early 1980's [37]. Despite many advances in both scoring functions and sampling algorithms, the receptor is still most often considered to be rigid, especially in the SBVS setting where hundreds of thousands, if not millions, of ligands are to be evaluated. Induced fit effects observed for many protein structures upon ligand binding are not usually treated explicitly [38].

In cognate ligand docking the co-crystallized complex is essentially separated and the ligand is docked back into the protein structure. The rigid receptor assumption is in this case valid, since all the side chains are positioned correctly to accommodate the ligand and form favorable interactions. In this setting it has been shown that 60–80% of protein–ligand complexes can be predicted correctly using various docking methods [10, 12, 39]. Unfortunately, in SBVS the more relevant experiment is that of non-cognate ligand docking or cross-docking. In cross-docking the ligand is docked against an X-ray structure of its target that was co-crystallized with a different compound or in the apo form. In this situation all rigid receptor docking methods perform significantly less well, with success rates generally well below 50% [10, 12, 40, 41]. The cognate docking success rates can be further enhanced by optimizing the protein–ligand complex before embarking on re-docking the ligand [42, 43], again emphasizing that the rigid receptor assumption used by most docking algorithms is not valid. Although a number of groups have focused on including receptor flexibility [44–47] these methods are generally too slow to be used in high-throughput SBVS settings.

The known difficulties with non-cognate docking led us to investigate results from SBVS experiments by analyzing enrichments with and without analogs of the co-crystallized ligand, and with and without large clusters of compounds. Although many aspects of virtual screening validation studies have been addressed in the literature in recent years, including physical–chemical properties of actives and decoys, [48, 49], and the need for publicly-available benchmarking sets like the DUD dataset [49], the influence of retrieval of clusters of compounds on the measured enrichment factors has not been explicitly

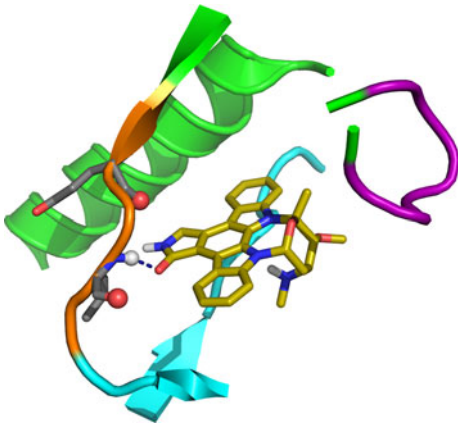

Fig. 2 Close-up of ATP-binding site of PI3K-γ in complex with staurosporine (1E8Z.pdb). Close-up of the ATP binding site of kinases highlighting the hydrogen bonding interaction of staurosporine with the backbone of the hinge region. The backbone-NH of the central hinge region donates a hydrogen bond to staurosporine in the PI3K-γ structure. The two other potential hydrogen bonding partners are the carbonyl atoms of the hinge backbone and are shown in sphere representation

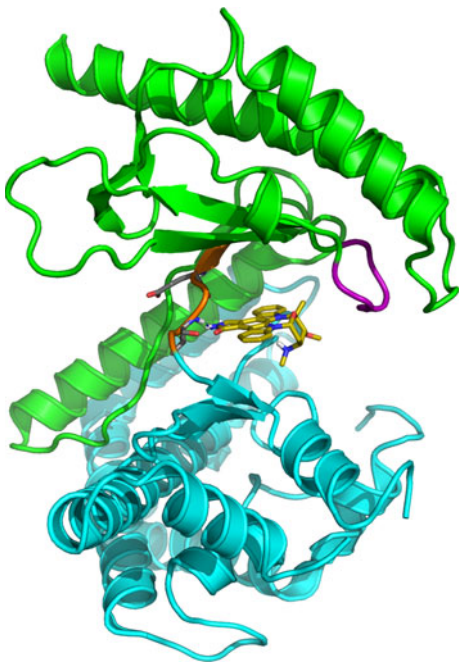

Fig. 1 Overview of kinase catalytic domain structure of PI3K-γ in complex with staurosporine (1E8Z.pdb [63]). Staurosporine is shown in *yellow* stick representation. The N-terminal lobe of the catalytic domain is shown in *green* and the C-terminal lobe is shown in *cyan*. The hinge region connecting the two lobes is highlighted in *orange*. The glycine-rich loop is shown in *purple*

investigated. Our results show that a single cluster of compounds, i.e. a chemical series, can make a large contribution to the calculated enrichment. Somewhat unexpectedly, the largest cluster in the top 1% is not necessarily chemically similar to the co-crystallized ligand. These results suggest that an assessment of the diversity of retrieved hits needs to be taken into account when evaluating the performance of SBVS experiments.

Methods and datasets

Diverse kinase dataset and virtual screening

The kinase subset in the Directory of Useful Decoys [49] (DUD) was used as a virtual screening dataset covering a range of diverse kinases. This dataset consists of eight kinase structures, CDK2, EGFr, FGFr1, p38 MAP, PDGFrb, SRC, TK, and VEGFr2, as well as active and decoy compounds associated with each target. The PDB accession codes for the targets are listed in Table 1. Decoys in the DUD were chosen to have physical properties similar to the active compounds, while remaining topologically distinct based on Tanimoto coefficients calculated using CACTVS fingerprints [50]. This makes the DUD dataset a challenge and a valuable benchmark for virtual screening, despite known issues and limitations [50]. In this study, only the decoys associated with a target were docked to that target (DUD-self), rather than all decoys for all targets.

All protein, active ligand, and decoy structures were downloaded from the publicly available DUD website (http://dud.docking.org). Protein preparation was performed with the Protein Preparation Wizard workflow available in Maestro [51]. Hydrogen atoms were added to the protein and subsequently minimized, keeping the heavy atom coordinates fixed, and in the absence of the cognate ligand. Protonation states for histidine residues and side chain rotamers for asparagine and glutamine residues were visually inspected and adjusted accordingly. The active and decoy ligand datasets were used as provided through the DUD website. The DUD kinase datasets vary from 22 to 444 actives and 784 to 14,894 decoys across the 8 kinase targets (see Table 1), representing a 20-fold range in dataset size. As described in Huang et al. [49], all ligand structures were generated using Corina and protonation/tautomeric states were calculated in the pH range 5.75–8.25 using LigPrep [52].

Table 1 Enrichments for the diverse kinases from the DUD at several early false positive rates

Target	PDB ID	DUD composition		Enrichment factors				
		Actives	Decoys	0.5%	1%	2%	5%	10%
CDK2	1CKP	50	1,779	23.3	23.3	17.5	9.6	5.6
EGFr	1M17	444	14,894	27.8	30.3	24.3	14.1	7.9
FGFr1	1AGW	118	4,205	5.0	3.3	2.9	2.5	2.2
p38 MAP	1KV2	256	8,387	4.7	3.5	2.4	1.9	1.4
PDGFrb	Model	157	5,614	19.1	9.5	4.8	2.1	1.1
SRC	2SRC	155	5,793	54.5	32.3	17.9	8.7	4.8
TK	1KIM	22	784	0.0	0.0	0.0	0.9	0.9
VEGFr2	1VR2	74	2,641	16.2	9.5	8.8	6.8	5.3

Docking was carried out using GLIDE v4.5 [53, 54]. Grids were generated by defining a receptor site centered at the centroid of the cognate ligand coordinates as defined by the X-ray co-structure. Standard precision GLIDE was used for docking, with default settings. For each compound, all tautomeric and protonation states were docked, however, only the best scoring example of each compound was retained for the virtual screening analysis. Compounds were then ranked according to their docking score.

PI3K-γ virtual screening

In addition to the kinases from the DUD set, both prospective and retrospective SBVS experiments were done on multiple crystal structures of a single target, namely phosphoinositide 3-kinase gamma (PI3K-γ). Based on visual assessment, five crystal structures were chosen from a larger set of eighteen public domain and proprietary structures available at the time. The chosen structures were diverse both from the perspective of the co-crystallized ligands (maximum similarity 0.21 as measured using the FCFP_4 fingerprint) and from the binding site, where several critical binding site residues were positioned differently [55]. The five chosen structures include a staurosporine complex structure (1E8Z.pdb), a pyrazolo-pyrimidine structure (labeled S1 or ligand 1 below, 3IBE.pdb), and three unpublished structures (labeled S2 to S4 or ligand 2 to ligand 4 below).

Hydrogens were added to the protein structures within the Maestro GUI and were subsequently prepared using the Protein Preparation Wizard workflow in Maestro [56]. Protonation states for histidine and rotamers for asparagine and glutamine residues were determined by the Protein Preparation Wizard workflow and inspected visually. Manual adjustment was not required. Hydrogens and heavy atoms were minimized in the presence of the cognate ligand.

For retrospective virtual screening an internal high throughput screening (HTS) data set was utilized. The $\sim$533,000 compounds that were tested in the HTS were filtered for lead-like properties (molecular weight between 180 and 400, number of rotatable bonds less than 7, less than 3 ionizable groups, only 1 chiral center, no reactive functionalities) and subsequently prepared using

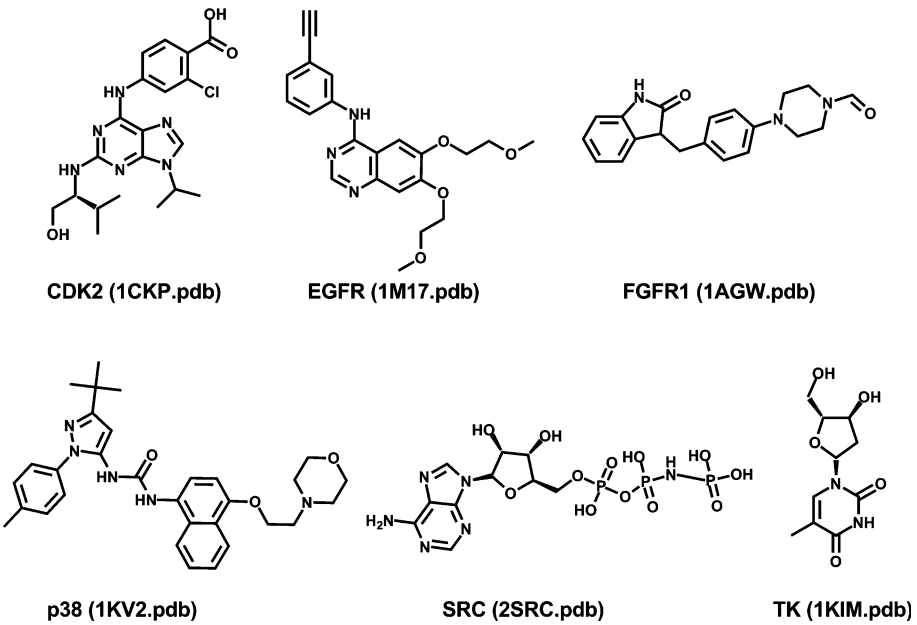

Fig. 3 Co-crystallized ligands present in the X-ray structures of the DUD targets used for docking

the LigPrep utility [57]. LigPrep was used to annotate tautomers and ionization states at a pH range between 5 and 9 with Ionizer. Application of all the filters resulted in a lead-like HTS set of ~179,000 compounds, which was subsequently expanded to 289,000 representations through OMEGA chiral expansion [58] and subsequent annotation of ionization and tautomeric states using LigPrep. HTS compounds were identified as "hits" or "actives" based on the confirmation screen performed after the initial HTS screen. HTS confirmation was done in triplicate and the average percent inhibition was used to determine activity. Both the initial screen and the confirmation screen utilized a cut-off of 20% inhibition at 10 μM to identify hits. The cut-off was based on the interquartile range (IQR) and inhibition equal to or greater than 3 IQR was deemed statistically significant. Of the ~179,000 tested lead-like HTS compounds, 558 actives were found. Lead-like compounds with less than (average) 20% inhibition in either the primary or the confirmation assay were used as decoys in the retrospective VS experiments.

Retrospective docking studies were performed using GLIDE v4.5 [59]. Grids were prepared centered on the center-of-mass for each of the co-crystallized ligands in the five X-ray structures that were used. The Virtual Screening Workflow (VSW) in Maestro was used to dock and score the lead-like HTS set first with the HTVS function in GLIDE, followed by a re-docking and scoring of the top 10% from HTVS with GLIDE Standard Precision (SP). The top-scoring 10,000 molecules from Glide-SP were retained. As for the DUD dataset, GLIDE was run using default parameters. Added to the HTS set were 41 active (≤20 μM) and 6 inactive analogs of the co-crystallized ligand of S1, 51 active and 9 inactive analogs of the co-crystallized ligand of S3, and 28 active and 18 inactive analogs of the co-crystallized ligand of S4. All added active and inactive analogs passed the lead-like filters. The VS analysis was performed for each structure with and without inclusion of the analogs of the co-crystallized chemical series and both the HTS actives and the active analogs of the co-crystallized ligands were considered to be actives in the enrichment analysis (Fig. 3).

For prospective VS the lead-like corporate VS database was screened. The lead-like filters used were the same as for the retrospective screens described above, but included a compound availability filter. This library consisted of ~225,000 compounds, which were again expanded using OMEGA and LigPrep as described above for the PI3K-γ retrospective screens. Similarly to the retrospective screens, the VSW workflow was used to first dock with GLIDE HTVS, followed by GLIDE SP re-docking of the top 10% from HTVS. The top 10,000 scoring molecules from the SP stage were saved for further (visual) analysis. Selected compounds were tested in the same enzyme assay as used during the HTS campaign and hits were identified using the same criteria. Hits were further confirmed by generating a full dose–response (11-points).

During the visual assessment of the hit lists from the prospective virtual screens, a number of known analogs of ligands 1, 3 and 4 were found within the top 300. Hit rates for the prospective SBVS experiments were analyzed with and without the already known actives present.

Evaluation of virtual screening performance

Early enrichments for all retrospective virtual screens were computed using the ROC enrichment factor (EF) [60],

which utilizes information from the Receiver Operating Characteristic (ROC) curve [61] rather than a recovery curve. This metric differs from the standard enrichment factor that is often encountered in virtual screening studies by referring to the fraction of decoys considered (i.e., false-positive rate or FPR) rather than the fraction of all compounds considered.

Success rates for the prospective virtual screens were measured using the hit-rate, which is the percentage of compounds selected for experimental testing that showed the desired activity.

Results and discussion

Diverse kinase dataset

The EFs at several early FPRs for the DUD kinases listed in Table 1 are values taken from a recently published study comparing docking methods [4] and serve as a baseline for additional analysis presented in this work. The enrichment varied widely between the kinases, but only one of the targets, thymidine kinase (TK), consistently yielded an enrichment factor lower than expected for random selection. CDK2, EGFr, and SRC all produced very high enrichments at 0.5 and 1%, in excess of 20-fold at both of these FPRs. The enrichments for PDGFrb and VEGFr2 were marginally lower and showed a more marked decline between the 0.5 and 1% FPRs, dropping from 19.1 to 9.5 and 16.2 to 9.5, respectively. While the enrichments for FGFr1 and p38 MAP were more modest, ranging from 3.3 to 5.0 at the 0.5 and 1% FPRs, their enrichments stayed consistently superior to random selection up to a 10% FPR. TK was the only kinase target in the DUD for which Glide SP failed to identify any active compounds at the early FPRs, and even at the 5 and 10% FPR levels the performance was still worse than random compound selection.

Further analysis of these virtual screening results was performed by removing clusters of active ligands and recomputing the enrichments at 1% FPR (corresponding to the top 10–150 scoring compounds depending on the dataset). Good and Oprea performed clustering of the DUD using reduced graphs; however, a lead-like filter was also applied that resulted in the removal of a significant number of actives [62]. This filtering process had the effect of removing some of the closest structural analogs of the p38 MAP kinase X-ray bound ligand from the dataset, therefore an alternate clustering strategy was employed in this study. (Enrichments using the DUD reduced graph clusters are included in Supplementary Table I.) Compounds were clustered using MACCS keys and Tanimoto cut-offs using the MOE fingerprint clustering algorithm without lead-like filtering based on physical properties. The similarity and overlap cut-offs for clustering were each set to 0.75. Three different methods were used to select the clusters for removal: (a) the cluster containing the X-ray co-ligand if present in the list of actives, (b) the largest cluster from the entire set of active ligands, and (c) the cluster with the most representatives in the top 1% FPR. In each case all compounds in the cluster were removed, including the cluster centroid. Table 2 lists the total sizes of these clusters, the number of compounds in the selected cluster within the top 1% FPR, and the EFs with each of these clusters removed (cluster membership for all the DUD actives can be found in the Supplementary Table II).

Table 2 Cluster sizes and enrichment factors at 1% FPR for removal of X-ray ligand and largest clusters

Target	Cluster size			Cluster size in top 1%			EF 1% with cluster removed[c]		
	X-ray ligand	Largest	Largest in top 1%	X-ray ligand	Largest	Largest in top 1%	X-ray ligand	Largest	Largest in top 1%
CDK2	4	6	4	2	0	2	17.8 (23.6%)	23.3 (0.0%)	17.8 (23.6%)
EGFr	2	42	34	0	5	14	30.3 (0.0%)	29.3 (3.3%)	27.2 (10.2%)
FGFr1	11	25	20	0	1	2	3.3 (0.0%)	2.5 (24.2%)	1.7 (48.5%)
p38 MAP	21	34	28	0	1	1	3.5 (0.0%)	3.1 (11.4%)	3.1 (11.4%)
PDGFrb	–[a]	18	7	–[a]	0	7	–[a]	9.5 (0.0%)	1.9 (80.0%)
SRC	–[a]	35	35	–[a]	20	20	–[a]	19.3 (40.2%)	19.3 (40.2%)
TK	15	15	–[b]	0	0	0	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)
VEGFr2	–[a]	6	2	–[a]	0	2	–[a]	9.5 (0.0%)	6.9 (27.4%)

[a] These targets had no clusters containing X-ray ligands. The PDGFrb structure is a homology model, the VEGFr2 structure has no co-crystallized ligand, and there were no compounds similar to the SRC ligand (AMPPNP) in the dataset

[b] There were no actives identified in the top 1%

[c] Values in parentheses represent the percent reduction in enrichment when the cluster is removed compared to the enrichment with the cluster present

Removal of the cluster containing the X-ray ligand was only possible for five of the eight kinase targets, as the PDGFrb protein structure in the DUD is a homology model with no associated ligand, the VEGFr2 crystal structure (1KV2) is an apo structure without a co-crystallized ligand, and the SRC crystal structure (2SRC) contains a co-crystallized ligand, AMPPNP, that is not part of the active ligand set. For the other five targets, only CDK2 had active ligands from the X-ray ligand cluster in the top 1% FPR. In this case, the enrichment dropped from 23.3 to 17.8-fold, a reduction of 23.6%. The other targets had no actives from the X-ray ligand cluster ranked sufficiently high to impact the enrichment factor at 1% FPR.

The removal of the largest cluster of actives resulted in more targets having reduced enrichment factors at 1% FPR, but since the largest cluster does not necessarily have representation at the top of the ranked list of docked compounds, not all targets exhibited this effect. In addition to TK, which does not have actives present in the top 1%, CDK2, PDGFrb, and VEGFr2 also saw no reduction in their enrichments due to a lack of representation of the largest cluster. For these four targets, an important chemical series was thus completely missed. EGFr and p38 MAP showed a modest decrease at 1% FPR, with enrichments reduced by 3.3 and 11.4%, respectively. FGFr1 and SRC had much larger declines in enrichment at 24.2 and 40.2%, respectively.

Each of the DUD kinase targets, other than TK, necessarily has a cluster of actives with the most representatives in the top 1% FPR; therefore, each of the targets shows a reduction in enrichment when these clusters are removed from consideration. The smallest decreases were seen for EGFr and p38 MAP at 10.2 and 11.4%, respectively. A larger effect was seen for CDK2 and VEGFr2 (23.6 and 27.4%, respectively), while significant reduction in enrichment was observed for FGFr1 and SRC (48.5 and 40.2%, respectively). There was a near elimination of enrichment for PDGFrb with the enrichment factor dropping from 9.5 to 1.9-fold, a reduction of 80.0%.

PI3K-γ retrospective screening

Early recovery (0.05% FPR, $\sim$100 top-scoring compounds) of active compounds identified through HTS varied approximately 2.5-fold between the worst performer (Structures 1 and 5) and the best performer (Structure 3), but for all structures better than random performance was observed (Table 3a). The spread in EFs was smaller at higher FPRs, with significantly better than random enrichment observed at all levels. Practically, however, the most relevant EFs with this large dataset are those at FPR of 0.05% ($\sim$100 top-scoring compounds) and 0.5% ($\sim$1,000 top-scoring compounds) as experimental follow-up is usually on the order of hundreds or perhaps a couple of thousand compounds.

Similarly to the analysis performed with the DUD dataset, the PI3K-γ HTS actives were clustered using the MOE fingerprint method and the largest highly ranked cluster was removed and enrichments were recalculated. The largest cluster contained 13 members and was the top-ranked cluster when docking to Structures 1, 4, and 5. The largest highly ranked cluster when docking to Structures 2 and 3 contained 10 members, as the cluster with 13 members was not found among the top 10,000 ranked

Table 3 (a) Enrichments for PI3K-γ retrospective screening at several early false positive rates for each structure and (b) after removal of the largest top-scoring cluster

Structure	0.05%	0.5%	1%	2%	5%
(a)					
Structure 1	7.1	16.1	14.0	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	15.0	11.8	9.3	7.0
Structure 5	7.1	17.2	14.7	11.5	7.7
(b)					
Structure 1	7.3 (+2.7)	14.7 (8.7)	13.4 (4.3)	10.4 (1.9)	6.7 (0.0)
Structure 2	14.5 (+1.4)	12.0 (16.0)	10.2 (11.3)	6.5 (8.5)	4.2 (4.5)
Structure 3	18.2 (+2.2)	20.4 (5.1)	16.2 (6.9)	10.1 (5.6)	6.0 (3.2)
Structure 4	11.0 (+2.8)	15.0 (0.0)	11.9 (+0.8)	9.1 (2.2)	6.9 (1.4)
Structure 5	7.3 (+2.8)	16.1 (6.4)	14.1 (4.1)	11.2 (2.6)	7.6 (1.3)

In between brackets the percentage change in enrichment factor compared to the full PI3K-γ HTS dataset (a) is indicated. A positive sign for the percentage change in EF indicates that removal of the largest cluster led to an increase in EF, whereas no sign in front of the percentage change indicates enrichment was reduced upon removal of the cluster

compounds. Removal of the largest highly ranked cluster resulted in decreases in enrichments at the 0.5 and 1.0% FPR, but because the largest clusters were not ranked among the top-scoring compounds (<250), enrichment at the 0.05% FPR generally increased. The decrease in enrichment at the 0.5 and 1.0% FPR rate is between 0.6 and 2.3-fold, which constitutes a percentage change in enrichment factor between 4 and 16%. This is consistent with the changes in enrichment observed for the DUD dataset upon removal of the largest cluster, despite the fact that the clusters in the PI3K-γ HITS dataset are significantly smaller in size.

The PI3K-γ HTS set did not contain any analogs of the co-crystallized ligands, however, in-house data on analogs of the ligands of Structure 1, 3, and 4 was available. Addition of these analogs to the HTS set, after application of the same lead-like filters as used to prepare the HTS dataset and recalculation of the EFs, resulted in significantly higher EFs (Table 4). The EF dropped from 29.9 to 7.1 (76.3% change) between the set with and without the analogs of the ligand of Structure 1. For Structure 3 an $\sim$ 85% drop in the EF is observed and for Structure 4 the change in EF is 37% at the 0.05% FPR upon removal of the X-ray ligand analogs. At 1% FPR the percentage changes in EF varied between 11 and 42% (Fig. 4).

PI3K-γ prospective screening

While the ligand preparation and docking procedures for the PI3K-γ prospective SBVS experiments were identical to those used in the retrospective screens, experimental follow-up was not based on the GLIDE scores alone. Instead, the top 150 scoring compounds against each structure were visualized and their compatibility with the binding site was evaluated. Since this target is a kinase, an important criterion was the presence of hydrogen bonding interactions with the so-called hinge region. Known inhibitors of PI3K-γ that scored among the top 150 were removed from the list before experimental follow-up. Finally, prioritized compounds were assessed by the chemistry project leader for chemical desirability.

A total of 127 compounds were tested in the PI3K-γ enzyme assay, resulting in 3 compounds with IC50 $\leq$ 20 μM (Table 5). Each of the tested compounds was prioritized by a single crystal structure. The average hit rate across the five structures is only 2.4%. Strikingly, several of the structures did not yield any hits, although enrichment for all receptors in the retrospective experiments was above random. However, when the known actives that scored well were taken into account for Structures 1, 3 and 5, hit rates were significantly higher, ranging from 17 to 28%. Comparison of the hit rates with and without the known actives of the co-crystallized ligand shows a drop between 72 and 100% when the known analogs are removed (Table 6).

Conclusions

In both the retrospective and prospective SBVS studies with the DUD kinase targets and PI3K-γ, Glide SP consistently yielded high early enrichment for all targets except TK. Detailed analyses of the composition of the retrieved actives highlights that generally a single cluster of related compounds accounts for a significant percentage (10–80%) of total enrichment. While one might expect this to be a result of a higher preference for the X-ray ligand and its analogs, as exemplified by CDK2 and PI3K-γ, this is not necessarily the case. The removal of a single cluster consequently can have a significant effect on the EF, resulting in decreases between 10 and 80% at the 1% FPR for the DUD and PI3K-γ. The impact can be even larger at earlier enrichment, as exemplified by the PI3K-γ retrospective screens where at 0.05% FPR decreases in EF between 37 and 85% are observed. Although it is possible that minimization of the PI3K-γ heavy atoms with the ligand present contributes to the large change in EF when

Table 4 Enrichments at various false positive rates for PI3K-γ retrospective screening with and without active analogs of the co-crystallized ligand

Compound set	0.05%	0.5%	1%	2%	5%
S1 HTS + S1 analogs[a]	29.9	28.0	19.9	30.0	7.6
S1 HTS, no S1 analogs[b]	7.1 (76.3%)	16.1 (42.5%)	14.0 (29.6%)	10.6 (18.5%)	6.7 (11.8%)
S3 HTS + S3 analogs[a]	114.4	42.0	30.2	18.1	10.2
S3 HTS, no S3 analogs[b]	17.8 (84.4%)	21.5 (48.8%)	17.4 (42.4%)	10.7 (40.9%)	6.2 (39.2%)
S4 HTS + S4 analogs[a]	17.0	16.7	13.3	10.0	7.4
S4 HTS, no S4 analogs[b]	10.7 (37.1%)	15.0 (10.2%)	11.8 (11.3%)	9.3 (7.0%)	7.0 (5.4%)

[a] The HTS set with added analogs of the co-crystallized ligand (active and inactive)

[b] The HTS set without added analogs of the co-crystallized ligand

In between brackets the percentage change in enrichment factor is indicated

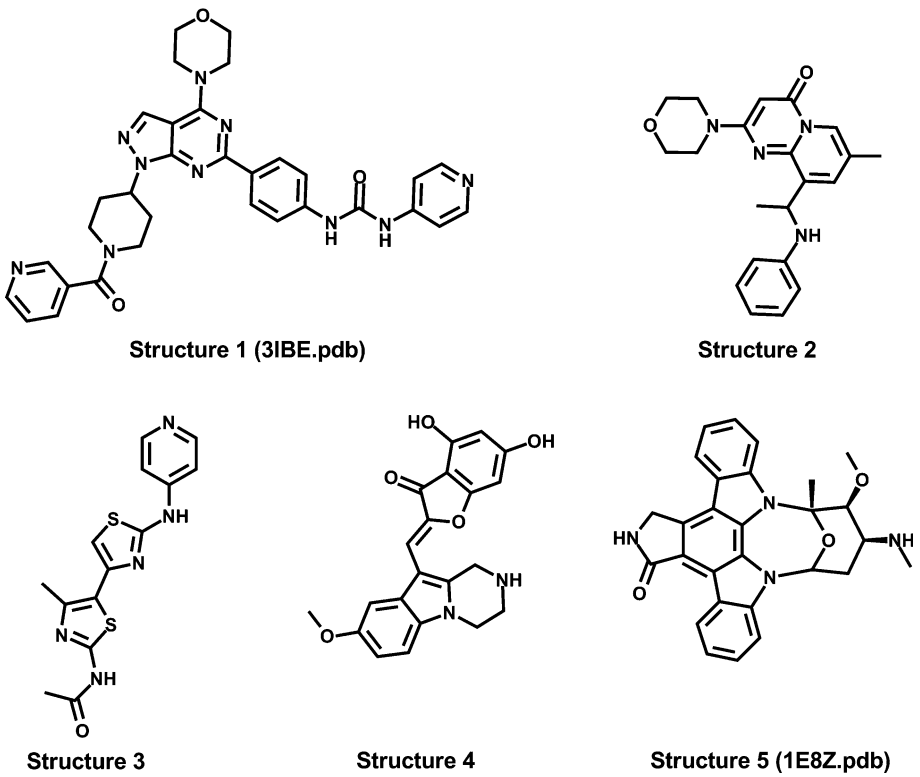

Fig. 4 Structures of co-crystallized ligands of PI3K-γ X-ray structures used in the virtual screening experiments

Table 5 Hit rates for PI3K-γ prospective screening for each structure and overall and average hit rate

Structure	Tested	Selected for IC50	IC50 $\leq$ 20 μM	Hit rate
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

Table 6 Hit rates for PI3K-γ prospective screening for Structure 1, 2 and 3 with and without addition of known analogs of the co-crystallized ligand for each structure to the corporate virtual screening collection

Structure	Compound set	Tested	IC50 $\leq$ 20 μM	Hit rate
Structure 1	CORP-VS	37	9	24.3
Structure 1[a]	No S1 analogs	30	2	6.7 (72.4%)
Structure 3	CORP-VS	34	6	17.6
Structure 3[a]	No S3 analogs	27	1	3.7 (79.0%)
Structure 4	CORP-VS	11	3	27.3
Structure 4[a]	No S4 analogs	8	0	0.0 (100%)

[a] For the results without structural analogs of the co-crystallized ligand the percentage change in Hit Rate is compared to the full compound set is shown in between brackets

the active and inactive analogs of the X-ray ligand are removed, based on the fact that a wide variety of weak, active chemotypes are enriched, we believe that this is only a contributing factor and not the sole determinant of the change in EF. Another contributing factor to the large change in EF with X-ray series removal could be that the X-ray ligand analogs are highly optimized and potent PI3K-γ inhibitors, while the HTS actives are fairly weak and not optimized for PI3K-γ in any way. This is further illustrated by the removal of the largest highest ranked clusters of HTS hits from PI3K-γ. The changes in enrichment factor are now smaller, although still significant (up to 16% reduction in EF).

The biased enrichment of a particular chemical series is not necessarily a result of co-crystallization, as evidenced by the decrease in EF of more than 25% in VEGFr2, which was an apo structure. Similarly, PDGFrb, an unbound homology model, saw a reduction of 80% when the largest cluster in the top 1% FPR was removed.

The PI3K-γ prospective and retrospective studies utilizing multiple co-crystal structures highlight that different crystal structures prioritize different chemical series, and are thus complementary to each other.

While none of the recent, large-scale virtual screening studies have analyzed the impact of clusters on enrichments explicitly, careful analysis of the data in Tables 5 and 6 of Warren et al. [12] shows that essentially all docking algorithms have targets with moderate to good

enrichment factors where over 10% of the test set needs to be sampled to recover a representative of each active chemotype. This suggests that early enrichment in these cases is dominated by a particular chemotype.

In summary, we have highlighted the need for more in-depth analyses of SBVS validation studies, as it is not only important to have high early enrichment of actives, but also to retrieve a wide variety of chemotypes for further follow-up. The design of virtual screening benchmarking sets would also benefit from inclusion of a wide variety of chemotypes, including analogs of the co-crystallized ligands.

Acknowledgments The authors thank David C. Thompson, Brajesh K. Rai, J. Christian Baber, Kristi Yi Fan, and Yongbo for participating in the virtual screening studies using the DUD.

References

1. Berman H, Henrick K, Nakamura H, Markley JL (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res 35:D301–D303
2. Joseph-McCarthy D, Baber JC, Feyfant E, Thompson DC, Humblet C (2007) Lead optimization via high-throughput molecular docking. Curr Opin Drug Discov Dev 10(3):264–274
3. Cavasotto CN, Orry AJW (2007) Ligand docking and structure-based virtual screening in drug discovery. Curr Top Med Chem 7(10):1006–1014
4. 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
5. Bissantz C, Folkers G, Rognan D (2000) Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. J Med Chem 43(25):4759–4767
6. Stahl M, Rarey M (2001) Detailed analysis of scoring functions for virtual screening. J Med Chem 44(7):1035–1042
7. Schulz-Gasch T, Stahl M (2003) Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. J Mol Model 9(1):47–57
8. Kellenberger E, Rodrigo J, Muller P, Rognan D (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins Struct Funct Bioinf 57(2):225–242
9. Perola E, Walters WP, Charifson PS (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins Struct Funct Bioinf 56(2): 235–249
10. 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
11. Kontoyianni M, Sokol Glenn S, McClellan Laura M (2005) Evaluation of library ranking efficacy in virtual screening. J Comput Chem 26(1):11–22
12. Warren GL, Andrews CW, Capelli A-M, 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
13. McGaughey GB, Sheridan RP, Bayly CI, Culberson JC, Kreatsoulas C, Lindsley S, Maiorov V, Truchon J-F, Cornell WD (2007) Comparison of topological, shape, and docking methods in virtual screening. J Chem Inf Model 47(4):1504–1519
14. Muegge I, Enyedy IJ (2004) Virtual screening for kinase targets. Curr Med Chem 11(6):693–707
15. McInnes C (2006) Improved lead-finding for kinase targets using high-throughput docking. Curr Opin Drug Discov Dev 9(3): 339–347
16. 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
17. 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
18. 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
19. 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
20. 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
21. 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
22. 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
23. 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
24. Fu D-H, Jiang W, Zheng J-T, Zhao G-Y, Li Y, Yi H, Li Z-R, Jiang J-D, Yang K-Q, Wang Y, Si S-Y (2008) Jadomycin B, an Aurora-B kinase inhibitor discovered through virtual screening. Mol Cancer Ther 7(8):2386–2393
25. 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
26. 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
27. 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
28. 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
29. 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
30. 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
31. 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
32. Golub AG, Yakovenko OY, Bdzhola VG, Sapelkin VM, Zien P, Yarmoluk SM (2006) Evaluation of 3-Carboxy-4(1H)-quinolones as inhibitors of human protein kinase CK2. J Med Chem 49(22): 6443–6450
33. 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
34. Kang NS, Lee GN, Kim CH, Bae MA, Kim I, Cho YS (2009) Identification of small molecules that inhibit GSK-3beta through virtual screening. Bioorg Med Chem Lett 19(2):533–537
35. Sun D, Chuaqui C, Deng Z, Bowes S, Chin D, Singh J, Cullen P, Hankins G, Lee W-C, Donnelly J, Friedman J, Josiah S (2006) A kinase-focused compound collection: compilation and screening strategy. Chem Biol Drug Des 67(6):385–394
36. Gozalbes R, Simon L, Froloff N, Sartori E, Monteils C, Baudelle R (2008) Development and experimental validation of a docking strategy for the generation of kinase-targeted libraries. J Med Chem 51(11):3124–3132
37. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161(2):269–288
38. Najmanovich R, Kuttner J, Sobolev V, Edelman M (2000) Side-chain flexibility in proteins upon ligand binding. Proteins 39(3): 261–268
39. Kontoyianni M, McClellan LM, Sokol GS (2004) Evaluation of docking performance: comparative data on docking algorithms. J Med Chem 47(3):558–565
40. Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW (2008) Protein-ligand docking against non-native protein conformers. J Chem Inf Model 48(11):2214–2225
41. 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
42. Jain AN (2008) Bias, reporting, and sharing: computational evaluations of docking methods. J Comput Aided Mol Des 22(3–4):201–212
43. Rapp CS, Schonbrun C, Jacobson MP, Kalyanaraman C, Huang N (2009) Automated site preparation in physics-based rescoring of receptor ligand complexes. Proteins 77:52–61
44. 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
45. Sherman W, Beard HS, Farid R (2006) Use of an induced fit receptor structure in virtual screening. Chem Biol Drug Des 67(1):83–84
46. Claussen H, Buning C, Rarey M, Lengauer T (2001) FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 308(2):377–395
47. Knegtel RM, Kuntz ID, Oshiro CM (1997) Molecular docking to ensembles of protein structures. J Mol Biol 266(2):424–440
48. 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
49. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. J Med Chem 49(23):6789–6801
50. Irwin JJ (2008) Community benchmarks for virtual screening. J Comput Aided Mol Des 22(3–4):193–199
51. Maestro (2006) Schrodinger, LLC, Portland, OR, USA
52. Schrodinger (2008) LigPrep 2.2. Manual. p 108
53. 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
54. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem 47(7):1750–1759
55. Brooijmans N, Humblet C (2010) Chemical space sampling by different scoring functions and crystal structures. J Comput Aided Mol Des 24(5):433–447
56. Maestro (2007) Schrodinger, LLC, New York, NY
57. Venturelli A, Tondi D, Cancian L, Morandi F, Cannazza G, Segatore B, Prati F, Amicosante G, Shoichet BK, Costi MP (2007) Optimizing cell permeation of an antibiotic resistance inhibitor for improved efficacy. J Med Chem 50(23):5644–5654
58. OMEGA (2007) OpenEye Scientific Software, Santa Fe, NM
59. Glide (2007) In: Manual. Schrodinger, LLC, Portland, OR, USA, p 112
60. Nicholls A (2008) What do we know and when do we know it? J Comput Aided Mol Des 22(3–4):239–255
61. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29–36
62. Good AC, Oprea TI (2008) Optimization of CAMD techniques 3. Virtual screening enrichment studies: a help or hindrance in tool selection? J Comput Aided Mol Des 22(3–4):169–178
63. Walker EH, Pacold ME, Perisic O, Stephens L, Hawkins PT, Wymann MP, Williams RL (2000) Structural determinants of phosphoinositide 3-kinase inhibition by wortmannin, LY294002, quercetin, myricetin, and staurosporine. Mol Cell 6(4):909–919