Published in final edited form as:
J Chem Inf Model. 2010 October 25; 50(10): 1839–1854. doi:10.1021/ci100235n.

Comprehensive structural and functional characterization of the human kinome by protein structure modeling and ligand virtual screening

Michal Brylinski and **Jeffrey Skolnick**
Center for the Study of Systems Biology, School of Biology, Georgia Institute of Technology, Atlanta, Georgia, USA

Michal Brylinski: michal@gatech.edu; Jeffrey Skolnick: skolnick@gatech.edu

Abstract

The growing interest in the identification of kinase inhibitors, promising therapeutics in the treatment of many diseases, has created a demand for the structural characterization of the entire human kinome. At the outset of the drug development process, the lead-finding stage, approaches that enrich the screening library with bioactive compounds are needed. Here, protein structure-based methods can play an important role, but despite structural genomics efforts, it is unlikely that the three-dimensional structures of the entire kinome will be available soon. Therefore, at the proteome level, structure-based approaches must rely on predicted models, with a key issue being their utility in virtual ligand screening. In this study, we employ the recently developed FINDSITE/Q-Dock Ligand Homology Modeling approach, which is well suited for proteome-scale applications using predicted structures, to provide extensive structural and functional characterization of the human kinome. Specifically, we construct structure models for the human kinome; these are subsequently subject to virtual screening against a library of more than 2 million compounds. To rank the compounds, we employ a hierarchical approach that combines ligand- and structure-based filters. Modeling accuracy is carefully validated using available experimental data with particularly encouraging results found for the ability to identify, without prior knowledge, specific kinase inhibitors. More generally, the modeling procedure results in a large number of predicted molecular interactions between kinases and small ligands that should be of practical use in the development of novel inhibitors. The dataset is freely available to the academic community a user-friendly web interface at http://cssb.biology.gatech.edu/kinomelhm/as well as the ZINC website (http://zinc.docking.org/applications/2010Apr/Brylinski-2010.tar.gz).

1. INTRODUCTION

One of the largest enzyme families, the protein kinase family, comprises about ~2% of the human proteome [1]. Each member of this family contains a highly conserved kinase catalytic domain responsible for the reversible phosphorylation of protein substrates, a major regulatory process in both prokaryotic and eukaryotic organisms [2, 3]. The transfer of the γ-phosphate of ATP to serine, threonine and tyrosine residues in many enzymes and receptors turns them on and off; thus, the dysfunction of kinase activity is implicated in various pathological conditions. The regulation of kinase activity has been recognized by the pharmaceutical industry as an important therapeutic strategy in the treatment of many

Correspondence to: Jeffrey Skolnick, skolnick@gatech.edu.

The full set of modeled protein structures, docked ligand conformations and compound rankings are available at http://cssb.biology.gatech.edu/kinomelhm/ and http://zinc.docking.org/applications/2010Apr/Brylinski-2010.tar.gz diseases including cancer, Alzheimer's disease, diabetes, inflammation, multiple sclerosis and cardiovascular disease [4–8]. Currently, an estimated one-third of drug discovery programs focus on protein kinases [9], with already approved drugs such as imatinib [10] (*Gleevec*, Novartis), gefitinib [11] (*Iressa*, AstraZeneca), lapatinib [12] (*Tykerb/Tyverb*, GlaxoSmithKline) or sunitinib [13] (*Sutent*, Pfizer). These are just a few of the more than a hundred successfully developed compounds with kinase inhibition as their mode of action [14].

To speed up the development of new biopharmaceuticals, computational techniques for the identification of lead compounds are widely used [15]. In particular, virtual screening, a technique that shows great promise for lead discovery, is becoming an integral part of modern drug design pipelines [16, 17]. Due to advances in computer technology resulting in constantly increasing computational power, virtual libraries comprising millions of compounds can be rapidly evaluated in silico prior to experimental screens and at the fraction of the cost. Virtual screening approaches, historically divided into ligand- and structure-based algorithms [18], prioritize drug candidates by estimating the probability of binding to the target receptor. Among many methods developed to date, docking-based techniques are valuable tools for lead identification [19]. These algorithms rank compounds by predicting the binding mode for a query molecule in the binding pocket of the target protein [20–22]; this is followed by the prediction of binding affinity from molecular interactions [23–25]. Recent successful applications of structure-based virtual screening to kinase targets include the identification of potent inhibitors for death-associated protein kinases (DAPKs) [26], protein kinase B (PKB/AKT) [27], Janus kinase 2 (JAK2) [28], Met tyrosine kinase (RTK Met) [29] and Aurora kinase A (AurA) [30].

Notwithstanding the practical value of virtual screening by ligand docking for lead identification, there are significant flaws in current methods. Most salient is the fact that the predicted binding affinity is strongly correlated with the molecular weight of the ligand, independent of whether or not the ligand really binds to its target [31, 32]. Furthermore, to achieve satisfactory performance, many commonly used docking algorithms require the X-ray structure of their receptor target, preferably in the ligand-bound conformational state [33]. Such high-resolution structural information is available only for the fraction of the druggable proteome. At 90% sequence identity, Figure 1 shows that the coverage of the human kinome by protein crystal structures from the PDB [34] is ~20%. On the other hand, the popularity of kinase inhibitors as novel therapeutics has significantly increased. Since 1995, when one of five published papers on inhibitor development was related to kinases, the interest in kinase inhibitors has grown significantly; in 2008, approximately one-third of publications reporting on inhibitor development can be linked to protein kinases (Figure 1, inset). This evident trend in pharmaceutical research creates a great demand for the structural data that would cover the entire human kinome. The gap between the availability of protein sequences and structures can be filled by protein structure prediction, particularly comparative modeling [35, 36]. For a target sequence, given a set of evolutionarily related protein structures, state-of-the-art template-based algorithms can construct a model whose quality is often comparable to that of a low-resolution experimentally determined structure [37]. However, despite having the correct global topology, theoretically predicted protein structures may still have significant structural inaccuracies in their ligand binding regions. It has been demonstrated that even moderate structural errors in the backbone and side chain coordinates interfere with traditional ligand docking approaches and cause a critical deterioration in the ability to accurately reproduce binding poses [32, 33].

On that account, the use of protein models as target receptors for ligand docking in structure-based drug development requires appropriate computational techniques that may be different from those designed to operate on the crystal structures. The recently developed

FINDSITE/Q-dock ligand homology modeling (LHM) methodology is one such approach that has been demonstrated to exhibit the desired tolerance to receptor structure deformation [38, 39]. Conceptually similar to protein comparative modeling, LHM extends template-based techniques to the modeling of protein-ligand interactions and provides a detailed functional annotation of the target proteins. As schematically depicted in Figure 2, following protein structural characterization, the functional characterization can be considered as a three-stage process. First, functional relationships between proteins are detected by sensitive methods such as sequence profile-driven threading [40, 41] in order to identify essential features associated with ligand binding, i.e. functionally important residues, common molecular substructures in binding ligands and the structural conservation of their binding modes [39]. These insights are subsequently exploited during the initial docking of ligands by a similarity-based approach [39, 42]. Finally, drug candidates placed into the target binding pockets are subject to a refinement procedure to optimize the interactions with the protein and to rank the predicted poses [38, 43]. To deal with the problem of structural deformations when protein models are used as the target structures, low-resolution ranking and scoring techniques have been developed [44–46].

In this study, we present the results of the large-scale structure modeling and virtual screening of the entire human kinome. All-atom structural models of all kinase domains in humans have been constructed by a state-of-the-art protein structure prediction approach [40, 41, 47, 48]. Next, ATP-binding pockets were identified and used as the target sites in ligand-based virtual screening against a large (>2×10^6) collection of commercially available drug-like compounds [49] followed by ligand docking/refinement applied to the top 1×10^4 molecules for each kinase. Ligand homology modeling [38, 39] produced >1×10^9 molecular fingerprint-based similarity assessments of drug-kinase pairs and >5×10^6 3D models of drug-kinase complexes. The latter were subsequently evaluated by various scoring functions and finally, the ranked lists of compounds were compiled for each human kinase. Modeling accuracy is validated for protein structure prediction, binding residues identification and ligand docking using available experimental data. Compound ranking is assessed in retrospective benchmarks against several commonly used ligand libraries, including BindingDB [50], MDL Drug Data Report [51] and the Directory of Useful Decoys [52]. Furthermore, in a case study, we discuss the possible application of machine learning on virtual screening data to support the development of isoform-specific protein kinase inhibitors.

The full set of modeled protein structures, docked ligand conformations and compound rankings are freely available to the academic community via a user-friendly web interface that can be accessed from http://cssb.biology.gatech.edu/kinomelhm/as well as from the ZINC website (http://zinc.docking.org/applications/2010Apr/Brylinski-2010.tar.gz).

2. MATERIALS AND METHODS

2. 1. Kinase structure modeling

The sequences of all kinase domains identified in the human genome were taken from [1]. This repository contains 516 putative protein kinase genes; 409 of which are grouped into 8 major kinase families (AGC, CAMK, CK1, CMGC, RGC, STE, TK and TKL), 82 are classified as “others” and 25 are considered atypical. Protein structure modeling was carried out as follows: First, for each kinase domain structure templates were selected from a non-redundant template library by our threading algorithm PROSPECTOR_3 [40, 41], which was designed to detect close as well as remote homologous templates. Subsequently, threading templates were submitted to TASSER [47, 48], a coarse-grained structure assembly/refinement procedure guided by tertiary restraints extracted from the template structures. All-atom models were constructed from Cα coordinates obtained from the TASSER simulations by

PULCHRA [53]. Finally, the kinase structures were energy minimized in the CHARMM22 force field [54] using the Jackal modeling package [55]. Modeled kinase structures were then taken as targets for the prediction of ATP-binding sites by FINDSITE [56, 57], a threading-based method that identifies ligand-binding sites based on binding site similarity among superimposed groups of functionally and structurally related template structures. The ATP-binding pockets were used as the target sites to dock ligands.

2.2. Ligand docking and ranking

The ligand docking procedure consisted of initial ligand placement by FINDSITE^{LHM} [39] followed by low-resolution refinement by Q-Dock^{LHM} [38] and all-atom refinement using AMMOS [58]. FINDSITE^{LHM} is a fast ligand homology modeling approach that docks flexible ligands by a simple superpositioning procedure. It uses a collection of template-bound ligands extracted from binding sites predicted by FINDSITE to derive the common molecule substructures, viz. the anchor functional groups. Subsequently, the consensus binding poses of the anchor substructures are used for target ligand superposition, where the flexibility of a ligand is accounted for by the superposition of multiple low-energy conformations generated by BALLOON [59]. The conformation that can be superimposed onto the reference coordinates with the lowest RMSD structure to the predicted anchor pose is selected as the final model. Initial binding poses generated by FINDSITE^{LHM} were submitted to low-resolution refinement by Q-Dock^{LHM}. Q-Dock^{LHM} is a direct extension of Q-Dock [44] that additionally includes harmonic RMSD restraints imposed on the predicted anchor-binding pose. The lowest-energy conformation generated during the Replica Exchange Monte Carlo sampling was selected as the final docking result. Ligand poses provided by Q-Dock^{LHM} were transformed into the all-atom representation and further refined by molecular mechanics optimization using AMMOS [58]. AMMOS uses the AMMP molecular simulation package [60] to carry out automatic refinement of the protein-ligand complexes. We used the sp4 force field in all simulations; protein atoms within a 12 Å sphere around the ligand were allowed to be flexible (AMMOS Case 4).

To provide compound ranking in virtual screening, we applied the following scoring functions: ligand-based molecular fingerprints implemented in FINDSITE [56, 61], anchor substructure coverage, where the anchor substructures were identified by FINDSITE^{LHM} [39], structure-based scoring by the total energy and the pocket-specific component from Q-Dock^{LHM}'s force field [38] and the total docked energy provided by AMMOS [58].

2.3. Datasets

2.3.1. ZINC—Each protein kinase was screened against 2,095,759 compounds from the ZINC7 library [49]. In the first step, a fast ligand-based screening was applied using molecular fingerprints provided by FINDSITE [56, 57], as described above. Subsequently, for each target, the top 10,000 compounds (0.5% of the library) were selected based on the modified Tanimoto score [39, 62, 63] and submitted to molecular docking by FINDSITE^{LHM} followed by Q-Dock^{LHM} and AMMOS. Finally, the compounds were re-ranked by the structure-based scoring functions.

2.3.2. PDB—Protein structure modeling, binding residue prediction and docking accuracy were assessed for 326 kinase crystal structures taken from [64]. The dataset consists of 57 different human kinases with a ligand bound in the ATP-binding site (278 unique protein-ligand pairs) and 48 ligand-free forms.

Kinase structure modeling accuracy was assessed by the global Cα RMSD and the TM-score [65]. Local structural distortions of the binding pockets were evaluated by their Cα and all-atom RMSD calculated over the binding residues identified by LPC [66]. The accuracy of

ATP-binding site detection by FINDSITE was expressed as the distance of the predicted site from the ligand geometric center in the crystal structures and the Matthew's correlation coefficient (MCC) calculated for the binding residues:

$$MCC=\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}} \quad \text{(Eq. 1)}$$

where *TP*, *TN*, *FP* and *FN* denote respectively: true positives (correctly predicted binding residues), true negatives (residues correctly predicted not to bind a ligand), false positives (overpredicted binding residues) and false negatives (missing binding residues).

To evaluate docking accuracy, we use the fraction of correctly predicted binding residues as well as the fraction of recovered native specific protein-ligand contacts [38]. In theoretical protein models, the local geometry of the binding pocket frequently deviates from the experimental structure. Therefore, ligand poses transferred from the crystal structures upon the superposition of the binding residues roughly estimate the upper bound for ligand docking accuracy against protein models. Ligands randomly placed into the ATP-binding pockets within a distance of 7 Å (docking sphere) from the predicted pocket center delineate the lower bound of docking accuracy.

2.3.3. BindingDB—Ranking accuracy in virtual screening was assessed for 362 known active compounds selected from BindingDB [50]. The top 10,000 compounds from virtual screening against the ZINC7 library were used as background compounds. For each known kinase inhibitor, we assess the improvement of ranking by structure-based scoring using Q-Dock^{LHM} and AMMOS over the fingerprint-based scoring by FINDSITE.

2.3.4. KEGG—The rank of ATP for each kinase target was calculated versus 12,158 background molecules from the KEGG compound library [67].

2.3.5. DUD—The Directory of Useful Decoys [52] was designed for benchmarking virtual screening approaches and contains 40 protein targets, 2,950 active compounds and 36 decoy molecules per one active compound with similar physical properties. Seven targets from DUD belong to the human kinase family: CDK2, EGFR, FGFR1, KDR, p38a, PDGFRb and SRC. Here, we use these targets to provide a comparative assessment of the screening protocols used in this study and in state-of-the-art virtual screening using DOCK [68]. The energy-based ligand rankings by DOCK3.5 applied to the crystal structures of the target kinases were taken from [52]. In addition, we carried out docking simulations using DOCK6 against the crystal as well as modeled kinase structures. Target receptor structures were prepared by Chimera [69] using the default set of parameters. Ligand preparation including the Gasteiger-Marsili partial charge assignment and the calculation of hydrogen positions were done using OpenBabel [70]. Binding poses generated by flexible ligand docking simulations using a default "anchor and grow" protocol were ranked by the total grid score. The results provided by DOCK3.5/6 were compared to ligand rankings obtained by low-resolution docking/scoring by Q-Dock^{LHM} [38, 44] (knowledge-based potential) and FINDSITE^{LHM} [39] (anchor coverage) using modeled structures. Furthermore, we applied data fusion to combine the results from virtual screening using the pocket-specific potential (Q-Dock^{LHM}) and the anchor coverage (FINDSITE^{LHM}). Here, we use the *SUM* rule that is expected to be less sensitive to noisy input than both extreme rules [71] and is preferred when fusion is by rank [72]. For a given library compound *k*, a combined score (*CS*) is calculated from:

$$CS^{k}=\sum_{i=1}^{n}r_{i} \quad \text{(Eq. 2)}$$

where n is the number of ranked lists (in our case, n=2: Q-Dock^{LHM} and FINDSITE^{LHM}) and r_i denotes the rank position of the library compound k in the i-th ranked list.

The performance of DOCK3.5/6 and Q-Dock^{LHM}/FINDSITE^{LHM} in virtual screening for kinase inhibitors is assessed by EF_{10} (enrichment factor calculated for the top 10% of the ranked screening library) [39, 73], BEDROC20 (Boltzmann-enhanced discrimination of ROC) [73], AUAC (area under the accumulation curve) [73] and ACT-50% (the top fraction of ranked library that contains 50% of the active compounds). Random ligand ranking yields EF_{10}, BEDROC20, AUAC and ACT-50% of 1.0, 0.1, 0.5 and 0.5, respectively.

2.3.6. MDDR—MDL Drug Data Report provides comprehensive information on bioactive compounds compiled from published and unpublished sources [51]. 562 protein kinase C (PKC) inhibitors were selected from MDDR (MDL activity index: 78374) and used in virtual screening against 9 isoenzymes of PKC: α, β, γ, delta;, ε, η, θ, ι and ζ. For each PKC isoform, 10,000 compounds randomly selected from the ZINC7 database [49] were used as the background library.

2.3.7. PKC—In addition to the assessment of the ligand ranking capability for protein kinase C, we also investigated the possibility of the prediction of inhibitor specificity toward different isoenzymes of PKC by a machine learning approach. Here, we use 10 inhibitors collected from the literature, for which half-maximal inhibition constants (IC_{50}) values toward PKC isoforms were determined experimentally: corallidictyal [74], GF-109203X [75], Gö-6976 [76], JTT-010 [77], K252a [78], midostaurin [79], rottlerin [80], ruboxistaurin [81], staurosporine [82] and UCN-01 [83]. A simple three-state classification model was constructed; for each PKC isoenzyme, the inhibitors were divided into three classes based on the IC_{50} values: class I, good binders (IC_{50} < 100 nM), class II, weak binders (100nM < IC_{50} < 1 μM) and class III, non-binders (IC_{50} > 1 μM). The Supporting Vector Machine (SVM, nu-SVC type with a polynomial kernel) [84] was trained on the following features: docking scores (raw score and the Z-score from virtual screening): fingerprint-based (FINDSITE), final docked energy (Q-Dock^{LHM}), pocket specific component (Q-Dock^{LHM}), and the chemo-physical properties of the inhibitors: molecular weight (MW), octanol/water partition coefficient (logP) and topological polar surface area (PSA). The molecular properties were calculated by OpenBabel [70]. The classification model was validated using the following leave-one-out procedure: in each round, one inhibitor was removed from the dataset, the SVM model was trained on the inhibition data for the remaining compounds and the excluded inhibitor was assigned a binding class for each PKC isoenzyme. The accuracy is assessed in terms of the fraction of correct assignments. Finally, the SVM model was trained on all experimental data and the prediction was made for PKC isoenzyme–inhibitor pairs for which no inhibition constants are reported in the literature.

3. RESULTS

3. 1. Modeled structures for the human kinome

Template-based modeling is one of the most frequently used techniques in protein structure prediction and has the capability of providing reliable models in the presence of evolutionarily related template structures [35, 36]. In this study, we constructed structure models for all kinase sequences identified in the human kinome by our protein structure prediction protocol: threading by PROSPECTOR_3 [40, 41] followed by structure assembly/

refinement using TASSER [47, 48]. Figure 3 presents the global Cα root-mean-square-deviation, RMSD, TM-score [65] and binding pocket RMSD from the crystal structure for the set of 57 ligand-bound and 48 ligand-free human kinases [64] that have experimentally determined structures in the PDB. The global structures of kinase domains have an average Cα RMSD (TM-score) from the holo and apo crystal structures of 2.75Å (0.92) and 3.13Å (0.90), respectively. The lower RMSD and higher TM-score values calculated for holo vs. apo structures reflect the fact that most of the template structures in the PDB are in the ligand-bound functional state (see Figure 1) and the force field used by TASSER for structure refinement favors conformations that are typically more compact and contain more inter residue contacts than the open conformational states. Figure 3C shows the local deviations from the experimental structure for ATP-binding pockets; the accuracy of these regions is critical for ligand docking and ranking. The average Cα (all-atom) RMSD calculated over the binding residues is 1.27 Å (2.36 Å). Despite progress in the prediction of residue rotamers [85–87], side chain modeling still needs further improvement. Nevertheless, these values concur with the estimated plasticity of the binding sites that have the capability to bind the same ligand (or class of ligands) in the kinase family [88] and proteins in general [39]. In contrast to many ligand-docking algorithms that require highly accurate experimental structures, the local distortions of ligand binding regions are tolerated to some extent by docking approaches that use a lower resolution description [38, 44–46].

3. 2. ATP-binding pocket prediction by FINDSITE

To dock ligands into the modeled kinase structures, we used binding pockets predicted by FINDSITE, a threading-based binding site prediction/protein functional inference/ligand screening algorithm that detects common ligand binding sites in a set of evolutionarily related proteins [56, 57]. The average number of binding sites predicted by FINDSITE for a kinase target is 32. Here, we use only the top-ranked pockets with the majority of low ranked sites likely involved in nonspecific ligand binding. The results of ATP-binding pocket prediction carried out for 57 different human kinases and 278 ligands are shown in Figure 4. Considering a cutoff distance of 4 Å as the hit criterion, the success rate for all complexes and for a non-redundant set with respect to the protein sequences is 86.7% and 94.7%, respectively. In most of the cases, the predicted distance is less than 2.5 Å. This very high accuracy of binding site prediction results in high Matthew's correlation coefficients (MCC) calculated for the binding residues; for most of the complexes, the MCC is >0.80 (Figure 4, inset). Two major factors account for the exceptional efficiency of ATP-binding site detection: the kinase structures have been modeled by TASSER to very high accuracy and most of the currently available kinase inhibitors, whose complexes are present in the PDB [34], target ATP-binding sites [64, 89].

3.3. Ligand binding pose prediction

Low-resolution docking techniques are frequently used to dock ligands into the distorted binding sites of the modeled receptor structures [38, 44–46]. In Figure 5, we assess the accuracy of ligand docking into the ATP-binding sites of modeled kinase structures for 278 unique protein-ligand pairs using FINDSITE^{LHM}, Q-Dock^{LHM} and an all atom refinement procedure, AMMOS [58]. The upper bound for docking accuracy is estimated by transferring ligands from the crystal structures into the modeled structures upon the local superposition of the binding residues. The fraction of correctly predicted binding residues (Figure 5A) is the highest for Q-Dock^{LHM} and is very close to the estimated upper bound. All-atom refinement by AMMOS recovers less binding residues, and is comparable in performance to FINDSITE^{LHM}. The fraction of correctly predicted specific protein-ligand contacts, (essential for effective ligand ranking), provides a more detailed assessment of the docking accuracy. Previous benchmark simulations demonstrated that ligand homology modeling by FINDSITE^{LHM} followed by an anchor-constrained low-resolution refinement by Q-

Dock^{LHM} outperforms other approaches in ligand binding pose prediction against modeled receptor structures [38]. Figure 5B shows that FINDSITE^{LHM} provides an approximately correct binding pose, which is subsequently improved by low-resolution refinement using Q-Dock^{LHM}. This procedure recovers significantly more specific protein-ligand contacts than all-atom refinement using AMMOS. It is noteworthy that all programs used for ligand docking perform significantly better than random ligand placement in terms of the recovered binding residues as well as the specific protein-ligand contacts.

The success of a refinement procedure depends on the quality of the force field used. The latter can be assessed by the correlation between the native-likeness, e.g. RMSD from the crystal ligand binding pose and the energy score, and the location of the energy minimum; the lowest energy pose should correspond to a conformation close to native. Here, for four representative examples, we evaluate the quality of the Q-Dock^{LHM}'s force field that impacts refinement outcome. In Figures 6A for cyclin-dependent kinase 2, CDK2, and in Figure 6B for proto-oncogene serine/threonine protein kinase, PIM1, we show that when the docking energy score is well correlated with RMSD and the energy minimum is located close to the ligand-binding pose in the crystal structure, not surprisingly, low-resolution refinement improves docking results; the fraction of specific contacts increases from 0.65 (using FINDSITE^{LHM}) to 0.70 (using Q-Dock^{LHM}) and from 0.45 to 0.60 respectively. On the other hand, in some cases, the energy score is not correlated with the native-likeness of the ligand poses; this results in minor (from 0.41 to 0.50 of the fraction of specific native contacts that are recovered for tyrosine kinase FGFR2, Figure 6C) or no improvement by Q-Dock^{LHM} over FINDSITE^{LHM} (0.40 for both methods for CDK2, Figure 6D). Nevertheless, significantly better ligand binding poses are generated by Q-Dock^{LHM} for most of the modeled complexes, which is critical for ligand ranking. As shown in Figure 5B, the fraction of complexes with 0.40, 0.50, 0.60 and 0.70 of the specific native contacts recovered by low-resolution, Q-Dock^{LHM}, refinement is 0.83, 0.72, 0.56 and 0.30, respectively.

We next consider some specific examples:

3.4. Staurosporine binding mode in modeled kinase structures

A natural product of *S. staurosporeus*, staurosporine (STU), was first described as an inhibitor of protein kinase C [82]. Later on, STU was demonstrated to have nanomolar potency toward a variety of other protein kinases [90, 91]. STU non-selectively inhibits protein kinases by competitively binding to the ATP-binding site. Highly conserved across the protein kinase family, the position of STU in the ATP-binding pocket (see Figure 7) is stabilized by predominantly hydrophobic interactions and hydrogen bonds [92, 93]. The inhibitor mimics several aspects of adenosine binding; the lactam ring of STU occupies a similar position to the amino group of ATP and the sugar moiety of STU binds to the region occupied by the ribose of ATP, pointing out of the binding site. Despite the structural distortions of ATP-binding sites in modeled kinase structures (see Figure 3C), similar binding modes of STU and ATP were recovered by the low-resolution docking using Q-Dock^{LHM}. This is shown in Figure 8 for nine protein kinases whose crystal structures are not available in the PDB [34]. High accuracy of STU docking into the ATP-binding sites of homology models has been reported previously for eight protein kinases [88]. Furthermore, it is noteworthy that structure-based virtual screening against protein models using the pocket-specific potential as a scoring function assigned very high Z-scores and corresponding ranks to both compounds (Figure 8). This high ranking efficiency is encouraging since staurosporine, as a potent and promiscuous kinase inhibitor, represents a prototypical ATP-competitive lead compound [94].

3.5. Ligand ranking

The goal of virtual screening is to rapidly assess a large library of compounds in order to identify those molecules that most likely bind to a drug target. To estimate the reliability of ligand ranking, known active molecules are typically included in the screening library; high ranks assigned to these compounds by a virtual screening approach indicate that the top fraction of the ranked library is significantly enriched in biologically active compounds. Here, we assess the accuracy of ligand- and structure-based virtual screening for a set of 362 known kinase inhibitors selected from the BindingDB [50]. We note that only compounds that are not present in the PDB [34] are used in this analysis. The results in terms of the ranks assigned to known active molecules in the screening library of the top 10,000 ranked compounds of the ZINC7 library are presented in Figure 9. First, we assess the improvement in ligand ranking of structure-based over ligand-based virtual screening. For most of the compounds, docking-based scores provide better (lower) ranks than the fingerprint-based scoring using FINDSITE, with the low-resolution scoring by Q-Dock^{LHM} providing the most effective ligand ranking. The number of compounds assigned with ranks <100 (the top 1% of the library) is 3, 68 and 2 for FINDSITE, Q-Dock^{LHM} and AMMOS, respectively. Q-Dock^{LHM} assigned ranks lower than 1,000 (the top 10% of the library) to almost twice as many known inhibitors as AMMOS and four times more inhibitors than FINDSITE. Separately, we assess the ranking of ATP that binds to all kinases (Figure 9, inset). For 95% of the protein kinases, ATP was ranked by Q-Dock^{LHM} within the top 1% of the screening library. Strong evolutionary relationships between protein kinases are easily detected by sequence profile-driven threading; this results in similar sets of templates identified for individual members. Hence, the ranks assigned to ATP by FINDSITE using the molecular fingerprints extracted from template-bound ligands are invariant across the kinase family. The improved ranking provided by Q-Dock^{LHM} over FINDSITE provides a very strong justification for the more CPU-expensive Q-Dock^{LHM}-based ligand docking. We note that the top 10,000 compounds selected by FINDSITE from the ZINC7 database [49] have been re-ranked by Q-Dock^{LHM} and AMMOS for all 516 kinases identified in the human proteome.

3. 6. Performance on the DUD dataset

The Directory of Useful Decoys (DUD) provides a large unbiased benchmark set to test the performance of virtual screening approaches [52]. In contrast to many other datasets, the decoy compounds included in DUD are physically similar to active compounds, yet they have a different topology from their active counterparts. This important feature helps avoid the artificial enrichment often seen in virtual screening studies [95]; hence DUD is frequently used in the assessment of the performance of virtual screening approaches [96–100]. In Table 1, we compare the performance of the ligand homology modeling approach (FINDSITE^{LHM}/Q-Dock^{LHM}) used in this study to DOCK3.5/6, the all-atom docking/screening tool on a set of 7 protein kinases from DUD. First, we note that for receptor crystal structures, DOCK6 provides higher enrichment with respect to the previous version, DOCK3.5. In benchmarks against modeled structures, considering single scoring functions, FINDSITE^{LHM} performs better on average than DOCK6, Q-Dock^{LHM} and AMMOS with an average EF_{10}, BEDROC20, AUAC and ACT-50% (the top fraction of ranked library that contains 50% of the active compounds) of 1.905, 0.133, 0.625 and 0.285, respectively. Moreover, the performance of FINDSITE^{LHM} for protein models is close to or depending on the metric used exceeds the performance of DOCK6 applied to the crystal structures, 1.955, 0.173, 0.383 and 0.779. The two docking algorithms, DOCK6 and Q-Dock^{LHM} perform quite comparably against modeled structures; DOCK6 outperforms Q-Dock^{LHM} with respect to EF_{10} and BEDROC20; however, the average AUAC and ACT-50% are notably better for Q-Dock^{LHM}. Poor AUAC and ACT-50% measures calculated for ligands ranked by DOCK6 suggest that active compounds are not equally well distributed across the screening library and low ranks are assigned to a significant fraction of known inhibitors. In addition, we find that high-resolution refinement and scoring using AMMOS applied to ligand poses generated by Q-Dock^{LHM} does not improve ligand ranking. The combined approach, data fusion using the *SUM* rule applied to ligand rankings from FINDSITE^{LHM} and Q-Dock^{LHM}, performs significantly better than the other approaches used in this study and yields an average EF_{10}, BEDROC20, AUAC and ACT-50% of 2.378, 0.162, 0.624 and 0.316, respectively. The most important conclusion emerging from this study is that ligand homology modeling by FINDSITE^{LHM}/Q-Dock^{LHM} using predicted protein structures is a competitive alternative to classical structure-based virtual screening with better or at least comparable efficacy in ligand ranking to approaches that require solved protein crystal structures with bound ligands.

3. 7. Virtual screening for isoform-specific PKC inhibitors

An early event in signal transduction pathways, the activation of the protein kinase C family (PKC), leads to many biological responses that regulate major cellular functions [101]. Different PKC isoenzymes are considered to be promising targets in the treatment of many diseases, including diabetes, multiple sclerosis, cardiovascular disease, cancer and Alzheimer's [5, 6, 8]. Based on their structure and regulation mechanisms, the isoforms of protein kinase C can be divided into three categories: conventional calcium-dependent PKCs (α, β_I, β_{II} and γ) that are activated by both phospholipids and diacylglycerol (DAG), novel PKCs (δ, ε, η and θ) that require phospholipids and DAG for activation but do not require Ca^{2+} and atypical PKCs (ι/λ and ζ) that are unresponsive to both activators [102, 103]. Most of the compounds inhibit PKC isoforms non-selectively; to exploit the distinct function of different PKC isoenzymes, isoenzyme-specific inhibitors are highly desired. Here, in a benchmark scenario, we demonstrate how virtual screening data can be used to support the development of isoform-specific PKC inhibitors.

In the first step, we carried out the retrospective evaluation of the virtual screening for the PKC inhibitors using 562 active compounds from the MDDR database [51] and 10,000 random decoys from the ZINC7 library [49]. We note that MDDR does not specify the selectivity of PKC inhibitors toward different isoenzymes. Therefore, the results in terms of the enrichment behavior plots are presented in Figure 10 for each isoform of the PKC. This example shows that the compound ranking using an all-atom scoring function such as the one used by AMMOS [58] is ineffective when modeled protein structures are used as the target receptors. It has been already demonstrated in more representative benchmarks that all-atom approaches for ligand docking and ranking are highly sensitive to structural distortions in ligand binding regions [38, 39, 44]. Molecular fingerprints provided by FINDSITE perform better that random ligand selection with 4.8% and 24.0% of the known inhibitors recovered in the top 1% and 10% of the screening library, respectively. Since PKC isoforms are closely related to each other, the ranks of library compounds by FINDSITE are identical for all isoenzymes; similar behavior was seen when FINDSITE is applied to the prediction of ATP binding (see Figure 9, inset), as FINDSITE emphasizes the conserved binding features across a protein family; here, we are interested in their differences. Quite similar performance is observed for structure-based virtual screening by the total energy reported by Q-Dock^{LHM} (which includes both generic and protein specific components, see Methods, below) Here, the percentage of active compounds recovered in the top 1% (10%) of the library varies from 2.8% (12.6%) for PKC-γ to 10.1% (27.6%) for PKC-ι. Undoubtedly, the best performance is obtained using the pocket-specific component of the Q-Dock^{LHM}'s force field as a scoring function to rank ligands. The fraction of known PKC inhibitors ranked within the top 1% and 10% of the library varies from 11.7% (PKC-α) to 13.9% (PKC-ι) and from 34.9% (PKC-α) to 42.3% (PKC-ε), respectively. Furthermore, using the pocket specific scoring function, ligand ranking is very stable across different isoforms of the PKC.

Next, we employed a simple machine learning model to demonstrate that virtual screening data can be used for the prediction of the inhibitor specificity toward different PKC isoenzymes. Leave-one-out cross validation (Table 2, in italics) shows that for 7 out of 10 inhibitors (GF-109203X, Gö-6976, K252a, midostaurin, rottlerin, staurosporine and UCN-01) the three-state binding assignment of good, weak and non-binders (see Materials and Methods) was better than random (random accuracy is 33.3%). The highest benchmark accuracy (60%) is observed for the indolocarbazole Gö-6976, which is the first discovered PKC inhibitor that was shown in vivo to discriminate between Ca^{2+}-dependent and Ca^{2+}-independent PKC isoenzymes [76]. In the validation of our model, Gö-6976 is predicted to inhibit α and β isoforms with high affinity <100 nM (experimental IC_{50} values are 2.3 nM and 6.2 nM, respectively). PKC isoenzymes d and e are false positives i.e. predicted to be inhibited, while the experimental data shows no inhibition. Gö-6976 is correctly assigned as a non-active compound against the isoform ζ. The activity of three other Ca^{2+}-independent PKC isoenzymes, η, θ and ι, is also predicted to be unaffected by Gö-6976; this is in good agreement with its class-selective inhibition profile. Another interesting example is rottlerin that was predicted as a weak/non-inhibitor for most PKC isoforms. In the recent study of protein kinases and inhibitors, rottlerin failed to show any PKC inhibitory activity against the α and delta; PKC isotypes [104, 105], which is consistent with our results. Considering the relatively high prediction accuracy, we used all experimental data to predict IC_{50} values for PKC isoenzyme–inhibitor pairs for which no inhibition constants are reported in the literature (Table 3, in bold).

Finally, we apply the SVM model to assign the selectivity toward PKC isoenzymes to 562 known inhibitors from MDDR. Since no information on the selectivity profile is provided by MDDR, we indirectly validate the results using the Google search engine. The results are shown in Figure 11. Most of the compounds were predicted by the SVM to inhibit the conventional PKC isoforms with an IC_{50} <100 nM, whereas relatively few inhibitors were predicted to be atypical PKC specific (Figures 11A and B). This trend is in good qualitative agreement with the number of hits reported by Google (Figure 11C). The highest number of hits was obtained using “protein kinase C alpha inhibitors” as the query phrase. Significantly fewer hits are reported for the novel and particularly for the atypical PKC isoenzymes. This simple study on the isoform selectivity of PKC inhibitors demonstrates that virtual screening using protein models can provide useful information for the development of biopharmaceuticals with desired specificity. Despite showing a classification accuracy that is better than random, there is still the possibility of further improvements. However, these would require an alternate approach that focuses on the variability across homologues rather than on their conserved features.

3. 8. Simulation times

Computational procedures were carried out on IBM cluster with 2.0GHz AMD Opteron processors and deploying Linux OS. Figure 12 shows docking times for the programs used in this study. FINDSITE^{LHM} is the least CPU-expensive procedure with an average docking time of less than 2 min per compound. Q-Dock^{LHM} requires ~8 min to dock a ligand on average. High-resolution refinement by AMMOS typically uses less than 5 min of CPU time.

4. DISCUSSION

The increasing interest in kinase inhibitors as novel therapeutics has created a demand for the structural characterization of the human kinase family. Targeting the entire family rather than individual members gives better prospects for developing compounds with improved selectivity [106, 107] or, in some cases, inhibitors that are “selectively unselective” i.e. modulate activity of multiple kinase targets associated with the selfsame pathological process [88, 108]. Despite progress in protein crystallography and structural genomics efforts that doubled the rate of experimental structure determination [109], the structural coverage of the kinase family remains poor and unequally distributed [110]. Propitiously, the presence of a sufficient number of template structures in the PDB [34] and the high structural conservation of kinase domains make the members of the kinome family perfect targets for template-based structure modeling. A wide range of highly accurate protein models would not only contribute directly to the structure-based drug design [111], but also to the initial experimental structure determination of new kinases by molecular replacement techniques [112].

In this study, we constructed reliable three-dimensional models for all kinase sequences identified in the human proteome for use in structure-based drug design. Structure modeling was followed by a detailed functional characterization, starting from the identification of ATP-binding pockets that are the primary target sites for most of the currently available kinase inhibitors [64, 89, 113]. Highly accurate protein models and the availability of ligand-bound template structures resulted in precisely annotated binding residues, which constitute a practical dataset to guide further mutational studies. Next, for each kinase family member, we applied fast fingerprint-based virtual screening to rank a collection of $>2\times10^6$ compounds from the ZINC database [49]. By selecting the top 10,000 molecules for each kinase, a kinase-focused library of ~30,000 unique compounds was compiled. This collection, representing reasonable chemical coverage of kinase inhibitor space, should improve the efficiency of drug development. In high throughput screens, large combinatorial libraries are frequently supplemented with the target-oriented libraries [114, 115]. Recent screening experiments on 41 kinases demonstrated that the overall hit enrichment is significantly higher for a target class focused library compared to generic drug-like compounds [116]. Our kinase-focused, 30,000-compound library compiled from the top virtual screening hits may be of practical use for the selection of compounds for high-throughput screens by providing scaffolds with high kinase inhibitory potential.

Docking benchmarks carried out for modeled kinase structures demonstrate that ligand homology modeling often produces approximately correct binding poses, which recover most of the native protein-ligand contacts. These results, nota bene non-trivial, since the distorted binding sites in protein models represent a considerable challenge for many ligand-docking algorithms, are in good agreement with our previous studies [38, 39]. We note that over five million distinct models of three-dimensional protein-drug complexes have been constructed; these can be used for rapid binding affinity assessment by any structure-based scoring function.

Our retrospective virtual screening analyses validate the modeled kinase structures as valuable targets in structure-based drug development. Here, we applied a hierarchical virtual screening approach. First, a large collection of compounds was assessed by a fast fingerprint-based approach. Subsequently, the top-ranked fraction of the screening library was submitted to more CPU-expensive ligand homology modeling followed by low-resolution docking/refinement. In the end, lead candidates were re-ranked using structure-based scoring functions. Such a workflow is very common in modern virtual screening protocols that typically consist of a cascade of different filter approaches [117]. The least computationally expensive ligand-based techniques applied at the outset of in silico screening allow for a rapid assessment of large compound libraries, with the top fraction of the ranked library enriched with active compounds [39, 56, 100]. These pre-filtered subsets are subject to structure-based virtual screening by flexible ligand docking. Predicted binding modes in the target receptor pockets are re-ranked according to the energy of binding estimated from molecular interactions. Finally, the top fraction of the library, typically containing hundreds to thousands molecules, is submitted for experimental validation. Following a protocol of consecutive hierarchical filters, lead candidates that show IC_{50}

values in the micro to nanomolar range have been successfully identified for, e.g., the human aldose reductase [118] and the human carbonic anhydrase [119]. Our approach to virtual screening that combines ligand homology modeling and low-resolution docking can be applied to theoretically modeled receptor structures and yields accuracy at least comparable to structure-based virtual screening against high quality X-ray structures using state-of-the-art docking algorithms.

5. CONCLUSIONS

Considering the accelerated pace of genome sequencing and the much slower rate of experimental protein structure determination, it is unlikely that three-dimensional structures will be soon available for all potential drug targets. Therefore, modern drug development at the proteome level must rely on modeled structures provided by state-of-the-art protein structure prediction techniques. In this study, we show that hierarchical virtual screening combining fast fingerprint-based filtering with structure-based ligand homology modeling emerges as a powerful compound prioritization technique applicable to the early stages of proteome-scale drug design projects. By applying this approach to all kinase domains in humans, we have provided the scientific community with a very extensive structural and functional characterization of the human kinome to support the discovery of novel kinase inhibitors.

Acknowledgments

We thank Drs. Brian K. Shoichet and John J. Irwin (UCSF) for making the dataset available from the ZINC website. This work was supported in part by grant No. GM-48835 and GM-37408 of the Division of General Medical Sciences of the National Institutes of Health.

References

1. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002; 298(5600):1912–34. [PubMed: 12471243]
2. Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J. 1995; 9(8):576–96. [PubMed: 7768349]
3. Kennelly PJ. Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. FEMS Microbiol Lett. 2002; 206(1):1–8. [PubMed: 11786249]
4. Adcock IM, Chung KF, Caramori G, Ito K. Kinase inhibitors and airway inflammation. Eur J Pharmacol. 2006; 533(1–3):118–32. [PubMed: 16469308]
5. Basu A. The potential of protein kinase C as a target for anticancer treatment. Pharmacol Ther. 1993; 59(3):257–80. [PubMed: 8309991]
6. Bradshaw D, Hill CH, Nixon JS, Wilkinson SE. Therapeutic potential of protein kinase C inhibitors. Agents Actions. 1993; 38(1–2):135–47. [PubMed: 8480534]
7. Leclerc S, Garnier M, Hoessel R, Marko D, Bibb JA, Snyder GL, Greengard P, Biernat J, Wu YZ, Mandelkow EM, Eisenbrand G, Meijer L. Indirubins inhibit glycogen synthase kinase-3 beta and CDK5/p25, two protein kinases involved in abnormal tau phosphorylation in Alzheimer's disease. A property common to most cyclin-dependent kinase inhibitors? J Biol Chem. 2001; 276(1):251–60. [PubMed: 11013232]
8. Sasase T. PKC - a target for treating diabetic complications. Drugs of the Future. 2006; 31(6):503–11.
9. Weinmann H, Metternich R. Drug discovery process for kinase inhibitors. Chembiochem. 2005; 6(3):455–9. [PubMed: 15742380]
10. Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, Fanning S, Zimmermann J, Lydon NB. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nat Med. 1996; 2(5):561–6. [PubMed: 8616716]

11. Barker AJ, Gibson KH, Grundy W, Godfrey AA, Barlow JJ, Healy MP, Woodburn JR, Ashton SE, Curry BJ, Scarlett L, Henthorn L, Richards L. Studies leading to the identification of ZD1839 (IRESSA): an orally active, selective epidermal growth factor receptor tyrosine kinase inhibitor targeted to the treatment of cancer. Bioorg Med Chem Lett. 2001; 11(14):1911–4. [PubMed: 11459659]
12. Burris HA 3rd. Dual kinase inhibition in the treatment of breast cancer: initial experience with the EGFR/ErbB-2 inhibitor lapatinib. Oncologist. 2004; 9(Suppl 3):10–5. [PubMed: 15163842]
13. Sun L, Liang C, Shirazian S, Zhou Y, Miller T, Cui J, Fukuda JY, Chu JY, Nematalla A, Wang X, Chen H, Sistla A, Luu TC, Tang F, Wei J, Tang C. Discovery of 5-[5-fluoro-2-oxo-1,2-dihydroindol-(3Z)-ylidenemethyl]-2,4- dimethyl-1H-pyrrole-3-carboxylic acid (2-diethylaminoethyl)amide, a novel tyrosine kinase inhibitor targeting vascular endothelial and platelet-derived growth factor receptor tyrosine kinase. J Med Chem. 2003; 46(7):1116–9. [PubMed: 12646019]
14. Noble ME, Endicott JA, Johnson LN. Protein kinase inhibitors: insights into drug design from structure. Science. 2004; 303(5665):1800–5. [PubMed: 15031492]
15. Terstappen GC, Reggiani A. In silico research in drug discovery. Trends Pharmacol Sci. 2001; 22(1):23–6. [PubMed: 11165668]
16. Jain AN. Virtual screening in lead discovery and optimization. Curr Opin Drug Discov Devel. 2004; 7(4):396–403.
17. Zoete V, Grosdidier A, Michielin O. Docking, virtual high throughput screening and in silico fragment-based drug design. J Cell Mol Med. 2009; 13(2):238–48. [PubMed: 19183238]
18. McInnes C. Virtual screening strategies in drug discovery. Curr Opin Chem Biol. 2007; 11(5):494–502. [PubMed: 17936059]
19. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004; 3(11):935–49. [PubMed: 15520816]
20. Abagyan RA, Totrov MM, Kuznetsov DN. ICM - a new method for protein modelling and design. Applications to docking and structure prediction from the distorted native conformation. J Comput Chem. 1994; 15(5):488–506.
21. Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput-Aided Mol Des. 2001; 15(5):411–28. [PubMed: 11394736]
22. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem. 1998; 19(14):1639–1662.
23. Chen H, Lyne PD, Giordanetto F, Lovell T, Li J. On evaluating molecular-docking methods for pose prediction and enrichment factors. J Chem Inf Model. 2006; 46(1):401–15. [PubMed: 16426074]
24. Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP. Comparison of automated docking programs as virtual screening tools. J Med Chem. 2005; 48(4):962–76. [PubMed: 15715466]
25. Kroemer RT. Structure-based drug design: docking and scoring. Curr Protein Pept Sci. 2007; 8(4): 312–28. [PubMed: 17696866]
26. Okamoto M, Takayama K, Shimizu T, Ishida K, Takahashi O, Furuya T. Identification of death-associated protein kinases inhibitors using structure-based virtual screening. J Med Chem. 2009; 52(22):7323–7. [PubMed: 19877644]
27. Medina-Franco JL, Giulianotti MA, Yu Y, Shen L, Yao L, Singh N. Discovery of a novel protein kinase B inhibitor by structure-based virtual screening. Bioorg Med Chem Lett. 2009; 19(16): 4634–8. [PubMed: 19604696]
28. Kiss R, Polgar T, Kirabo A, Sayyah J, Figueroa NC, List AF, Sokol L, Zuckerman KS, Gali M, Bisht KS, Sayeski PP, Keseru GM. Identification of a novel inhibitor of JAK2 tyrosine kinase by structure-based virtual screening. Bioorg Med Chem Lett. 2009; 19(13):3598–601. [PubMed: 19447617]

29. Peach ML, Tan N, Choyke SJ, Giubellino A, Athauda G, Burke TR Jr, Nicklaus MC, Bottaro DP. Directed discovery of agents targeting the Met tyrosine kinase domain by virtual screening. J Med Chem. 2009; 52(4):943–51. [PubMed: 19199650]

30. Coumar MS, Leou JS, Shukla P, Wu JS, Dixit AK, Lin WH, Chang CY, Lien TW, Tan UK, Chen CH, Hsu JT, Chao YS, Wu SY, Hsieh HP. Structure-based drug design of novel Aurora kinase A inhibitors: structural basis for potency and specificity. J Med Chem. 2009; 52(4):1050–62. [PubMed: 19140666]

31. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL 3rd. Assessing scoring functions for protein-ligand interactions. J Med Chem. 2004; 47(12):3032–47. [PubMed: 15163185]

32. Kim R, Skolnick J. Assessment of programs for ligand binding affinity prediction. J Comput Chem. 2008; 29(8):1316–31. [PubMed: 18172838]

33. McGovern SL, Shoichet BK. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. J Med Chem. 2003; 46(14):2895–907. [PubMed: 12825931]

34. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000; 28(1):235–42. [PubMed: 10592235]

35. Cozzetto D, Kryshtafovych A, Fidelis K, Moult J, Rost B, Tramontano A. Evaluation of template-based models in CASP8 with standard measures. Proteins. 2009; 77(Suppl 9):18–28. [PubMed: 19731382]

36. Ginalski K. Comparative modeling for protein structure prediction. Curr Opin Struct Biol. 2006; 16(2):172–7. [PubMed: 16510277]

37. Moult J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. Curr Opin Struct Biol. 2005; 15(3):285–9. [PubMed: 15939584]

38. Brylinski M, Skolnick J. Q-Dock(LHM): Low-resolution refinement for ligand comparative modeling. J Comput Chem. 2009

39. Brylinski M, Skolnick J. FINDSITE(LHM): a threading-based approach to ligand homology modeling. PLoS Comput Biol. 2009; 5(6):e1000405. [PubMed: 19503616]

40. Skolnick J, Kihara D. Defrosting the frozen approximation: PROSPECTOR--a new approach to threading. Proteins. 2001; 42(3):319–31. [PubMed: 11151004]

41. Skolnick J, Kihara D, Zhang Y. Development and large scale benchmark testing of the PROSPECTOR_3 threading algorithm. Proteins. 2004; 56(3):502–18. [PubMed: 15229883]

42. Marialke J, Korner R, Tietze S, Apostolakis J. Graph-based molecular alignment (GMA). J Chem Inf Model. 2007; 47(2):591–601. [PubMed: 17381175]

43. Marialke J, Tietze S, Apostolakis J. Similarity based docking. J Chem Inf Model. 2008; 48(1):186–96. [PubMed: 18044949]

44. Brylinski M, Skolnick J. Q-Dock: Low-resolution flexible ligand docking with pocket-specific threading restraints. J Comput Chem. 2008; 29(10):1574–1588. [PubMed: 18293308]

45. Vakser IA. Low-resolution docking: prediction of complexes for underdetermined structures. Biopolymers. 1996; 39(3):455–64. [PubMed: 8756522]

46. Wojciechowski M, Skolnick J. Docking of small ligands to low-resolution and theoretically predicted receptor structures. J Comput Chem. 2002; 23(1):189–97. [PubMed: 11913386]

47. Zhang Y, Skolnick J. Automated structure prediction of weakly homologous proteins on a genomic scale. Proc Natl Acad Sci U S A. 2004; 101(20):7594–9. [PubMed: 15126668]

48. Zhang Y, Skolnick J. Tertiary structure predictions on a comprehensive benchmark of medium to large size proteins. Biophys J. 2004; 87(4):2647–55. [PubMed: 15454459]

49. Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005; 45(1):177–82. [PubMed: 15667143]

50. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res. 2007; 35(Database issue):D198–201. [PubMed: 17145705]

51. MDL Drug Data Report. Prous Science. 2007. http://www.mdl.com/

52. Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. J Med Chem. 2006; 49(23):6789–801. [PubMed: 17154509]

53. Rotkiewicz P, Skolnick J. Fast procedure for reconstruction of full-atom protein models from reduced representations. J Comput Chem. 2008; 29(9):1460–5. [PubMed: 18196502]

54. MacKerell AD, Bashford D, Bellott, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B. 1998; 102(18):3586–3616.

55. Xiang Z, Honig B. Extending the accuracy limits of prediction for side-chain conformations. J Mol Biol. 2001; 311(2):421–30. [PubMed: 11478870]

56. Brylinski M, Skolnick J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. Proc Natl Acad Sci U S A. 2008; 105(1):129–34. [PubMed: 18165317]

57. Skolnick J, Brylinski M. FINDSITE: a combined evolution/structure-based approach to protein function prediction. Brief Bioinform. 2009; 10(4):378–91. [PubMed: 19324930]

58. Pencheva T, Lagorce D, Pajeva I, Villoutreix BO, Miteva MA. AMMOS: Automated Molecular Mechanics Optimization tool for in silico Screening. BMC Bioinformatics. 2008; 9:438. [PubMed: 18925937]

59. Vainio MJ, Johnson MS. Generating conformer ensembles using a multiobjective genetic algorithm. J Chem Inf Model. 2007; 47(6):2462–74. [PubMed: 17892278]

60. Harrison RW. Stiffness and Energy Conservation in Molecular Dynamics: an Improved Integrator. J Comput Chem. 1993; 14(9):11122–1122.

61. Brylinski M, Skolnick J. Comparison of structure-based and threading-based approaches to protein functional annotation. Proteins. 2009

62. Tanimoto, TT. An elementary mathematical theory of classification and prediction. 1958.

63. Xue L, Godden JW, Stahura FL, Bajorath J. Profile scaling increases the similarity search performance of molecular fingerprints containing numerical descriptors and structural keys. J Chem Inf Comput Sci. 2003; 43(4):1218–25. [PubMed: 12870914]

64. Kinnings SL, Jackson RM. Binding site similarity analysis for the functional classification of the protein kinase family. J Chem Inf Model. 2009; 49(2):318–29. [PubMed: 19434833]

65. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins. 2004; 57(4):702–10. [PubMed: 15476259]

66. Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M. Automated analysis of interatomic contacts in proteins. Bioinformatics. 1999; 15(4):327–32. [PubMed: 10320401]

67. Goto S, Okuno Y, Hattori M, Nishioka T, Kanehisa M. LIGAND: database of chemical compounds and reactions in biological pathways. Nucleic Acids Res. 2002; 30(1):402–4. [PubMed: 11752349]

68. Lorber DM, Shoichet BK. Hierarchical docking of databases of multiple ligand conformations. Curr Top Med Chem. 2005; 5(8):739–49. [PubMed: 16101414]

69. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25(13):1605–12. [PubMed: 15264254]

70. Guha R, Howard MT, Hutchison GR, Murray-Rust P, Rzepa H, Steinbeck C, Wegner J, Willighagen EL. The Blue Obelisk-interoperability in chemical informatics. J Chem Inf Model. 2006; 46(3):991–8. [PubMed: 16711717]

71. Ginn CMR, Willett P, Bradshaw J. Combination of molecular similarity measures using data fusion. Perspect Drug Discov Design. 2000; 20:1–16.

72. Hert J, Willett P, Wilton DJ, Acklin P, Azzaoui K, Jacoby E, Schuffenhauer A. Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures. J Chem Inf Comput Sci. 2004; 44(3):1177–85. [PubMed: 15154787]

73. Truchon JF, Bayly CI. Evaluating virtual screening methods: good and bad metrics for the “early recognition” problem. J Chem Inf Model. 2007; 47(2):488–508. [PubMed: 17288412]

74. Chan JA, Freyer AJ, Carte BK, Hemling ME, Hofmann GA, Mattern MR, Mentzer MA, Westley JW. Protein kinase C inhibitors: novel spirosesquiterpene aldehydes from a marine sponge Aka (= Siphonodictyon) coralliphagum. J Nat Prod. 1994; 57(11):1543–8. [PubMed: 7853003]

75. Toullec D, Pianetti P, Coste H, Bellevergue P, Grand-Perret T, Ajakane M, Baudet V, Boissin P, Boursier E, Loriolle F, et al. The bisindolylmaleimide GF 109203X is a potent and selective inhibitor of protein kinase C. J Biol Chem. 1991; 266(24):15771–81. [PubMed: 1874734]

76. Martiny-Baron G, Kazanietz MG, Mischak H, Blumberg PM, Kochs G, Hug H, Marme D, Schachtele C. Selective inhibition of protein kinase C isozymes by the indolocarbazole Go 6976. J Biol Chem. 1993; 268(13):9194–7. [PubMed: 8486620]

77. Sasase T, Yamada H, Sakoda K, Imagawa N, Abe T, Ito M, Sagawa S, Tanaka M, Matsushita M. Novel protein kinase C-beta isoform selective inhibitor JTT-010 ameliorates both hyper- and hypoalgesia in streptozotocin- induced diabetic rats. Diabetes Obes Metab. 2005; 7(5):586–94. [PubMed: 16050952]

78. Geiges D, Meyer T, Marte B, Vanek M, Weissgerber G, Stabel S, Pfeilschifter J, Fabbro D, Huwiler A. Activation of protein kinase C subtypes alpha, gamma, delta, epsilon, zeta, and eta by tumor-promoting and nontumor-promoting agents. Biochem Pharmacol. 1997; 53(6):865–75. [PubMed: 9113106]

79. Marte BM, Meyer T, Stabel S, Standke GJ, Jaken S, Fabbro D, Hynes NE. Protein kinase C and mammary cell differentiation: involvement of protein kinase C alpha in the induction of beta-casein expression. Cell Growth Differ. 1994; 5(3):239–47. [PubMed: 8018556]

80. Gschwendt M, Muller HJ, Kielbassa K, Zang R, Kittstein W, Rincke G, Marks F. Rottlerin, a novel protein kinase inhibitor. Biochem Biophys Res Commun. 1994; 199(1):93–8. [PubMed: 8123051]

81. Jirousek MR, Gillig JR, Gonzalez CM, Heath WF, McDonald JH 3rd, Neel DA, Rito CJ, Singh U, Stramm LE, Melikian-Badalian A, Baevsky M, Ballas LM, Hall SE, Winneroski LL, Faul MM. (S)-13-[(dimethylamino)methyl]-10,11,14,15-tetrahydro-4,9:16, 21-dimetheno-1H, 13H-dibenzo[e, k]pyrrolo[3,4-h][1,4,13]oxadiazacyclohexadecene-1,3(2H)-d ione (LY333531) and related analogues: isozyme selective inhibitors of protein kinase C beta. J Med Chem. 1996; 39(14):2664–71. [PubMed: 8709095]

82. Tamaoki T, Nomoto H, Takahashi I, Kato Y, Morimoto M, Tomita F. Staurosporine, a potent inhibitor of phospholipid/Ca++dependent protein kinase. Biochem Biophys Res Commun. 1986; 135(2):397–402. [PubMed: 3457562]

83. Seynaeve CM, Kazanietz MG, Blumberg PM, Sausville EA, Worland PJ. Differential inhibition of protein kinase C isozymes by UCN-01, a staurosporine analogue. Mol Pharmacol. 1994; 45(6): 1207–14. [PubMed: 8022414]

84. Chang, C-C.; Lin, C-J. LIBSVM: a library for support vector machines. 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

85. Krivov GG, Shapovalov MV, Dunbrack RL Jr. Improved prediction of protein side-chain conformations with SCWRL4. Proteins. 2009; 77(4):778–95. [PubMed: 19603484]

86. Liang S, Grishin NV. Side-chain modeling with an optimized scoring function. Protein Sci. 2002; 11(2):322–31. [PubMed: 11790842]

87. Xiang Z, Steinbach PJ, Jacobson MP, Friesner RA, Honig B. Prediction of side-chain conformations on protein surfaces. Proteins. 2007; 66(4):814–23. [PubMed: 17206724]

88. Rockey WM, Elcock AH. Structure selection for protein kinase docking and virtual screening: homology models or crystal structures? Curr Protein Pept Sci. 2006; 7(5):437–57. [PubMed: 17073695]

89. Liao JJ. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. J Med Chem. 2007; 50(3):409–24. [PubMed: 17266192]

90. Fabian MA, Biggs WH 3rd, Treiber DK, Atteridge CE, Azimioara MD, Benedetti MG, Carter TA, Ciceri P, Edeen PT, Floyd M, Ford JM, Galvin M, Gerlach JL, Grotzfeld RM, Herrgard S, Insko DE, Insko MA, Lai AG, Lelias JM, Mehta SA, Milanov ZV, Velasco AM, Wodicka LM, Patel HK, Zarrinkar PP, Lockhart DJ. A small molecule-kinase interaction map for clinical kinase inhibitors. Nat Biotechnol. 2005; 23(3):329–36. [PubMed: 15711537]

91. Karaman MW, Herrgard S, Treiber DK, Gallant P, Atteridge CE, Campbell BT, Chan KW, Ciceri P, Davis MI, Edeen PT, Faraoni R, Floyd M, Hunt JP, Lockhart DJ, Milanov ZV, Morrison MJ, Pallares G, Patel HK, Pritchard S, Wodicka LM, Zarrinkar PP. A quantitative analysis of kinase inhibitor selectivity. Nat Biotechnol. 2008; 26(1):127–32. [PubMed: 18183025]

92. Lawrie AM, Noble ME, Tunnah P, Brown NR, Johnson LN, Endicott JA. Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with CDK2. Nat Struct Biol. 1997; 4(10):796–801. [PubMed: 9334743]

93. Prade L, Engh RA, Girod A, Kinzel V, Huber R, Bossemeyer D. Staurosporine-induced conformational changes of cAMP-dependent protein kinase catalytic subunit explain inhibitory potential. Structure. 1997; 5(12):1627–37. [PubMed: 9438863]

94. Gescher A. Analogs of staurosporine: potential anticancer drugs? Gen Pharmacol. 1998; 31(5): 721–8. [PubMed: 9809468]

95. Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW, Taylor RD, Watson P. Virtual screening using protein-ligand docking: avoiding artificial enrichment. J Chem Inf Comput Sci. 2004; 44(3):793–806. [PubMed: 15154744]

96. Cross JB, Thompson DC, Rai BK, Baber JC, Fan KY, Hu Y, Humblet C. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. J Chem Inf Model. 2009; 49(6):1455–74. [PubMed: 19476350]

97. Dror O, Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. Novel approach for efficient pharmacophore-based virtual screening: method and applications. J Chem Inf Model. 2009; 49(10):2333–43. [PubMed: 19803502]

98. Fan H, Irwin JJ, Webb BM, Klebe G, Shoichet BK, Sali A. Molecular docking screens using comparative models of proteins. J Chem Inf Model. 2009; 49(11):2512–27. [PubMed: 19845314]

99. Tawa GJ, Baber JC, Humblet C. Computation of 3D queries for ROCS based virtual screens. J Comput-Aided Mol Des. 2009

100. von Korff M, Freyss J, Sander T. Comparison of ligand- and structure-based virtual screening on the DUD data set. J Chem Inf Model. 2009; 49(2):209–31. [PubMed: 19434824]

101. Nishizuka Y. Studies and prospectives of the protein kinase c family for cellular regulation. Cancer. 1989; 63(10):1892–903. [PubMed: 2539241]

102. Hofmann J. The potential for isoenzyme-selective modulation of protein kinase C. FASEB J. 1997; 11(8):649–69. [PubMed: 9240967]

103. Mellor H, Parker PJ. The extended protein kinase C superfamily. Biochem J. 1998; 332(Pt 2): 281–92. [PubMed: 9601053]

104. Davies SP, Reddy H, Caivano M, Cohen P. Specificity and mechanism of action of some commonly used protein kinase inhibitors. Biochem J. 2000; 351(Pt 1):95–105. [PubMed: 10998351]

105. Soltoff SP. Rottlerin: an inappropriate and ineffective inhibitor of PKCdelta. Trends Pharmacol Sci. 2007; 28(9):453–8. [PubMed: 17692392]

106. Diller DJ, Li R. Kinases, homology models, and high throughput docking. J Med Chem. 2003; 46(22):4638–47. [PubMed: 14561083]

107. Goldstein DM, Gray NS, Zarrinkar PP. High-throughput kinase profiling as a platform for drug discovery. Nat Rev Drug Discov. 2008; 7(5):391–7. [PubMed: 18404149]

108. Wilhelm SM, Carter C, Tang L, Wilkie D, McNabola A, Rong H, Chen C, Zhang X, Vincent P, McHugh M, Cao Y, Shujath J, Gawlak S, Eveleigh D, Rowley B, Liu L, Adnane L, Lynch M, Auclair D, Taylor I, Gedrich R, Voznesensky A, Riedl B, Post LE, Bollag G, Trail PA. BAY 43–9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. Cancer Res. 2004; 64(19):7099–109. [PubMed: 15466206]

109. Grabowski M, Joachimiak A, Otwinowski Z, Minor W. Structural genomics: keeping up with expanding knowledge of the protein universe. Curr Opin Struct Biol. 2007; 17(3):347–53. [PubMed: 17587562]

110. Marsden BD, Knapp S. Doing more than just the structure-structural genomics in kinase drug discovery. Curr Opin Chem Biol. 2008; 12(1):40–5. [PubMed: 18267130]

111. Stout TJ, Foster PG, Matthews DJ. High-throughput structural biology in drug discovery: protein kinases. Curr Pharm Des. 2004; 10(10):1069–82. [PubMed: 15078142]

112. Argos P, Ford GC, Rossmann MG. An application of the molecular replacement technique in direct space to a known protein structure. Acta Crystallogr. 1975; A31:499–506.

113. Cohen P. The development and therapeutic potential of protein kinase inhibitors. Curr Opin Chem Biol. 1999; 3(4):459–65. [PubMed: 10419844]

114. Schnur DM. Recent trends in library design: 'rational design' revisited. Curr Opin Drug Discov Devel. 2008; 11(3):375–80.

115. Sun D, Chuaqui C, Deng Z, Bowes S, Chin D, Singh J, Cullen P, Hankins G, Lee WC, Donnelly J, Friedman J, Josiah S. A kinase-focused compound collection: compilation and screening strategy. Chem Biol Drug Des. 2006; 67(6):385–94. [PubMed: 16882313]

116. Gozalbes R, Simon L, Froloff N, Sartori E, Monteils C, Baudelle R. Development and experimental validation of a docking strategy for the generation of kinase-targeted libraries. J Med Chem. 2008; 51(11):3124–32. [PubMed: 18479119]

117. Muegge I, Enyedy IJ. Virtual screening for kinase targets. Curr Med Chem. 2004; 11(6):693–707. [PubMed: 15032724]

118. Kraemer O, Hazemann I, Podjarny AD, Klebe G. Virtual screening for inhibitors of human aldose reductase. Proteins. 2004; 55(4):814–23. [PubMed: 15146480]

119. Gruneberg S, Stubbs MT, Klebe G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. J Med Chem. 2002; 45(17):3588–602. [PubMed: 12166932]

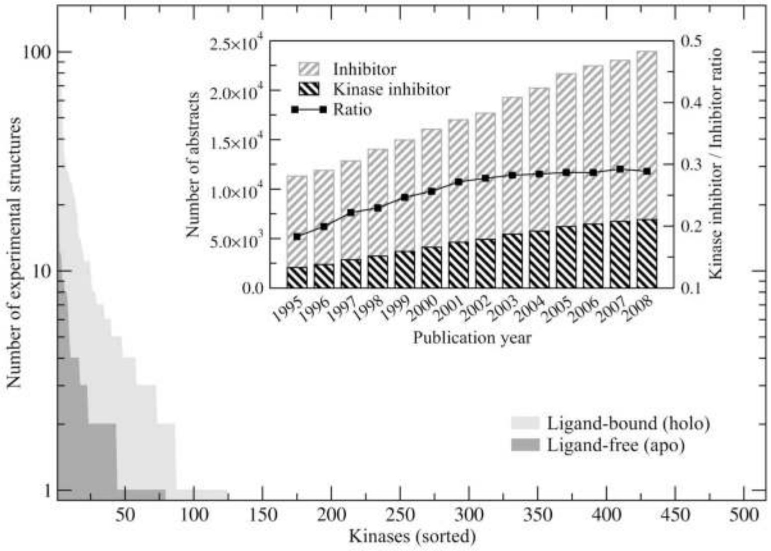

Figure 1.

Availability of the ligand-bound and ligand-free crystal structures for the human kinome. Inset: Histogram of the number of abstracts published since 1995 selected from the PubMed using following queries: (“inhibitor”[Text Word]) AND (“*YEAR*/01/01”[Publication Date]: “*YEAR*/12/31”[Publication Date]) and ((“inhibitor”[Text Word]) AND (“kinase”[Text Word])) AND (“*YEAR*/01/01”[Publication Date]: “*YEAR*/12/31”[Publication Date]).

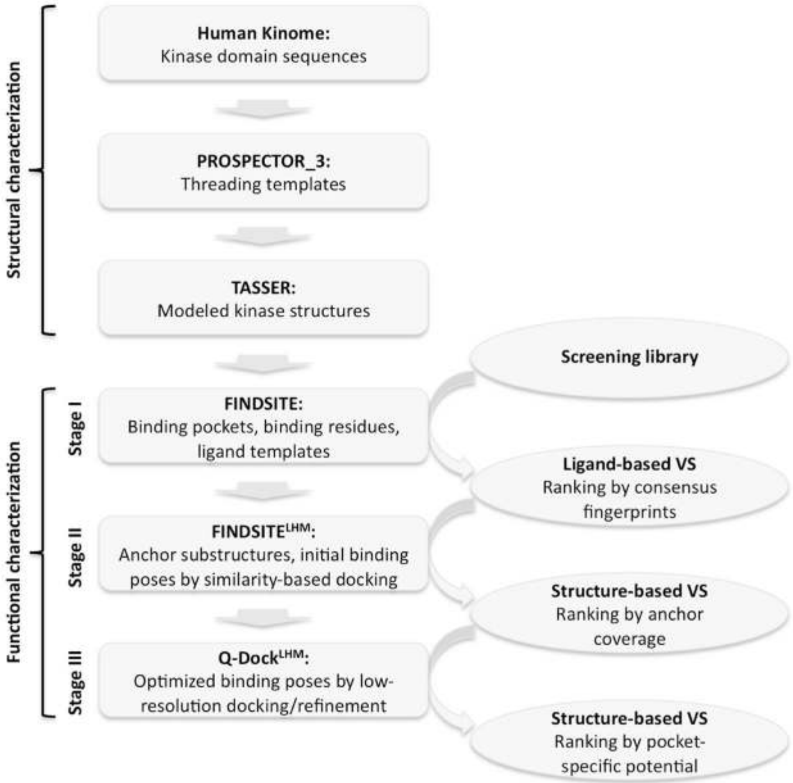
Structural characterization
Human Kinome:
Kinase domain sequences
PROSPECTOR_3:
Threading templates
TASSER:
Modeled kinase structures
Functional characterization
Stage I
FINDSITE:
Binding pockets, binding residues, ligand templates
Stage II
FINDSITE^LHM:
Anchor substructures, initial binding poses by similarity-based docking
Stage III
Q-Dock^LHM:
Optimized binding poses by low-resolution docking/refinement
Screening library
Ligand-based VS
Ranking by consensus fingerprints
Structure-based VS
Ranking by anchor coverage
Structure-based VS
Ranking by pocket-specific potential

Figure 2.

Hierarchical approach to structural and functional characterization of proteins using homology modeling techniques.

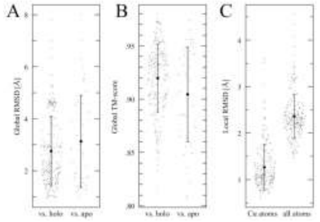

Figure 3.

Accuracy of kinase structure modeling using TASSER. Global Cα RMSD (A) and TM-score (B) are calculated versus ligand-bound (holo) and ligand-free (apo) structural forms of the target proteins. Local Ca and all-atom RMSD calculated over the binding residues are shown in C.

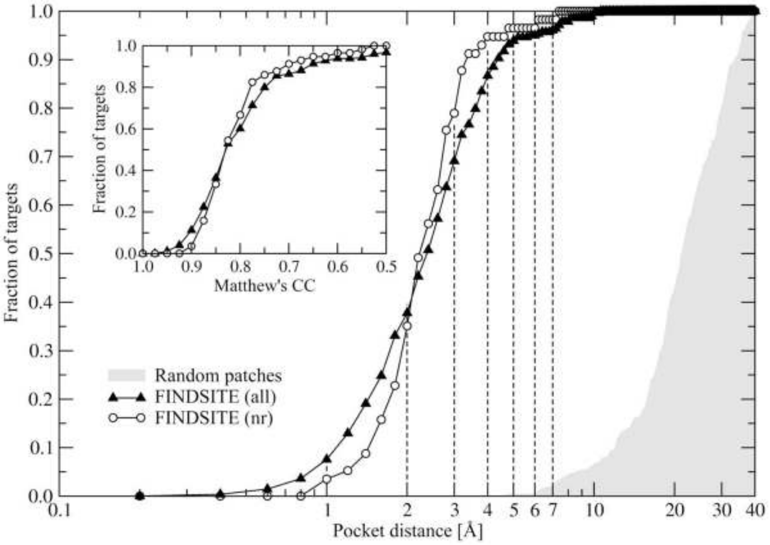

Figure 4.

ATP-binding pocket detection by FINDSITE. The results are presented as the cumulative fraction of kinase targets with a distance between the center of mass of an inhibitor in the crystal complex and the center of the predicted binding sites, less than or equal to the distance displayed on the *x* axis. Open circles show the results for a non-redundant (nr) dataset with respect to the target proteins. Gray area corresponds to randomly selected patches on the protein surface. Inset: Matthew's correlation coefficient calculated for the predicted binding residues.

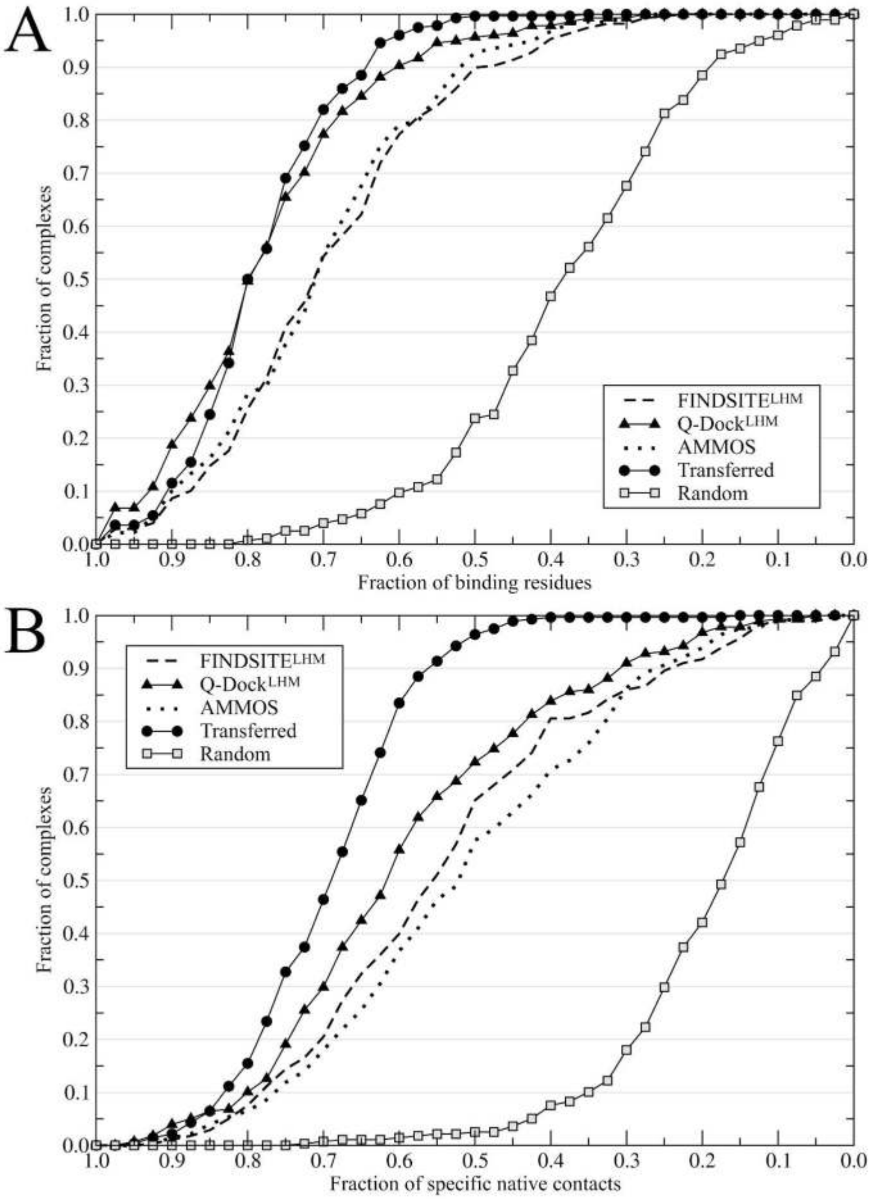

Figure 5.

Docking accuracy of the ligand homology modeling approach applied to the human kinome. Fraction of binding residues (A) and specific protein-ligand contacts (B) predicted by FINDSITE^{LHM}, Q-Dock^{LHM} and AMMOS is compared to the ligand poses directly transferred from the crystal structures as well as to ligands randomly placed into the binding pockets.

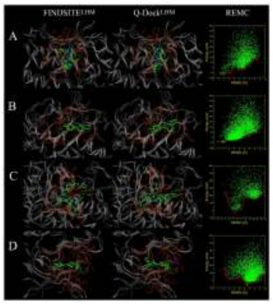

Figure 6.

Low-resolution docking/refinement by ligand homology modeling using protein models as the target receptors. A – CDK2, 1oiq; B – PIM1, 1yxx; C – FGFR2, 1oec and D – CDK2, 2btr. Left, middle: Inhibitor binding poses predicted by FINDSITE^{LHM} and Q-Dock^{LHM} (solid sticks, colored by atom type) are compared to the crystal structures (transparent sticks). Protein models (binding residues colored in red) are superposed onto the crystal structures of the target kinases (binding residues colored in orange). Right: correlation of the Q-Dock energy score and RMSD from the crystal binding pose for the ligand conformations sampled using Replica Exchange Monte Carlo (REMC). The red line highlights low-energy conformations for the broad range of RMSD values.

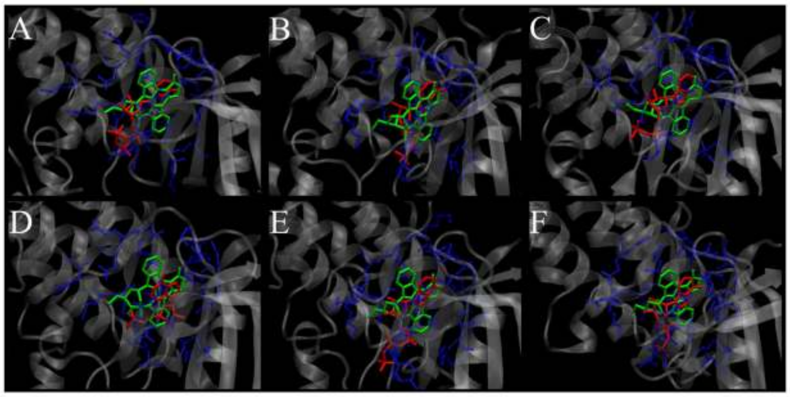

Figure 7.
Crystal structures of several protein kinases complexed with staurosporine (STU) and ATP. A – CDK2 (STU: 1aq1, ATP: 1b38), B – GSK3B (STU: 1q3d, ADP: 1j1c), C – LCK (STU: 1qpd, ANP: 1qpc), D – PIM1 (STU: 1yhs, AMP: 1yxu), E – PDK1 (STU; 1oky, ATP: 1h1w), F – MAPKAPK2 (STU: 1nxk, ADP: 1ny3). STU, the set ATP/ADP/AMP/ANP and selected binding residues are colored in green, red and blue, respectively.

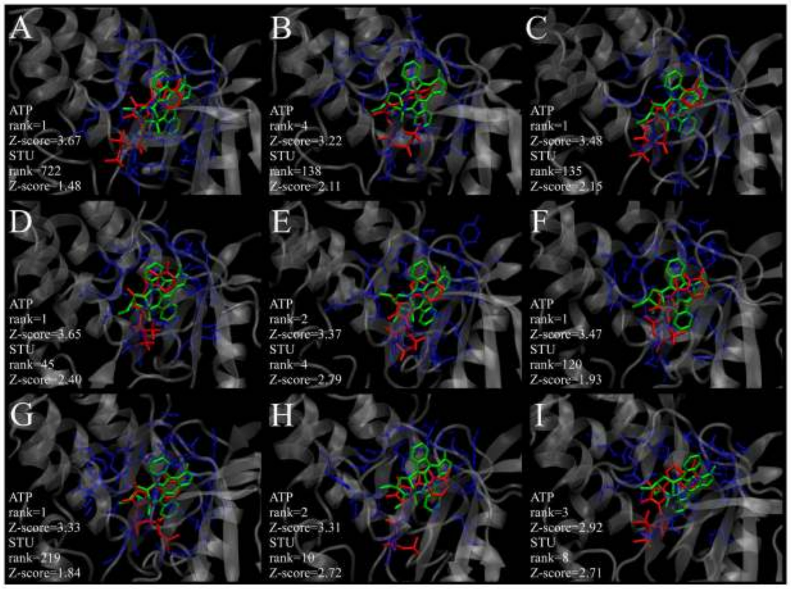

Figure 8.

Modeled structures of protein kinases bound to staurosporine (STU) and ATP. A – CDC2, B – Erk1, C – FGR, D – LYN, E – PKACa, F – PKCa, G – PKCg, H – PKG1, I – smMLCK. STU, ATP and selected binding residues are colored in green, red and blue, respectively. ATP and STU ranks and Z-scores from virtual screening using Q-Dock^{LHM} against modeled kinase structures are given.

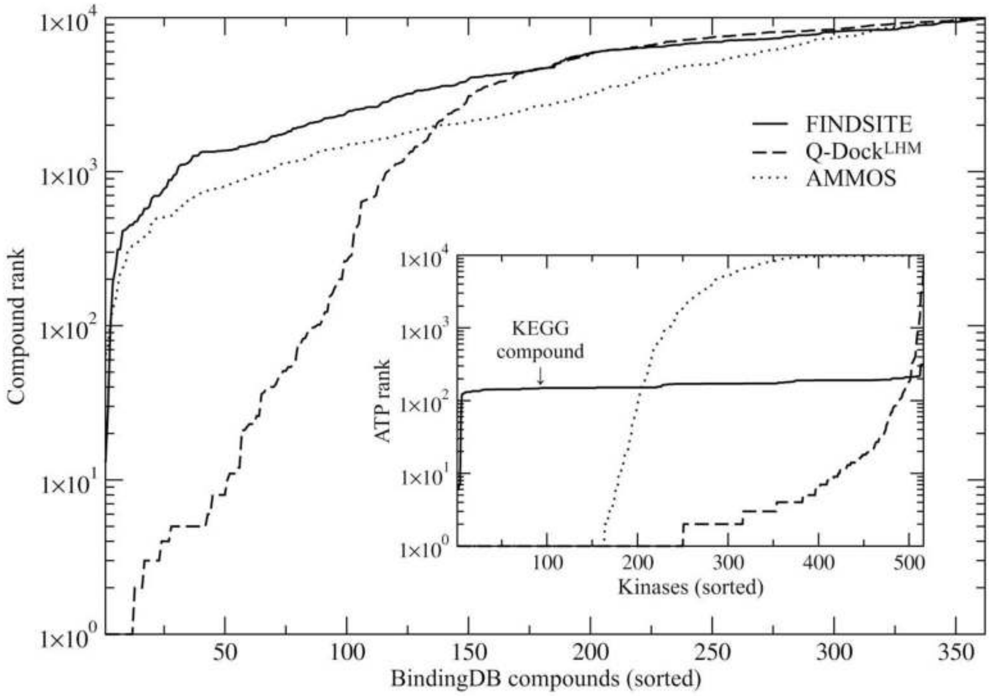

Figure 9.

Performance of virtual screening on the BindingDB dataset. Active compounds are sorted by increasing rank reported by FINDSITE fingerprints (ligand-based screening), Q-Dock^{LHM} (structure-based screening, low-resolution) and AMMOS (structure-based screening, high-resolution). Inset: ATP ranks for all protein kinases; for FINDSITE, the ranks in the KEGG compound library are used.

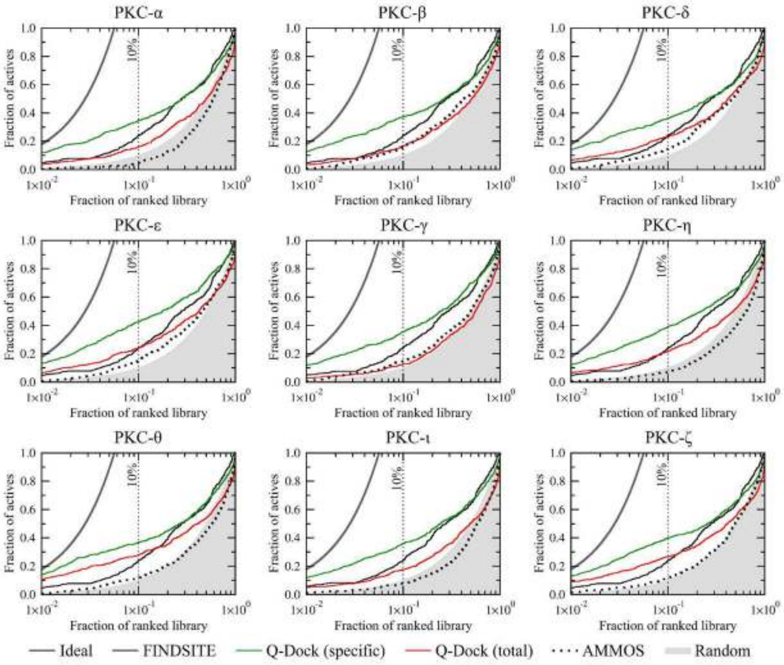

Figure 10.

Virtual screening for protein kinase C inhibitors. The enrichment behavior for FINDSITE (molecular fingerprints), Q-Dock^{LHM} (total energy score and the pocket-specific component) and AMMOS (all-atom scoring) is compared to a random ligand selection for different isoenzymes of PKC.

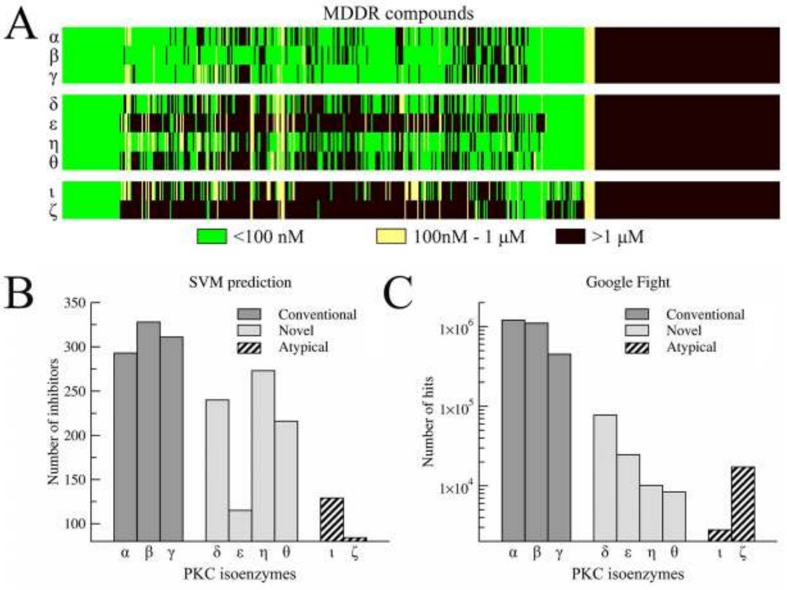

Figure 11.

Prediction of the PKC isoenzyme selectivity of known PKC inhibitors from MDDR. A – three-state binding assignment of good (IC_{50} <100 nM), weak (100nM < IC_{50} < 1 μM) and non-binders (IC_{50} >1 μM) by machine learning. B – number of MDDR compounds predicted to inhibit different PKC isoforms with IC_{50} <100 nM, C – number of hits returned by the Google search engine (http://www.googlefight.com/) using different PKC isoenzyme inhibitors as the query phrases.

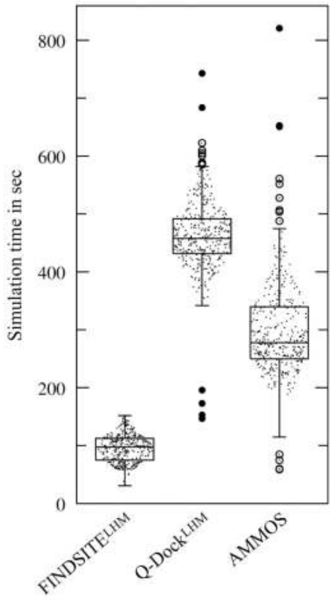

Figure 12.

Docking times for FINDSITE^{LHM}, Q-Dock^{LHM} and AMMOS. Boxes end at the quartiles Q_1 and Q_3; a horizontal line in a box is the median. “Whiskers” point at the farthest points that are within 3/2 times the interquartile range. Outliers and suspected outliers are presented as solid and blank circles, respectively.

Table 1

Performance of ligand homology modeling on seven protein kinases from the DUD dataset compared to the results obtained using DOCK. Ranking capability is assessed by the enrichment factor (EF_{10}), Boltzmann-enhanced discrimination of ROC (BEDROC20), the area under the accumulation curve (AUAC) and the top fraction of ranked library that contains 50% of the active compounds (ACT-50%).

		CDK2	EGFR	FGFR1	KDR	p38a	PDGFRb	SRC	Average ±SD
DOCK3.5 *crystal structures*	BEDROC20	0.189	0.200	0.003	0.085	0.115	0.009	0.026	0.090 ±0.082
	EF_{10}	2.200	2.545	0.085	1.081	1.992	0.197	0.323	1.203 ±1.038
	AUAC	0.549	0.565	0.201	0.402	0.532	0.323	0.448	0.431 ±0.134
	ACT-50%	0.340	0.274	0.885	0.682	0.463	0.742	0.494	0.554 ±0.223
DOCK6 *crystal structures*	BEDROC20	0.250	0.236	0.107	0.198	0.106	0.161	0.150	0.173 ±0.058
	EF_{10}	2.600	2.590	1.453	1.892	1.289	1.987	1.871	1.955 ±0.504
	AUAC	0.459	0.441	0.346	0.393	0.314	0.355	0.371	0.383 ±0.052
	ACT-50%	0.586	0.717	0.847	0.760	0.868	0.861	0.812	0.779 ±0.101
DOCK6 *protein models*	BEDROC20	0.104	0.255	0.003	0.119	0.070	0.306	0.090	0.135 ±0.107
	EF_{10}	1.000	2.568	0.085	1.351	0.898	2.930	1.161	1.428 ±0.992
	AUAC	0.341	0.433	0.196	0.399	0.236	0.404	0.290	0.328 ±0.091
	ACT-50%	0.781	0.725	0.863	0.711	0.923	0.832	0.911	0.821 ±0.085
AMMOS *protein models*	BEDROC20	0.049	0.014	0.058	0.038	0.069	0.072	0.033	0.048 ±0.021
	EF_{10}	1.400	0.158	0.932	0.405	1.055	1.338	0.581	0.838 ±0.472
	AUAC	0.671	0.466	0.611	0.422	0.510	0.475	0.518	0.525 ±0.087
	ACT-50%	0.299	0.537	0.350	0.590	0.501	0.527	0.427	0.462 ±0.107
Q-DOCK^{LHM} *protein models*	BEDROC20	0.163	0.062	0.105	0.076	0.088	0.099	0.020	0.088 ±0.044
	EF_{10}	2.400	0.968	1.610	1.081	1.289	1.210	0.194	1.250 ±0.668
	AUAC	0.665	0.553	0.613	0.533	0.526	0.577	0.466	0.562 ±0.064
	ACT-50%	0.225	0.418	0.364	0.423	0.477	0.438	0.529	0.411 ±0.097
FINDSITE$^{\mathbf{LHM}}$ *protein models*	BEDROC20	0.155	0.067	0.175	0.146	0.125	0.113	0.151	0.133 ±0.035
	EF_{10}	2.000	1.014	2.712	2.027	1.797	1.656	2.129	1.905 ±0.515
	AUAC	0.690	0.525	0.686	0.563	0.595	0.618	0.700	0.625 ±0.069
	ACT-50%	0.204	0.442	0.186	0.372	0.249	0.317	0.228	0.285 ±0.095

		CDK2	EGFR	FGFR1	KDR	p38a	PDGFRb	SRC	Average ±SD
Data fusion *protein models*	BEDROC20	0.321	0.107	0.210	0.114	0.127	0.161	0.096	0.162 ±0.080
	EF_{10}	4.400	1.689	2.966	1.486	1.875	2.229	2.000	2.378 ±1.010
	AUAC	0.724	0.552	0.698	0.565	0.584	0.627	0.619	0.624 ±0.066
	ACT-50%	0.155	0.420	0.184	0.420	0.391	0.314	0.328	0.316 ±0.109

Table 2

Benchmarking results for the prediction of the inhibitor selectivity toward protein kinase C isoenzymes. Experimental and benchmark values of IC_{50} are shown in regular font and italics, respectively. Correct and incorrect classifications are highlighted in green and red, respectively.

Inhibitor	**IC_{50} values for PKC isoenzymes**								
	α (50%)	**β (75%)**	**γ (71%)**	**δ (11%)**	**ε (30%)**	**η (80%)**	**θ**	**ι**	**ζ (20%)**
corallidictyal (0%)	🟥 30 µM *100nM – 1 µM*				🟥 89 µM *100nM – 1 µM*	🟥 >300 µM *100nM – 1 µM*			🟥 >300 µM *100nM – 1 µM*
GF-109203X (40%)	🟩 8.4 nM *<100 nM*	🟩 18 nM *<100 nM*		🟥 210 nM *<100 nM*	🟥 132 nM *<100 nM*				🟥 5.8 µM *<100 nM*
Gö-6976 (60%)	🟩 2.3 nM *<100 nM*	🟩 6.2 nM *<100 nM*		🟥 No inh *100nM – 1 µM*	🟥 No inh *<100 nM*				🟩 No inh *>1 µM*
JTT-010 (33%)	🟥 86 nM *100nM – 1 µM*	🟥 4 nM *100nM – 1 µM*	🟩 110 nM *100nM – 1 µM*	🟥 54 nM *100nM – 1 µM*	🟩 490 nM *100nM – 1 µM*				🟥 1.7 µM *100nM – 1 µM*
K252a (50%)	🟥 40 nM *100nM – 1 µM*		🟩 400 nM *100nM – 1 µM*	🟩 925 nM *100nM – 1 µM*	🟥 4.5 µM *100nM – 1 µM*	🟩 490 nM *100nM – 1 µM*			🟥 4.2 µM *100nM – 1 µM*
midostaurin (57%)	🟩 24 nM *<100 nM*	🟩 17 nM *<100 nM*	🟩 18 nM *<100 nM*	🟥 360 nM *<100 nM*	🟥 4.5 µM *<100 nM*	🟩 60 nM *<100 nM*			🟥 >10 µM *<100 nM*
rottlerin (57%)	🟩 30 µM *>1 µM*	🟥 42 µM *100nM – 1 µM*	🟩 40 µM *>1 µM*	🟥 6 µM *100nM – 1 µM*	🟩 100 µM *>1 µM*	🟩 82 µM *>1 µM*			🟥 100 µM *100nM – 1 µM*
ruboxistaurin (28%)	🟥 360 nM *<100 nM*	🟩 4.7 nM *<100 nM*	🟥 300 nM *<100 nM*	🟥 250 nM *<100 nM*	🟥 600 nM *<100 nM*	🟩 52 nM *<100 nM*			🟥 >10 µM *<100 nM*
staurosporine (50%)	🟥 8.7 nM *100nM – 1 µM*	🟩 11 nM *<100 nM*	🟩 11 nM *100nM – 1 µM*	🟥 4.3 nM *100nM – 1 µM*	🟩 7.4 nM *<100 nM*				🟥 1.7 µM *>1 µM*
UCN-01 (50%)	🟩 29 nM *<100 nM*	🟩 34 nM *<100 nM*	🟩 30 nM *<100 nM*	🟥 590 nM *<100 nM*	🟥 530 nM *<100 nM*				🟥 No inh *<100 nM*

Table 3

Prediction of the inhibitor selectivity toward protein kinase C isoenzymes by machine learning on virtual screening data. Experimental and predicted values of IC_{50} are shown in regular font and bold, respectively.

Inhibitor	IC_{50} values for PKC isoenzymes								
	α	β	γ	δ	ε	η	θ	ι	ζ
corallidictyal	30 μM	**>1 μM**	**>1 μM**	**>1 μM**	89 μM	>300 μM	**>1 μM**	**>1 μM**	>300 μM
GF-109203X	8.4 nM	18 nM	**<100 nM**	210 nM	132 nM	**<100 nM**	**<100 nM**	**100nM – 1 μM**	5.8 μM
Gö-6976	2.3 nM	6.2 nM	**>1 μM**	No inh	No inh	**>1 μM**	**>1 μM**	**>1 μM**	No inh
JTT-010	86 nM	4 nM	110 nM	54 nM	490 nM	**100nM –1 μM**	**<100 nM**	**>1 μM**	1.7 μM
K252a	40 nM	**<100 nM**	400 nM	925 nM	4.5 μM	490 nM	**100nM – 1 μM**	**>1 μM**	4.2 μM
midostaurin	24 nM	17 nM	18 nM	360 nM	4.5 μM	60 nM	**<100 nM**	**>1 μM**	>10 μM
rottlerin	30 μM	42 μM	40 μM	6 μM	100 μM	82 μM	**>1 μM**	**100nM – 1 μM**	100 μM
ruboxistaurin	360 nM	4.7 nM	300 nM	250 nM	600 nM	52 nM	**100nM – 1 μM**	**100nM – 1 μM**	>10 μM
staurosporine	8.7 nM	11 nM	11 nM	4.3 nM	7.4 nM	**100nM – 1 μM**	**<100 nM**	**<100 nM**	1.7 μM
UCN-01	29 nM	34 nM	30 nM	590 nM	530 nM	**<100 nM**	**<100 nM**	**>1 μM**	No inh