Published in final edited form as:

J Med Chem. 2016 May 12; 59(9): 4326–4341. doi:10.1021/acs.jmedchem.5b02041.

Delineation of Polypharmacology across the Human Structural Kinome Using a Functional Site Interaction Fingerprint Approach

Zheng Zhao[†], Li Xie[‡], Lei Xie^{*,§,||}, and Philip E. Bourne^{*,⊥}

[†]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, United States

[‡]Scripps Ranch, San Diego, California 92131, United States

[§]Department of Computer Science, Hunter College, The City University of New York, New York, New York 10065, United States

^{||}The Graduate Center, The City University of New York, New York, New York 10016, United States

[⊥]Office of the Director, National Institutes of Health, Bethesda, Maryland 20892, United States

Abstract

Targeted polypharmacology of kinases has emerged as a promising strategy to design efficient and safe therapeutics. Here, we perform a systematic study of kinase–ligand binding modes for the human structural kinome at scale (208 kinases, 1777 unique ligands, and their complexes) by integrating chemical genomics and structural genomics data and by introducing a functional site interaction fingerprint (Fs-IFP) method. New insights into kinase–ligand binding modes were obtained. We establish relationships between the features of binding modes, the ligands, and the binding pockets, respectively. We also drive the intrinsic binding specificity and which correlation with amino acid conservation. Third, we explore the landscape of the binding modes and highlight the regions of “selectivity pocket” and “selectivity entrance”. Finally, we demonstrate that Fs-IFP similarity is directly correlated to the experimentally determined profile. These improve our understanding of kinase–ligand interactions and contribute to the design of novel polypharmacological therapies targeting kinases.

*Corresponding Authors: L.X.: phone, +1-212-396-6550; lei.xie@hunter.cuny.edu. *P.E.B.: phone, +1-301-402-9818; philip.bourne@nih.gov.

Supporting Information: The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmed-chem.5b02041.

Notes: The authors declare no competing financial interest.

Introduction

Protein kinases are enzymes that phosphorylate other proteins (typical) and indeed other molecules (atypical) and hence play a vital role in signal transduction and cell differentiation.^{1,2} Clinical evidence from anticancer studies have suggested that many cancers involve mutations harbored in the kinase leading to signal disruption.^{3,4} The same is true of a variety of other disease states, and hence the design of safe and efficient kinase inhibitors is an extensive area of pharmaceutical research. Since the first kinase-targeted drug, imatinib, was approved in 2001,⁵ 30 kinase inhibitors have been approved (through Nov 2015) by the U.S. Food and Drug Administration (FDA).^{6–8} Notwithstanding this success, developing a kinase-targeted inhibitor with a desired selectivity profile across the human kinome remains a daunting task, as all kinases share a common catalytic domain and folding scaffold that binds ATP.^{4,9} Thus, it is vital to uncover the basic principles that govern the kinase–ligand binding specificity across the kinome so that we can rationally design safe and efficient kinase-targeted (poly)-pharmacological therapeutics.¹⁰

With the availability of an increasing number of diverse kinase–ligand complexes structures released by the Protein Data Bank (PDB),¹¹ many efforts have been devoted to establishing subtle binding modes by utilizing solved kinase–ligand complex structures. Kinase–ligand interaction patterns have been analyzed by directly comparing the binding similarity and the structural–activity relationship of kinase ligands.^{12–18} These studies have facilitated rational drug design and uncovered the inhibition mechanisms for certain drug molecules.^{19–27} However, many studies are only limited to the study of the binding mode of one kinase or a family of closely related kinases when a specific kinase is the focused target for drug design and discovery. A typical example of such structure-based drug design is crizotinib²⁰ for advanced non-small-cell lung cancer (NSCLC) with *ALK* rearrangement. Such work circumvents binding-pocket comparison among distantly related kinases and hence ignores selectivity across the kinome. Recently, several studies have analyzed the binding modes on a whole kinome scale using available X-ray structures of the human kinome.^{12,13,19,27,28} McGreg¹³ has used a set of 220 X-ray kinase–ligand structures to training a common pharmacophore for the ATP binding sit, and showed that common similar local structures exist for diverse kinases. Similarly, Kinnings et al.¹² have used a geometric hashing method to undertake the structural comparison of ATP binding sites and validated drug cross-reactivity. Huang et al.¹⁹ have provided a network analysis scheme of kinase selectivity

potential at the ATP binding site by mapping the sequences of 518 human kinases onto the structural alignment of 116 kinases of known three-dimensional structures and established a systematic guideline to develop ATP-competitive inhibitors. However, global sequence alignment of kinases may not reflect their relationships at the ligand binding site.⁶ Moreover, an increasing body of evidence suggests that the binding pocket of human kinases is not just in the region of the ATP binding site. A hydrophobic pocket adjacent to the ATP binding site and an allosteric pocket distal to the ATP binding site can also be used to achieve drug selectivity.^{17,29} Thus, the detailed analysis of all regions of the binding pocket may shed new light on the binding selectivity of kinases. Recently Bryan et al.²⁸ established that the structural features of the kinase binding site are correlated with bound inhibitors by using a combinatorial clustering of kinase residue position subsets method. The residues of the binding site are defined using a type II inhibitor imatinib-bound kinase complex as a template. Then all other kinases are mapped to the template structure using a Pfam kinase family multiple alignment.³⁰ van Linden et al.²⁷ have presented another systematic analysis of kinase–ligand interactions in all regions of the catalytic cleft of 1252 human kinase–ligand cocrystal structures that comprise only 85 kinases. The binding pocket of each human kinase was divided into three regions: front cleft, gate cleft, and back cleft. The ligand binding modes have been identified for each region. This divide-and-conquer analysis improves our understanding of kinase binding modes for several families of kinases. In spite of these advances, a whole kinome comparison and analysis of diverse kinases in all ligand-bound regions is needed to comprehensively study kinase inhibitor selectivity across the human kinome if kinase-targeted polypharmacology is to be successful.

Until June 2015, 2383 kinase complex structures belonging to 208 distinct human kinases had been released by the PDB. The fast-growing numbers of 3D complex structures combined with abundant kinase bioassay data not only provide us with new opportunities to study the binding modes across the human structural kinome but also present new challenges in fully utilizing these diverse data sets. In short, novel computational methods are needed to support kinome-wide comparison and analysis. Previously, for one or more ligands to be screened against a single target, a docking method³¹ would be applied to obtain the ligand bound complex and hence determine the binding mode. Another class of methods involves creating a pharmacophore model,³² in which the pharmacophore for the desired target protein is first trained. Then the trained pharmacophore is used to represent the structure–activity relationship (SAR) or to screen a compound library to select potential inhibitors. A promising method is to compare the binding modes of several kinases using a protein–ligand interaction fingerprint (IFP),^{25,27,33} which encodes the protein–ligand interfacial interaction as 1D fingerprints which discriminate the protein–ligand binding modes. The IFP method has been widely applied to docking pose analysis,^{25,34} scoring function refinement,³⁵ virtual screening,^{22,26,36,37} binding site comparison,³³ protein–ligand interaction mining,³⁸ binding mode prediction,³⁹ and other applications.³⁶ However, these studies have been limited to only one kinase or a family of high-homology kinases analyzed against multiple compounds. In these cases, the proteins can be aligned easily and the list of binding site residues, which constitute the binding pocket, are consistent for different protein structures. Thus, the encoded fingerprints for different compounds conserve the consensus sequences and can be compared easily. Here we extend the IFP method to determine the selectivity of kinase

inhibitors across the complete human structural kinome, which includes many diverse families. For IFPs to be comparable across the complete human kinome requires the accurate alignment of diverse binding pockets for all kinases so that the consensus sequences of binding site residues can be retrieved for further comparison.

It is not a trivial task to align a large number of diverse binding pockets using conventional global sequence or structure alignment methods.^{40,41} An alternative method is to use ligand-binding site (aka functional site) comparison that can detect remote structural, functional, and evolutionary relationships across fold space.⁴² Many methods have been developed to superimpose or detect binding sites (see recent reviews^{43–45} and herein). We have developed a sequence-order-independent profile–profile alignment (SOIPPA) method for ligand binding site comparison and implemented it in the software SMAP.^{40,42,46} SMAP has been successfully applied to side effect prediction,^{47,48} drug repositioning,^{49–51} polypharmacology drug design,^{52–55} and more.^{40,42,56,57} SMAP extracts aligned consensus sequence motifs, which consist of a list of binding site residues essential to derive the IFP. Moreover, the binding site residues are not only sequence-order-independent but also spatially matched providing pairwise aligned binding pockets. In this study, we further improve our SMAP method through combination with the protein–ligand IFP method. In our improved scheme, first, a comparable list of binding site residues is retrieved by superposing kinase complex structures using SMAP. Then the IFP is obtained by encoding the superposed function-site sequence motif for every kinase–ligand structure. Thus, the sequence order independent aligned function-site IFP strategy (termed Fs-IFP) provides a new solution to the comparison and analysis of diverse ligand binding modes (for details see the Method section).

In this paper, we use Fs-IFP to carry out the global analysis of ligand binding modes across the human kinome at an unprecedented scale and complexity (208 kinases, 1777 unique ligands, and their complexes) by integrating chemical genomics and structural genomics data. With what follows we first introduce the details of the function-site interaction fingerprint (Fs-IFP) approach. Second, we provide a detailed analysis of the relationship of ligand similarity, protein similarity, and Fs-IFP similarity. Third, we apply Fs-IFP to uncover the binding modes across the human structural kinome. Finally, we reveal the correlation of the binding mode and experimentally determined kinase binding profile. These findings provide new insights that can be applied to the *in silico* design of polypharmacology compounds targeting kinases.

Results and Discussions

Fs-IFP Data Set of Human Protein Kinase Structural Database (HPK-SD)

We obtained an Fs-IFP data set derived from a three-step protocol (see Method Section) as shown in Figure 1d, and the complete data set is shown in Supporting Information Table S1. In the Fs-IFPs data set, 2383 superimposed binding pockets are shown using the encoded bit strings. Each row represents the binding mode of one kinase–ligand complex structure using the bit strings. Each column represents the specific protein–ligand interactions of the same spatial position among the aligned kinase–ligand complex structures.

Distribution of Ligand-FP, Kinase-FP, and Fs-IFP Similarity

In HPK-SD, there are 2383 ligands, which contain 1777 unique small molecules. The top 20 most common ligands are shown in Figure 2a. It is not surprising that the top 3 most common molecules are ATP, ANP, and ADP because ATP is the natural ligand, and ANP and ADP are similar to ATP. The distribution of molecular weights for all ligands was shown in Figure 2b. The molecular weights of more than 80% of the ligands are between 300 and 550 Da, in keeping with Lipinski's rules.⁵⁸

On the basis of the ECFP fingerprint, the distribution of chemical structural similarity of ligands (ligand-FP) is shown in Figure 3 in red. We also compared the similarity of the binding sites of 2383 kinase structures in the HPK-SD (kinase-FP) (Figure 3, blue). In order to compare the distribution of ligand-FP similarity, kinase-FP similarity, and Fs-IFP similarity with each other, the distribution of the Fs-IFP similarity is shown in Figure 3 in green.

The distribution of the ligand-FP similarity has a mean TC of 0.12, a standard derivation of 0.085, and a long tail indicating there are some highly similar ligands in the HPK-SD, which is consistent with the results shown in Figure 2a. About 10 000 pairwise comparisons of ligand-FPs have TC values of 1.0 implying identical ligands, as our statistics show in Figure 2a. The top 5 redundant ligands are ANP, ADP, ATP, ACP, and STU. It will be interesting to subsequently analyze the binding modes of the Fs-IFPs based on the same ligands. However, the distribution of the ligand-FP similarity has a peak value when TC = 0.1, and the peak value is lower than that of the binding site or that of the kinase–ligand interaction as described in the following analysis.

The distribution of the similarity of the kinase binding sites as shown by kinase-FP (Figure 3 in blue) is close to a normal distribution with a mean of 0.18 and a standard derivation of 0.066. Given that the kinase structure has a highly similar scaffold,^{18,27} including one N-lobe consisting of five β sheets ($\beta 1$ – $\beta 5$) and one helix (C-helix), and a C-lobe consisting of four β sheets ($\beta 6$ – $\beta 9$) and additional helices (see labels in Figure 6b) and connected with a hinge-loop, the discriminatory power of kinase-FP is significant. Although the kinase scaffold is similar, the TC value of binding site fingerprints takes the difference of both amino acids and the direction of side chains into account. As a result, the similarity of the binding site is not high. The distinct property of the binding sites as described here by kinase-FP makes designing more selective kinase inhibitors possible. Different from the distribution of ligand similarity and binding-site similarity of kinases, the distribution of Fs-IFP similarity has a high mean value of 0.41 and a large standard deviation of 0.32. Comparatively, the average value of the similarity of the Fs-IFP is higher than that of ligand FPs or kinase pocket FPs across the human kinome. This implies that chemicals with diverse scaffolds achieve similar binding modes. It is noted that when TC = 0, there are about 12 000 pairs of Fs-IFPs. This suggests that pairs of protein–ligand interactions are in different regions of the kinase binding pocket, which do not overlap each other, or the pairs of protein–ligand interactions have different binding modes. For example, the ligand in PDB 1H8F (Figure 8d) and the ligand in 3Q4C (Figure 8h) bind to different locations of the binding pocket as shown in Figure 8i. The ligand in 3Q4C is at the entrance of ATP binding pocket clustered as cluster 5d in Figure 8i. The ligand in 1H8F binds to the allosteric site

clustered as cluster 4 in Figure 8i and interacts with the activation loop and the head of the C-helix. In addition, we noted that when the TC is larger than 0.75, the mean of the Fs-IFP similarity distribution is relatively lower than that of the ligand-FP similarity distribution. This implies that the binding mode is not highly similar even if there are many highly similar ligands. Thus, the same ligand may potentially accommodate a different pocket with different binding modes.

Correlation of Ligand-FP, Kinase-FP, and Fs-IFP

Figure 4 shows the correlation between the ligand structure, kinase binding pocket, and kinase–ligand binding mode by analyzing the relationship of pairwise similarity of ligand-FPs, kinase-FPs, and Fs-IFPs.

As expected, similar pockets bind similar ligands. The correlation of the kinase pocket and the ligand characteristics is shown in Figure 4a and Figure 4b, respectively. From the point of view of the ligand (Figure 4b), when the TC value of ligand similarity changes from 0.0 to 1.0, the distribution of the TC values of the kinase pocket similarity shows no significant change, with all medians approximately 0.19. This suggests that the ligand similarity cannot directly infer the binding pocket similarity. In contrast, there is some correspondence between kinase pocket similarity and ligand similarity for the most similar pockets as shown in Figure 4a. For the 0.12% of pockets with similarity larger than $TC = 0.5$, the median of the TC for ligand similarity increases from 0.3 to 1.0 with the increase in the TC of pocket similarity. The correlation coefficient for pairs within this range is $R^2 = 0.28$. This result is consistent with the conventional view that similar pockets bind similar ligands.⁵⁹

We further characterize how the protein–ligand binding modes are correlated with the characteristics of ligands and kinase pockets. Figure 4c and Figure 4d show the correlation of the ligand-FPs and the Fs-IFPs and that of the kinase pocket-FPs and the Fs-IFPs, respectively. The same basic features are observed in the two panels. The medians of both ligand-FP similarity and kinase pocket-FP similarity increase slowly with the increase of Fs-IFP similarity. This suggests that Fs-IFP is to some degree correlated not only with the ligand characteristics but also the kinase pocket properties. In Figure 4c, the medians of the TC value of ligand similarity slowly increase along with the TC value of Fs-IFP similarity. For the 1.04% of Fs-IFP's with similarity larger than $TC = 0.7$, the median of the TC for ligand similarity increases from 0.2 to 0.6 with the increase in the TC of Fs-IFP similarity. The correlation coefficient for pairs within this range is $R^2 = 0.24$. In addition, at each range of the TC value of Fs-IFP similarity, the TC value of ligand similarity has an extensive distribution from 0.0 to 1.0 as shown in the direction of Y-axis. This suggests that significantly different ligands (low ligand-FP value) may bind to the kinase pocket with similar binding modes (high Fs-IFP value). Alternatively, ligands may be promiscuous and adaptively bind to similar binding pockets. Moreover, in Figure 4e, where the ligand-similarity TC is 1.0, the median of the TC value for Fs-IFP is approximately 0.5. This implies that the same ligand can bind to proteins with different binding modes. Our data set includes several redundant ligands, such as ANP, ADP, ATP, and ACP, as aforementioned. They are highly similar to each other but have dissimilar binding modes. We can infer that the different binding modes enable these ligands to better accommodate different pockets.

The correlation coefficient for pairs within this range is $R^2 = 0.23$. It follows, and is confirmed by Figure 4f, that similar kinase pockets have similar Fs-IFPs. There is some correspondence between kinase pocket similarity and Fs-IFP similarity for similar pockets, as shown in Figure 4f. The correlation coefficient for pairs within this range of 0–1.0 is $R^2 = 0.23$, and the violin boxes of different bins have the larger overlap. We also noted that when $TC = 1.0$ of kinase-FP, there is 0.002% of Fs-IFP binding modes with the different TC from 0.4 to 1.0 for Fs-IFP, which implies that different ligands can bind the highly similar pockets with different binding modes.

Structural and Evolutionary Characteristics of Kinase–Ligand Interactions

In the process of encoding the Fs-IFP, each 7 bits represent 7 different contacts for one residue (see Method section). For HPK-SD, there are 37 954 contacts between the ligand and the kinase in the kinase–ligand binding interfaces, and these can be characterized as follows. They include 83.2% apolar (van de Waals) contacts, 8.82% contacts where the H-bond of the protein is a donor, 5.91% where the H-bond of the protein is an acceptor, 1.07% contacts of aromatic edge to face, 1.0% contacts of aromatic face to edge, and 0.01% electrostatic contacts, respectively, as shown in Figure 5a. On average, there are 16 contacts for each kinase–ligand interaction. The 16 contacts forming the kinase–ligand interaction typically consist of 13 apolar, 1 aromatic, and 2 H-bond contacts, respectively. We counted the number of interactions that occur on each residue of the superimposed protein–ligand pocket for different types of interactions. As shown in Figure 5b, the top 5 apolar interactions are located at positions $\beta 2$, $\beta 3$, hinge (two), and $\beta 7$. The top 3 aromatic interactions are located at the hinge (two) and the DFG peptide. The top 4 H-bond contacts are located at the positions $\beta 3$, hinge (two), and the DFG peptide. These observations suggest that $\beta 2$, $\beta 3$, $\beta 7$, hinge, and DFG provide the majority of the protein–ligand interactions.^{17,18}

Next we inspect which amino acids contribute to specific contacts. At each location the number of interactions is shown in Figure 6a. The amino acids involved in the top 6 most common interactions are located at positions $\beta 2$, $\beta 3$, hinge (two), $\beta 7$, and the gatekeeper (the first residue of hinge region). Their locations are marked using the top 6 largest balls, as shown in Figure 6b, where the size of the purple spheres is proportional to the number of interactions. The 6 amino acids that form the top 6 most conserved interactions are located at the binding site of adenosine from ATP and form the core of the binding pocket. The residues at the regions of $\beta 2$, $\beta 3$, hinge, and DFG also have higher hit frequencies than other locations and thus make major contributions to the binding. Typically there are one to three hydrogen bond contacts between the adenosine and the main-chain atoms of the hinge residues. The aromatic ring of adenosine forms aromatic interaction with an amino acid on the sheet of $\beta 3$ and an amino acid around DFG²⁷ and is in keeping with the observation that most of the current kinase-targeted drugs are ATP-competitive.^{17,29} Notwithstanding these similarities, other amino acids contribute to different binding mode in different structures and may play important roles in establishing selectivity for targeted polypharmacology.

We further delineated what kinds of residues contribute to the binding contacts at specific locations in order to reveal the role of residue conservation in binding specificity. At every

residue location of the superimposed binding sites, we inspected the constituent amino acid residues and calculated their ratios. The amino acid with the maximum ratio is defined as the conserved residue at that specific position. The results are shown in Figure 6c. The top 3 most conserved residues are Val, Ala, and Leu, located at $\beta 2$, $\beta 3$, and $\beta 7$, respectively. They have the highest interaction frequencies marked as ①, ②, and ④ in Figure 6a. Residue conservation is linearly correlated with the frequency of interactions on each residue as shown in Figure 6d (correlation coefficient 0.89). The location corresponding to these three most conserved amino acids forms the mouth of the core binding pocket, which traps the ligand, as shown in Figure 6b. Beyond that the conservation in the hinge region is also relatively high. Hinge interactions mainly take place through the main-chain atoms of the hinge residues.¹⁷ Mutation of hinge residues rarely affects ATP binding⁶⁰ but frequently changes the conformation of the kinase hydrophobic pocket resulting in drug resistance,⁶¹ such as the T315I mutation in Abl kinase and T670I mutation in KIT kinase.⁶² The hinge region may be responsible for subtle selectivity resulting from different side chain conformations and types of kinases. The C-helix region and the catalytic region have lower conservation as shown in Figure 6c and may be responsible for the specific selectivity of a ligand. The catalytic region is distal to the ATP site, and the C-helix contributes to forming the hydrophobic pocket and the allosteric site. Recently these regions have been increasingly utilized to design allosteric inhibitors and type II inhibitors to achieve improved selectivity.^{15,63}

Clustering the Fs-IFP Data Set

To further explore the kinase binding site, we cluster the Fs-IFPs of complexes (see Methods) and obtain a hierarchical clustering tree as shown in Figure 7. Setting the Fs-IFP similarity as the distance for clustering reveals five clusters, marked 1–5. All structures in each cluster are listed in Supporting Information Table S2. Among the five clusters, the binding modes are completely different. Such information may provide critical clues to targeted polypharmacology by utilizing binding pocket promiscuity and selectivity. To further show the spatial differences between binding pockets, we analyzed the binding modes for the five clusters individually. Approximately 90% of the structures in cluster 5 have a common character, i.e., have at least one hydrogen-bond interaction with hinge region. To further reveal the interaction characters, we divide cluster 5 into 4 subclusters: cluster 5a–d. We randomly choose eight PDB entries one from each cluster/subcluster to illustrate the protein–ligand interactions: 4ITJ, 2A5U, 3PXF, and 1H8F are chosen to represent clusters 1–4, and 3PRZ, 1BYG, 1ZZ2, and 3Q4C are chosen to represent clusters 5a–d. Their 3D structures are shown in Figure 8.

Figure 8a shows the binding mode for cluster 1 where the ligand is located within the hydrophobic cavity at the back of the ATP binding site, also shown in red in Figure 8i. Gatekeeper residues control the size of the cavity and the interaction with the ligand. The gatekeeper residue is different in different kinases, for example, threonine in EGFR but phenylalanine in CDK-2, providing the opportunity for high selectivity when the cavity is utilized. The success of the first generation inhibitors, including erlotinib, gefitinib, and lapatinib^{8,64} for non-small-cell lung cancer targeted EGFR,⁶⁵ is thought to make use of the interactions with the hydrophobic cavity.⁶⁶ Because of drug resistance, especially through

mutation of the gatekeeper T790M,⁶⁷ the new generation of EGFR inhibitors, such as *N*-[2-[2-(dimethylamino)-ethylmethylamino]-4-methoxy-5-[[4-(1-methylindol-3-yl)-pyrimidin-2-yl]amino]phenyl]prop-2-enamide (AZD9291),⁶⁸ *N*-[3-[[2-[4-(4-acetylpiperazin-1-yl)-2-methoxyanilino]-5-(trifluoromethyl)pyrimidin-4-yl]amino]phenyl]prop-2-enamide (CO-1686),⁶⁹ and *N*-[3-[5-chloro-2-[2-methoxy-4-(4-methyl-piperazin-1-yl)anilino]pyrimidin-4-yl]oxyphenyl]prop-2-enamide (WZ4002),⁷⁰ are designed to overcome the T790M mutation. Existing examples of the key gatekeeper pocket can be utilized to achieve the desired selectivity of kinase targeted inhibitors.

Figure 8b shows the location of cluster 2 in the center of the kinase pocket as shown in Figure 8i. The ligand has three direct interactions, with the Asp of the DFG tripeptide, Lys on the “roof” of $\beta 3$, and Glu on the C-helix. The location of the three interactions is just proximal to the ATP site and is at the catalytic center of the kinase, where the ATP is catalyzed to ADP.⁷¹ The occupation of this space inhibits the catalytic activity of the kinase as utilized by the FDA approved drugs, vemurafenib⁷² and ceritinib.⁷³ As a BRAF mutant (V600E) inhibitor to treat the late-stage melanoma, vemurafenib was designed using fragment-based lead discovery. Its selectivity was achieved by adding the sulfonamido group at the location of cluster 2. Similarly, ceritinib, a drug to treat ALK-positive metastatic non-small-cell lung cancer, also applied the same strategy of using a sulfonyl group.

Figure 8c shows the location of cluster 3, which accommodates the space at the back of the ATP binding site, which is between $\beta 5$ and the C-helix. There are H-bond interactions with the residue on $\beta 3$ and the residue on the C-helix. Between the $\beta 5$ and C-helix, typically, there is not too much space available. However, when the kinase is in the C-helix-out conformation such as in PDB 1XKK or 3BBT, the location of cluster 3 can be exposed and has been exploited.⁷⁴ The allosteric inhibitors 5-bromo-*N*-(2,3-dihydroxypropoxy)-3,4-difluoro-2-((2-fluoro-4-iodophenyl)amino)benzamide (PD318088)⁷⁵ and *N*-[5-[3,4-difluoro-2-[(2-fluoro-4-iodophenyl)amino]phenyl]-1,3,4-oxadiazol-2-yl]-4-morpholine-ethanamine (PD334581) (type III inhibitors)⁷⁵ have been reported to explore this location on cluster 3.

Figure 8d shows a binding mode between the head of the C-helix and the activation loop. The location of the subpocket is away from the ATP binding site. The tail of the FDA approved drug ponatinib⁷⁶ occupies this part of the subpocket. Presumably this subpocket can be further explored to improve the selectivity of the primary target of ponatinib, BCR-Abl, as well as other kinase targets.⁷⁷

For the aforementioned four clusters, a common pattern is that there is no interaction with the hinge region. These subpockets are located at the back of the ATP binding site and adjacent to each other. These regions are often used to improve selectivity, and the regions including cluster 1–3 are often called as the “selectivity pocket” to differentiate them from the ATP binding pocket. Typically, kinase inhibitors use these regions to achieve selectivity. Type I inhibitors often utilize the hydrophobic subpocket at the site of cluster 1 to improve the selectivity; types II and III occupy more space at the regions of cluster 1–4 to achieve the specific selectivity. Conversely, it is very challenging to design desired inhibitor using the regions corresponding to cluster 1–4 because these regions are more flexible than the ATP

binding site, depending on the conformational change of the inactive kinase state (DFG out or C-helix out).⁹ Currently, most kinase inhibitors are type I inhibitors, which are ATP-competitive and occupy the ATP binding site as shown in cluster 5 in Figure 7 and Figure 8e–h. Besides the common hinge interactions, they have distinguishing characteristics and thus divided into four subclusters.

Figure 8e shows the binding mode of cluster 5a, which interacts with the hinge region through two H-bonds. Moreover, the ligand interacts with the Lys and Glu of the C-helix. This interaction pattern is similar to that of cluster 2. In fact, cluster 5a is close to cluster 2 as described for the cluster 2 (Figure 8b). The important difference is the hinge interaction or not. Thus, to achieve the selectivity of cluster 2, fragments that occupy the location of the cluster 5 can be added to the inhibitor of cluster 2.

Figure 8f shows the typical binding mode of the ATP-competitive site of cluster 5b. Besides the interactions of the hinge region, there is an interaction with a residue on the activation loop. To achieve selectivity, the ligand at this location should be extended to other subpockets, such as the adjacent subpockets: cluster 2, cluster 5a, cluster 5c, and cluster 5d.

Figure 8g shows the binding site of cluster 5c, which has interactions with the activation loop and the DFG tripeptide. Similar to cluster 5b, the ligand has a longer tail and extensively utilizes the distal space of the ATP binding site. It will affect the phosphorylation of the substrate because the tail is close to the binding site of the substrate. For example, the approved drug afatinib is selective in this way (see the cocrystal structure PDB entry 4G5J).⁷⁸

Figure 8h illustrates the binding mode of cluster 5d at the entrance of the binding pocket adjacent to the adenosine binding site of the ATP binding pocket, as shown in Figure 8i. Besides the hinge interaction, the ligand has interactions with the loop following the hinge residues. Because the entrance is located between the loop following the hinge region and the P-loop, different kinases present diverse characteristics. This “selectivity entrance” can be utilized to achieve improved selectivity compared to the ATP binding site.¹⁷ Approximately 57% of the approved kinase-targeted drugs take advantage of this selectivity entrance.⁶⁴ In 2013 two irreversible kinase inhibitors ibrutinib⁷⁹ and afatinib⁸⁰ were approved that utilize the selectivity entrance.⁸¹ These inhibitors covalently inhibit their targets and compete with ATP. The covalent bond is formed by a Michael addition with a Cys residue of the kinase.⁸² The Cys residues involved in the two approved drugs are both located at the selectivity entrance.³ The third generation of EGFR inhibitors for NSCLC, osimertinib⁶⁸ and rociletinib,⁶⁹ are also irreversible, and both involved the Cys residue at the entrance. The Cys residue has been a hot-point amino acid for the design of irreversible kinase inhibitors and significantly improves the binding affinity.⁸³

From Figure 8a to Figure 8h, every cluster has a different binding site as shown in the cartoon panel of Figure 8i, which provides an overview of the relative locations in the kinase pocket. The promise is to design kinase inhibitors by choosing a type of binding mode or combining binding modes. Indeed, potential inhibitors have been developed by targeting the different position from the typical ATP site.¹⁸ For example, there are 30 kinase-targeted

drugs approved by FDA through Nov 2015, and one drug apatinib⁸⁴ approved by the Chinese FDA in Dec 2014. Apatinib has a typical binding mode for a type I inhibitor and occupies clusters 2, 5a, 5b, and 5d. The binding modes of 30 drugs approved by the FDA are summarized in Table 1 based on their binding positions. The 3D binding mode can be retrieved and checked based on the corresponding PDB⁸⁵ entries in Table 1 or from the reviews of FDA approved drugs.^{64,81} The types of drugs were also identified in Table 1 based on the binding position and the conformational state of the kinase.

Many drugs targeting kinases do so via the ATP binding site. Also the “selectivity pocket” and the “selectivity entrance” are targeted frequently. In the previous section we have shown that the amino acids involved in the binding of the adenosine of ATP are conserved, as shown in Figure 6b,c. Thus, it is possible to take advantage of a combination of different binding modes on different subpockets to achieve the desired selectivity profile using a fragment-based strategy.¹⁷ All existing drugs utilized more than one binding region to achieve their specificity (Table 1). It is noted that trametinib and cobimetinib,⁸⁶ novel allosteric inhibitors, do not use the ATP binding interaction with the hinge but use the cluster 1–3 and cluster 5c regions. This type of approved drug may become more prevalent in the future. Overall, the five different spatial locations not only reflect pocket promiscuity but ligand adaptability. There is no distinct boundary between clusters, which implies that the whole kinase pocket should be used to refine the design of the inhibitor in order to obtain the optimal specificity.

The flexibility of protein–ligand binding is another important factor in designing a desired inhibitor. Typically, type II inhibitor induces the dramatic change of active loop to rearrange the inhibitor-stabilized conformation.⁸⁷ The induced conformational change is not unique for type II inhibitors, and an example is axitinib, which induced different binding modes in different kinase structures (e.g., in PDB entries 4AGC and 4TWP).^{88,89} In wild-type VEGFR, axitinib inhibits the target with a type II binding mode and DFG-out state as shown in the crystal structure (PDB entry 4AGC). In the wide-type VEGFR, the gatekeeper is a small amino acid Val916; moreover the tail of axitinib consists of a phenyl fragment, which has a close interaction with Val916. Recently, a study shows axitinib can inhibit the T315I mutant bcr-abl1 with a distinct binding conformation.⁸⁹ Here the T315I mutation blocked axitinib from entering the region of the selectivity pocket; thus axitinib induced a new type I binding mode. The mutation prevents the phenyl fragment from entering the hydrophobic subpocket, resulting in a flip of the phenyl group to accommodate the mutation, as shown in the X-ray structure (PDB entry 4TWP). This poses another challenging task in designing the desired selectivity.

Analysis of the Correlation between the Fs-IFP Data Set and the Experimentally Determined Kinase Binding Profile

We further evaluate the correlation of the Fs-IFP and the experimentally determined kinase binding profile from KINOMEScan²¹ data taken from the HMS LINCS database.⁹⁰ We use the common small molecules in the KINOMEScan data and the ligands from HPK-SD to build relationships between the experimental activity values and the binding mode of Fs-IFP (see detailed procedures in the Method section). Figure 9a shows the pairwise correlation of

the thermodynamic constant K_d and Fs-IFP. The pairwise similarity of K_d , which is measured by Euclidean distance, is to a certain degree correlated with the binding mode of Fs-IFP as indicated from the trend in the curve fitting first-order exponential decay. The correlation coefficient is 0.40. When the pairwise similarity of Fs-IFP is higher than 0.5, the binding profile determined by the experimental K_d tends to be similar and vice versa. Figure 9b shows the relationship between percent of control (PoC) and Fs-IFP and trends in a similar way to Figure 9a. The correlation coefficient is 0.28. The two trend curves both show that the more similar are the binding modes, the more consistent the experimental values are with each other. Thus, Fs-IFP could be a useful in silico tool for virtual drug development, since Fs-IFP is related to the experimental profile. For a given drug as a reference, if a newly designed inhibitor has a high similar Fs-IFP, such as more than 0.5, it may imply the inhibitor has similar activity to the reference drug. In Supporting Information Figure S1, we also show that there is no significant correlation between the kinase pocket IFP and the experimental profile, further validating that Fs-IFP could be a better measure than pocket similarity to drive drug design.

It is worth noting that the encoded Fs-IFP interaction excludes interactions from any water molecules and metal ions in the binding site. Water molecules, metal ions, and molecular dynamics are important aspects of the ligand interaction. In future work the intent is that Fs-IFP will better mimic the realistic kinase–ligand interaction by including water molecules, metal ions, and molecular dynamics. In so doing the Fs-IFP method will better discriminate binding modes, subtle interactions, and conserved interactions.

In summary, understanding the structural basis of ligand binding selectivity and promiscuity across the human kinome provides critical information on drug repurposing and discovering targeted polypharmacology. We have demonstrated that Fs-IFP is a potentially powerful tool to study ligand binding profiles across the kinome. Currently Fs-IFP is encoded using the residues that cover all subpockets. Since a ligand may only occupy one or more, but not all, subpockets, an Fs-IFP that is based on subpockets may offer greater predictive power in studying protein–ligand interactions on a genome-wide scale. By integrating Fs-IFP with other computational tools such as historeceptomic fingerprints for druglike compounds⁹¹ and multiple omics data, in the future it should be possible for us to develop more powerful predictive models for polypharmacology prediction to support drug repositioning and target identification.

Conclusion

In this paper, we have presented an Fs-IFP method, which combines sequence-order-independent binding site alignment with a protein–ligand IFP technique to identify all interaction patterns for human kinases. We characterized the binding modes across the human structural kinome using sets of comparable Fs-IFPs and gained new insights into the structural discriminants of binding specificity and promiscuity. Moreover, we have analyzed the relationships between Fs-IFPs, ligand-FPs, and kinase-FPs. Our analysis supports the notion that similar binding pockets bind similar ligands with similar interaction patterns. However, similar ligands may bind to diverse binding pockets with different interaction patterns. The binding pocket clustering based on Fs-IFP similarity reveals several distinct

subpockets. The residue conservation in these subpockets is highly correlated with specific protein–ligand interactions. The selectivity of inhibitors can be achieved by combining the different subpockets. We should pay more attention to the “selectivity pocket” and “selectivity entrance” in order to obtain the desirable selectivity of a specific target or targeted polypharmacology. Furthermore, the Fs-IFP is correlated with kinome binding profiles from experimental data. Thus, the comparison of interaction patterns using the Fs-IFP across the kinome will facilitate the detailed understanding of the kinase binding mode and rational design of kinase inhibitors with desirable selectivity profiles and SAR. In principle, the Fs-IFP method can be applied to recognize protein–ligand interaction pattern across all of fold space, thereby providing a potentially powerful tool for revealing polypharmacology profiles of proteins beyond kinases.

Method

Fs-IFP Method

Our Fs-IFP method follows three steps as shown in Figure 1.

Step 1: To Prepare the Data Set of Human Protein Kinase Ligand-Bound Structures

—We first downloaded all UniProt entries of human protein kinases (HPK) from the Universal Protein Resource (UniProt) based on the document “Human protein kinases: classification and index” released by UniProt in June, 2015.^{92,93} Then for every kinase UniProt entry, we retrieved the PDB entries of all kinases and downloaded all kinase structures from the Protein Data Bank (PDB).^{11,85} Third, we excluded the PDB structures built by homology modeling and without the kinase catalytic domain; thus we obtained all kinase structures comprising 2786 structures from 235 human kinases. Finally, by exclusion of the nonligand bound kinase structures, as many as 2383 kinase–ligand complex structures, which we refer to as the HPK structural data set (HPK-SD), were obtained.

Step 2: To Compare Kinase Binding Sites Using SMAP—We superimposed the binding sites of all structures in the HPK-SD set by an all-against-all comparison using SMAP.^{40,42} Here the SMAP parameter of “Ligand Contact Distance Cutoff” was set as 6.5 Å, which is an interaction distance threshold used frequently in the literatures.^{94,95} Other SMAP parameters were set as default values. By use of SMAP, the function-site sequence motif, which constitutes the binding pocket, was obtained. Moreover, the function-site sequence motif is pairwise aligned. In addition, SMAP is a fast method for proteome-wide ligand-binding site comparison.⁴⁰

Then we chose a function-site sequence motif as the representative template. We noted the binding pocket of the structure of a lymphocyte-specific protein tyrosine kinase (LCK; PDB entry 3BYU) consisted of the highest number of amino acid residues (80) as shown in Figure 1b and covered all other aligned function-site sequence motifs. Thus, the function-site sequence motif in PDB 3BYU was selected as the reference template for the binding pocket.

For a given ligand-bound structure, we aligned it with the template 3BYU and retrieved the aligned amino acid residues to comprise the binding site. To present all binding pockets

using the same number of amino acids, we marked the position as “D-NULL” if the position corresponding to the template 3BYU did not have amino acid that contributes to the binding pocket. Thus, we could superimpose every binding pocket for all structures in HPK-SD and retrieve and align all binding site residues that constitute the binding pocket for these structures. We represented the binding pocket using a feature vector with the same number of elements (amino acid residues or “D-NULL”).

Step 3: To Encode Fs-IFP for HPK-SD—To encode the Fs-IFP, the polar hydrogen atoms were added for every function-site sequence motif and the corresponding ligand. Then we generated Fs-IFPs for the human kinase data set using PyPlif software by Radifar et al.⁹⁶ PyPlif is a python implementation of IFP.^{25,33,34,36,97} For every residue in the binding site, a 7-bit array was encoded. The seven bits represented seven types of interactions using the interactional geometric rules:³⁴ (1) apolar (van der Waals); (2) aromatic face to face; (3) aromatic edge to face; (4) hydrogen bond (protein as hydrogen bond donor); (5) hydrogen bond (protein as hydrogen bond acceptor); (6) electrostatic interaction (protein positively charged); (7) electrostatic interaction (protein negatively charged). Then an Fs-IFP with a length of 560 bits (7 bits $\times$ 80 residues) was encoded according to the list of binding site residues for every kinase–ligand complex structure. Thus, we obtained an Fs-IFP data set based on our HPK-SD. Here PyPlif software is used to encode and analyze IFP, which is very fast and does not need significant computer resource.⁹⁶

Fs-IFP Similarity and Clustering

For the HPK's Fs-IFP data set, we carried out an all-against-all comparison by calculating the pairwise Tanimoto coefficient (TC) between any two Fs-IFPs.⁹⁸ We further clustered all ligand-bound structures using the TC values of Fs-IFPs as similarity measure. The R software was used for clustering and drawing the hierarchical tree; the complete linkage clustering method was used using the R-function hclust with the A2R library.⁹⁹

Ligand Similarity

To determine the ligand characteristics in HPK-SD, we collected the information for all ligands extracted from all structures in HPK-SD. We described every ligand as a 1024-bit extended-connectivity fingerprint (ECFP), which is a circular topological fingerprint designed for molecular characterization.^{100–102} The ECFP was calculated using the Screenmd function Addin in Chemaxom.¹⁰² Then we compared the pairwise similarity of all ligands by calculating the pairwise TC.

Kinase Binding Site Similarity

Besides the attributes of protein–ligand binding modes and ligands, the features of all binding sites in HPK-SD were compared with each other to inspect the characteristics of the binding pockets, such as similarity and relationship with the attributes of ligand binding modes and ligands. First we defined the binding site for every ligand-bound structure using all amino acids that present one heavy atom closer than 6.5 Å to any heavy atom of the ligand.⁹⁵ Then the binding site was encoded into a fingerprint and compared using the FuzCav software,¹⁰³ which is a novel alignment-free high-throughput algorithm to compute pairwise similarities between the binding sites by calculating the TC.

KINOMEScan Data Analysis

We collected the compound-kinase profiling assay data from KINOMEScan,²¹ a large commercial kinase assay panels. KINOMEScan measures kinase-compound binding profiling shown by PoC (100 PoC means “no hit”; 0 PoC means “strong hit”) and the thermodynamic inhibitor constant K_d . Through Sep 22, 2015, the Harvard Medical School Library of Integrated Network-Based Cellular Signatures (HMS LINCS) contained 159 sets of small molecules profiling reports for use with KINOMEScan.⁹⁰ There are 87 compounds with the PoC profiling results in the presence of 10 μM assay compound concentration, 16 compounds with the PoC profiling results in the presence of 1 μM assay compound concentration, and 56 compounds with the profiling results of binding constants (K_d). Thus, we first evaluated the small molecule overlap between KINOMEScan data and our HPK-SD data set. We obtained 39 compounds with both X-ray complex structures and the compound-kinase PoC value at the screening concentration of 10 μM and 15 compounds with both X-ray complex structures and the compound-kinase PoC value at the screening concentration of 1 μM (shown in Supporting Information Table S3). There are 68 compounds with both X-ray complex structures and corresponding K_d values. Among them, 58 compounds have K_d less than 100 nM (Supporting Information Table S4).

We further studied the correlation of the experimental data and the binding mode derived from the Fs-IFP. Here we chose all PoC results from Supporting Information Table S3 at the screening concentration of 10 μM to obtain the correlation with the binding mode using Fs-IFP. The PoC results at the screening concentration of 1 μM were not included because the number of data points was too small. Meanwhile, we also counted all K_d results from Supporting Information Table S4. Then we obtained the pairwise similarity by calculating the Euclidean distance of any two experimental values in the PoC data (or in K_d data) corresponding to the pairwise similarity of Fs-IFP using the TC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health (Z.Z. and P.E.B.), the National Library of Medicine, National Institutes of Health under Award R01LM011986 (L.X.), and the National Institute on Minority Health and Health Disparities, National Institutes of Health, under Award G12MD007599 (L.X.). We also appreciate Dr. Enade Perdana Istyastono et al. for providing the IFP software, Pyplif and Dr. T. Exner et al. for providing the docking program PLANTS.

References

1. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science*. 2002; 298:1912–1934. [PubMed: 12471243]
2. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol*. 2011; 29:1046–1051. [PubMed: 22037378]
3. Ward, RA.; Goldberg, F. *Kinase Drug Discovery*. RSC Publishing; Cambridge, U.K.: 2012.
4. Klebl, B.; Müller, G.; Hamacher, M. *Protein Kinase as Drug Targets*. Vol. 49. Wiley-VCH Verlag GmbH & Co. KGaA; Weinheim, Germany: 2011.

5. Goldman JM, Melo JV. Chronic myeloid leukemia—advances in biology and new approaches to treatment. *N Engl J Med*. 2003; 349:1451–1464. [PubMed: 14534339]
6. Fedorov O, Muller S, Knapp S. The (un)targeted cancer kinome. *Nat Chem Biol*. 2010; 6:166–169. [PubMed: 20154661]
7. Sarnpitak P, Mujumdar P, Taylor P, Cross M, Coster MJ, Gorse AD, Krasavin M, Hofmann A. Panel docking of small-molecule libraries - Prospects to improve efficiency of lead compound discovery. *Biotechnol Adv*. 2015; 33:941–947. [PubMed: 26025037]
8. Wu P, Nielsen TE, Clausen MH. Small-molecule kinase inhibitors: an analysis of FDA-approved drugs. *Drug Discovery Today*. 2016; 21:5–10. [PubMed: 26210956]
9. Muller S, Chaikuad A, Gray NS, Knapp S. The ins and outs of selective kinase inhibitor development. *Nat Chem Biol*. 2015; 11:818–821. [PubMed: 26485069]
10. Jacoby E, Tresadern G, Bembenek S, Wroblowski B, Buyck C, Neefs JM, Rassokhin D, Poncelet A, Hunt J, van Vlijmen H. Extending kinome coverage by analysis of kinase inhibitor broad profiling data. *Drug Discovery Today*. 2015; 20:652–658. [PubMed: 25596550]
11. Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, Westbrook J. The Protein Data Bank and the challenge of structural genomics. *Nat Struct Biol*. 2000; 7(Suppl):957–959. [PubMed: 11103999]
12. Kinnings SL, Jackson RM. Binding site similarity analysis for the functional classification of the protein kinase family. *J Chem Inf Model*. 2009; 49:318–329. [PubMed: 19434833]
13. McGregor MJ. A pharmacophore map of small molecule protein kinase inhibitors. *J Chem Inf Model*. 2007; 47:2374–2382. [PubMed: 17941626]
14. Huse M, Kuriyan J. The conformational plasticity of protein kinases. *Cell*. 2002; 109:275–282. [PubMed: 12015977]
15. Zhao Z, Wu H, Wang L, Liu Y, Knapp S, Liu Q, Gray NS. Exploration of type II binding mode: a privileged approach for kinase inhibitor focused drug discovery? *ACS Chem Biol*. 2014; 9:1230–1241. [PubMed: 24730530]
16. Furtmann N, Hu Y, Bajorath J. Comprehensive analysis of three-dimensional activity cliffs formed by kinase inhibitors with different binding modes and cliff mapping of structural analogues. *J Med Chem*. 2015; 58:252–264. [PubMed: 25054653]
17. Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer*. 2009; 9:28–39. [PubMed: 19104514]
18. Liao JJ. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J Med Chem*. 2007; 50:409–424. [PubMed: 17266192]
19. Huang D, Zhou T, Lafleur K, Nevado C, Caflisch A. Kinase selectivity potential for inhibitors targeting the ATP binding site: a network analysis. *Bioinformatics*. 2010; 26:198–204. [PubMed: 19942586]
20. Cui JJ, Tran-Dube M, Shen H, Nambu M, Kung PP, Pairish M, Jia L, Meng J, Funk L, Botrous I, McTigue M, Grodsky N, Ryan K, Padrique E, Alton G, Timofeevski S, Yamazaki S, Li Q, Zou H, Christensen J, Mroczkowski B, Bender S, Kania RS, Edwards MP. Structure based drug design of crizotinib (PF-02341066), a potent and selective dual inhibitor of mesenchymal-epithelial transition factor (c-MET) kinase and anaplastic lymphoma kinase (ALK). *J Med Chem*. 2011; 54:6342–6363. [PubMed: 21812414]
21. 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; 23:329–336. [PubMed: 15711537]
22. Chuaqui C, Deng Z, Singh J. Interaction profiles of protein kinase-inhibitor complexes and their application to virtual screening. *J Med Chem*. 2005; 48:121–133. [PubMed: 15634006]
23. Patricelli MP, Nomanbhoy TK, Wu J, Brown H, Zhou D, Zhang J, Jagannathan S, Aban A, Okerberg E, Herring C, Nordin B, Weissig H, Yang Q, Lee JD, Gray NS, Kozarich JW. In situ kinase profiling reveals functionally relevant properties of native kinases. *Chem Biol*. 2011; 18:699–710. [PubMed: 21700206]

24. Shemon AN, Heil GL, Granovsky AE, Clark MM, McElheny D, Chimon A, Rosner MR, Koide S. Characterization of the Raf kinase inhibitory protein (RKIP) binding pocket: NMR-based screening identifies small-molecule ligands. *PLoS One*. 2010; 5:e10479. [PubMed: 20463977]
25. Deng Z, Chuaqui C, Singh J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J Med Chem*. 2004; 47:337–344. [PubMed: 14711306]
26. Witek J, Smusz S, Rataj K, Mordalski S, Bojarski AJ. An application of machine learning methods to structural interaction fingerprints—a case study of kinase inhibitors. *Bioorg Med Chem Lett*. 2014; 24:580–585. [PubMed: 24374279]
27. van Linden OP, Kooistra AJ, Leurs R, de Esch IJ, de Graaf C. KLIFS: a knowledge-based structural database to navigate kinase-ligand interaction space. *J Med Chem*. 2014; 57:249–277. [PubMed: 23941661]
28. Bryant DH, Moll M, Finn PW, Kavraki LE. Combinatorial clustering of residue position subsets predicts inhibitor affinity across the human kinome. *PLoS Comput Biol*. 2013; 9:e1003087. [PubMed: 23754939]
29. Cui JJ. Targeting receptor tyrosine kinase MET in cancer: small molecule inhibitors and clinical progress. *J Med Chem*. 2014; 57:4427–4453. [PubMed: 24320965]
30. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. The Pfam protein families database. *Nucleic Acids Res*. 2012; 40:D290–301. [PubMed: 22127870]
31. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. A critical assessment of docking programs and scoring functions. *J Med Chem*. 2006; 49:5912–5931. [PubMed: 17004707]
32. Koes DR, Camacho CJ, Pharmmer: efficient and exact pharmacophore search. *J Chem Inf Model*. 2011; 51:1307–1314. [PubMed: 21604800]
33. Mordalski S, Kosciolk T, Kristiansen K, Sylte I, Bojarski AJ. Protein binding site analysis by means of structural interaction fingerprint patterns. *Bioorg Med Chem Lett*. 2011; 21:6816–6819. [PubMed: 21974955]
34. Marcou G, Rognan D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J Chem Inf Model*. 2007; 47:195–207. [PubMed: 17238265]
35. Mpamhanga CP, Chen B, McLay IM, Willett P. Knowledge-based interaction fingerprint scoring: a simple method for improving the effectiveness of fast scoring functions. *J Chem Inf Model*. 2006; 46:686–698. [PubMed: 16562999]
36. Perez-Nueno VI, Rabal O, Borrell JJ, Teixido J. APIF: a new interaction fingerprint based on atom pairs and its application to virtual screening. *J Chem Inf Model*. 2009; 49:1245–1260. [PubMed: 19364101]
37. Wassermann AM, Geppert H, Bajorath J. Searching for target-selective compounds using different combinations of multiclass support vector machine ranking methods, kernel functions, and fingerprint descriptors. *J Chem Inf Model*. 2009; 49:582–592. [PubMed: 19249858]
38. Deng Z, Chuaqui C, Singh J. Knowledge-based design of target-focused libraries using protein-ligand interaction constraints. *J Med Chem*. 2006; 49:490–500. [PubMed: 16420036]
39. Chupakhin V, Marcou G, Baskin I, Varnek A, Rognan D. Predicting ligand binding modes from neural networks trained on protein-ligand interaction fingerprints. *J Chem Inf Model*. 2013; 53:763–772. [PubMed: 23480697]
40. Xie L, Xie L, Bourne PE. A unified statistical model to support local sequence order independent similarity searching for ligand-binding sites and its application to genome-based drug discovery. *Bioinformatics*. 2009; 25:i305–312. [PubMed: 19478004]
41. Briedis KM, Starr A, Bourne PE. Analysis of the human kinome using methods including fold recognition reveals two novel kinases. *PLoS One*. 2008; 3:e1597. [PubMed: 18270584]
42. Xie L, Bourne PE. Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. *Proc Natl Acad Sci U S A*. 2008; 105:5441–5446. [PubMed: 18385384]

43. Konc J, Janezic D. Binding site comparison for function prediction and pharmaceutical discovery. *Curr Opin Struct Biol.* 2014; 25C:34–39. [PubMed: 24878342]
44. Xie L, Ge X, Tan H, Xie L, Zhang Y, Hart T, Yang X, Bourne PE. Towards structural systems pharmacology to study complex diseases and personalized medicine. *PLoS Comput Biol.* 2014; 10:e1003554. [PubMed: 24830652]
45. Nisius B, Sha F, Gohlke H. Structure-based computational analysis of protein binding sites for function and druggability prediction. *J Biotechnol.* 2012; 159:123–134. [PubMed: 22197384]
46. Xie L, Bourne PE. A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites. *BMC Bioinf.* 2007; 8(Suppl. 4):S9.
47. Xie L, Li J, Xie L, Bourne PE. Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. *PLoS Comput Biol.* 2009; 5:e1000387. [PubMed: 19436720]
48. Chang RL, Xie L, Xie L, Bourne PE, Palsson BO. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput Biol.* 2010; 6:e1000938. [PubMed: 20957118]
49. Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L, Bourne PE. Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput Biol.* 2009; 5:e1000423. [PubMed: 19578428]
50. Xie L, Xie L, Bourne PE. Structure-based systems biology for analyzing off-target binding. *Curr Opin Struct Biol.* 2011; 21:189–199. [PubMed: 21292475]
51. Garijo D, Kinnings S, Xie L, Xie L, Zhang Y, Bourne PE, Gil Y. Quantifying reproducibility in computational biology: the case of the tuberculosis drugome. *PLoS One.* 2013; 8:e80278. [PubMed: 24312207]
52. Durrant JD, Amaro RE, Xie L, Urbaniak MD, Ferguson MA, Haapalainen A, Chen Z, Di Guilmi AM, Wunder F, Bourne PE, McCammon JA. A multidimensional strategy to detect polypharmacological targets in the absence of structural and sequence homology. *PLoS Comput Biol.* 2010; 6:e1000648. [PubMed: 20098496]
53. Kinnings SL, Xie L, Fung KH, Jackson RM, Xie L, Bourne PE. The Mycobacterium tuberculosis drugome and its polypharmacological implications. *PLoS Comput Biol.* 2010; 6:e1000976. [PubMed: 21079673]
54. Xie L, Evangelidis T, Xie L, Bourne PE. Drug discovery using chemical systems biology: weak inhibition of multiple kinases may contribute to the anti-cancer effect of nelfinavir. *PLoS Comput Biol.* 2011; 7:e1002037. [PubMed: 21552547]
55. Xie L, Xie L, Kinnings SL, Bourne PE. Novel computational approaches to polypharmacology as a means to define responses to individual drugs. *Annu Rev Pharmacol Toxicol.* 2012; 52:361–379. [PubMed: 22017683]
56. Kinnings SL, Liu N, Tonge PJ, Jackson RM, Xie L, Bourne PE. A machine learning-based method to improve docking scoring functions and its application to drug repurposing. *J Chem Inf Model.* 2011; 51:408–419. [PubMed: 21291174]
57. Chang RL, Xie L, Bourne PE, Palsson BO. Antibacterial mechanisms identified through structural systems pharmacology. *BMC Syst Biol.* 2013; 7:102. [PubMed: 24112686]
58. Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technol.* 2004; 1:337–341.
59. Klebe G. Applying thermodynamic profiling in lead finding and optimization. *Nat Rev Drug Discovery.* 2015; 14:95–110. [PubMed: 25614222]
60. Constantine KL, Mueller L, Metzler WJ, McDonnell PA, Todderud G, Goldfarb V, Fan Y, Newitt JA, Kiefer SE, Gao M, Tortolani D, Vaccaro W, Tokarski J. Multiple and single binding modes of fragment-like kinase inhibitors revealed by molecular modeling, residue type-selective protonation, and nuclear overhauser effects. *J Med Chem.* 2008; 51:6225–6229. [PubMed: 18771253]
61. Chen H, Ma J, Li W, Eliseenkova AV, Xu C, Neubert TA, Miller WT, Mohammadi M. A molecular brake in the kinase hinge region regulates the activity of receptor tyrosine kinases. *Mol Cell.* 2007; 27:717–730. [PubMed: 17803937]

62. Carter TA, Wodicka LM, Shah NP, Velasco AM, Fabian MA, Treiber DK, Milanov ZV, Atteridge CE, Biggs WH 3rd, Edeen PT, Floyd M, Ford JM, Grotzfeld RM, Herrgard S, Insko DE, Mehta SA, Patel HK, Pao W, Sawyers CL, Varmus H, Zarrinkar PP, Lockhart DJ. Inhibition of drug-resistant mutants of ABL, KIT, and EGF receptor kinases. *Proc Natl Acad Sci U S A*. 2005; 102:11011–11016. [PubMed: 16046538]
63. Foda ZH, Seeliger MA. Kinase inhibitors: an allosteric addon. *Nat Chem Biol*. 2014; 10:796–797. [PubMed: 25195010]
64. Wu P, Nielsen TE, Clausen MH. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol Sci*. 2015; 36:422–439. [PubMed: 25975227]
65. Cadranel J, Ruppert AM, Beau-Faller M, Wislez M. Therapeutic strategy for advanced EGFR mutant non-small-cell lung carcinoma. *Crit Rev Oncol Hematol*. 2013; 88:477–493. [PubMed: 23911281]
66. Engel J, Richters A, Getlik M, Tomassi S, Keul M, Termathe M, Lategahn J, Becker C, Mayer-Wrangowski S, Grutter C, Uhlenbrock N, Krull J, Schaumann N, Eppmann S, Kibies P, Hoffgaard F, Heil J, Menninger S, Ortiz-Cuaran S, Heuckmann JM, Tinnefeld V, Zahedi RP, Sos ML, Schultz-Fademrecht C, Thomas RK, Kast SM, Rauh D. Targeting drug resistance in EGFR with covalent inhibitors: a structure-based design approach. *J Med Chem*. 2015; 58:6844–6863. [PubMed: 26275028]
67. Siegelin MD, Borczuk AC. Epidermal growth factor receptor mutations in lung adenocarcinoma. *Lab Invest*. 2014; 94:129–137. [PubMed: 24378644]
68. Greig SL. Osimertinib: First Global Approval. *Drugs*. 2016; 76:263–273. [PubMed: 26729184]
69. Walter AO, Sjin RT, Haringsma HJ, Ohashi K, Sun J, Lee K, Dubrovskiy A, Labenski M, Zhu Z, Wang Z, Sheets M, St Martin T, Karp R, van Kalken D, Chaturvedi P, Niu D, Nacht M, Petter RC, Westlin W, Lin K, Jaw-Tsai S, Raponi M, Van Dyke T, Etter J, Weaver Z, Pao W, Singh J, Simmons AD, Harding TC, Allen A. Discovery of a mutant-selective covalent inhibitor of EGFR that overcomes T790M-mediated resistance in NSCLC. *Cancer Discovery*. 2013; 3:1404–1415. [PubMed: 24065731]
70. Zhou W, Ercan D, Chen L, Yun CH, Li D, Capelletti M, Cortot AB, Chirieac L, Iacob RE, Padera R, Engen JR, Wong KK, Eck MJ, Gray NS, Janne PA. Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature*. 2009; 462:1070–1074. [PubMed: 20033049]
71. Adams JA. Kinetic and catalytic mechanisms of protein kinases. *Chem Rev*. 2001; 101:2271–2290. [PubMed: 11749373]
72. Bollag G, Tsai J, Zhang J, Zhang C, Ibrahim P, Nolop K, Hirth P. Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat Rev Drug Discovery*. 2012; 11:873–886. [PubMed: 23060265]
73. Shaw AT, Kim DW, Mehra R, Tan DS, Felip E, Chow LQ, Camidge DR, Vansteenkiste J, Sharma S, De Pas T, Riely GJ, Solomon BJ, Wolf J, Thomas M, Schuler M, Liu G, Santoro A, Lau YY, Goldwasser M, Boral AL, Engelman JA. Ceritinib in ALK-rearranged non-small-cell lung cancer. *N Engl J Med*. 2014; 370:1189–1197. [PubMed: 24670165]
74. Fabbro D. 25 years of small molecular weight kinase inhibitors: potentials and limitations. *Mol Pharmacol*. 2015; 87:766–775. [PubMed: 25549667]
75. Ohren JF, Chen H, Pavlovsky A, Whitehead C, Zhang E, Kuffa P, Yan C, McConnell P, Spessard C, Banotai C, Mueller WT, Delaney A, Omer C, Sebolt-Leopold J, Dudley DT, Leung IK, Flamme C, Warmus J, Kaufman M, Barrett S, Tecle H, Hasemann CA. Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. *Nat Struct Mol Biol*. 2004; 11:1192–1197. [PubMed: 15543157]
76. Huang WS, Metcalf CA, Sundaramoorthi R, Wang Y, Zou D, Thomas RM, Zhu X, Cai L, Wen D, Liu S, Romero J, Qi J, Chen I, Banda G, Lentini SP, Das S, Xu Q, Keats J, Wang F, Wardwell S, Ning Y, Snodgrass JT, Broudy MI, Russian K, Zhou T, Commodore L, Narasimhan NI, Mohemmad QK, Iulucci J, Rivera VM, Dalgarno DC, Sawyer TK, Clackson T, Shakespeare WC. Discovery of 3-[2-(imidazo[1,2-b]pyridazin-3-yl)ethynyl]-4-methyl-N-{4-[(4-methylpiperazin-1-yl)methyl]-3-(trifluoromethyl)phenyl}benzamide (AP24534), a potent, orally active pan-inhibitor of breakpoint cluster region-abelson (BCR-ABL) kinase including the T315I gatekeeper mutant. *J Med Chem*. 2010; 53:4701–4719. [PubMed: 20513156]

77. O'Hare T, Shakespeare WC, Zhu X, Eide CA, Rivera VM, Wang F, Adrian LT, Zhou T, Huang WS, Xu Q, Metcalf CA 3rd, Tyner JW, Loriaux MM, Corbin AS, Wardwell S, Ning Y, Keats JA, Wang Y, Sundaramoorthi R, Thomas M, Zhou D, Snodgrass J, Commodore L, Sawyer TK, Dalgarno DC, Deininger MW, Druker BJ, Clackson T. AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell*. 2009; 16:401–12. [PubMed: 19878872]
78. Solca F, Dahl G, Zoephel A, Bader G, Sanderson M, Klein C, Kraemer O, Himmelsbach F, Haaksma E, Adolf GR. Target binding properties and cellular activity of afatinib (BIBW 2992), an irreversible ErbB family blocker. *J Pharmacol Exp Ther*. 2012; 343:342–350. [PubMed: 22888144]
79. Pan Z, Scheerens H, Li SJ, Schultz BE, Sprengeler PA, Burrill LC, Mendonca RV, Sweeney MD, Scott KC, Grothaus PG, Jeffery DA, Spoerke JM, Honigberg LA, Young PR, Dalrymple SA, Palmer JT. Discovery of selective irreversible inhibitors for Bruton's tyrosine kinase. *ChemMedChem*. 2007; 2:58–61. [PubMed: 17154430]
80. Li D, Ambrogio L, Shimamura T, Kubo S, Takahashi M, Chirieac LR, Padera RF, Shapiro GI, Baum A, Himmelsbach F, Rettig WJ, Meyerson M, Solca F, Greulich H, Wong KK. BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. *Oncogene*. 2008; 27:4702–4711. [PubMed: 18408761]
81. [accessed Dec 21, 2015] Food and Drug Administration (FDA) approved drug products. <https://www.accessdata.fda.gov/scripts/cder/drugsatfda/>
82. Garuti L, Roberti M, Bottegoni G. Irreversible protein kinase inhibitors. *Curr Med Chem*. 2011; 18:2981–2994. [PubMed: 21651479]
83. Liu Q, Sabnis Y, Zhao Z, Zhang T, Buhrlage SJ, Jones LH, Gray NS. Developing irreversible inhibitors of the protein kinase cysteinome. *Chem Biol*. 2013; 20:146–159. [PubMed: 23438744]
84. Mi YJ, Liang YJ, Huang HB, Zhao HY, Wu CP, Wang F, Tao LY, Zhang CZ, Dai CL, Tiwari AK, Ma XX, To KK, Ambudkar SV, Chen ZS, Fu LW. Apatinib (YN968D1) reverses multidrug resistance by inhibiting the efflux function of multiple ATP-binding cassette transporters. *Cancer Res*. 2010; 70:7981–7991. [PubMed: 20876799]
85. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28:235–242. [PubMed: 10592235]
86. Salama AK, Kim KB. Trametinib (GSK1120212) in the treatment of melanoma. *Expert Opin Pharmacother*. 2013; 14:619–627. [PubMed: 23432625]
87. Liu Y, Gray NS. Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol*. 2006; 2:358–364. [PubMed: 16783341]
88. McTigue M, Murray BW, Chen JH, Deng YL, Solowiej J, Kania RS. Molecular conformations, interactions, and properties associated with drug efficiency and clinical performance among VEGFR TK inhibitors. *Proc Natl Acad Sci U S A*. 2012; 109:18281–18289. [PubMed: 22988103]
89. Pemovska T, Johnson E, Kontro M, Repasky GA, Chen J, Wells P, Cronin CN, McTigue M, Kallioniemi O, Porkka K, Murray BW, Wennerberg K. Axitinib effectively inhibits BCR-ABL1(T315I) with a distinct binding conformation. *Nature*. 2015; 519:102–105. [PubMed: 25686603]
90. [accessed Sep 22, 2015] Harvard Medical School (HMS) LINCS Database. <http://lincs.hms.harvard.edu/explore/>
91. Shmelkov E, Grigoryan A, Swetnam J, Xin J, Tivon D, Shmelkov SV, Cardozo T. Historeceptomic fingerprints for drug-like compounds. *Front Physiol*. 2015; 6:371. [PubMed: 26733872]
92. UniProt Consortium. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res*. 2013; 41:D43–D47. [PubMed: 23161681]
93. UniProt Consortium. [accessed May 15, 2015] Human and mouse protein kinases: classification and index. <http://www.uniprot.org/docs/pkinfam.txt>
94. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*. 1996; 261:470–489. [PubMed: 8780787]
95. Kellenberger E, Muller P, Schalon C, Bret G, Foata N, Rognan D. sc-PDB: an annotated database of druggable binding sites from the Protein Data Bank. *J Chem Inf Model*. 2006; 46:717–727. [PubMed: 16563002]

96. Radifar M, Yuniarti N, Istyastono EP. PyPLIF: python-based protein-ligand interaction fingerprinting. *Bioinformatics*. 2013; 9:325–328.
97. Uchikoga N, Hirokawa T. Analysis of protein-protein docking decoys using interaction fingerprints: application to the reconstruction of CaM-ligand complexes. *BMC Bioinf*. 2010; 11:236.
98. Jaccard P. The distribution of the flora in the alpine zone. *New Phytol*. 1912; 11:37–50.
99. Hartigan JA, Wong MA. Algorithm AS 136: A K-means clustering algorithm. *Appl Stat*. 1979; 28:100–108.
100. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model*. 2010; 50:742–754. [PubMed: 20426451]
101. Hu Y, Lounkine E, Bajorath J. Improving the search performance of extended connectivity fingerprints through activity-oriented feature filtering and application of a bit-density-dependent similarity function. *ChemMedChem*. 2009; 4:540–548. [PubMed: 19263458]
102. JChem 6.2.0. ChemAxon; 2014. JChem Base was used for structure searching and chemical database access and management. <http://www.chemaxon.com>
103. Weill N, Rognan D. Alignment-free ultra-high-throughput comparison of druggable protein-ligand binding sites. *J Chem Inf Model*. 2010; 50:123–135. [PubMed: 20058856]

Abbreviations Used

IFP	interaction fingerprint
Fs-IFP	functional site interaction fingerprint
FDA	U.S. Food and Drug Administration
NSCLC	non-small-cell lung cancer
ALK	anaplastic lymphoma kinase
PoC	percent of control
HPK-SD	human protein kinase structural data set
TC	Tanimoto coefficient
ECFP	extended-connectivity fingerprint

Figure 1.

Flowchart describing the Fs-IFP method. (a) The protein kinases that have released apo and holo structures of catalytic domains are marked with blue and red filled circles, respectively, in the evolutionary tree.¹ (b) PDB entry 3BYU, which is used as a reference structure for all kinases. (c). Superimposed binding site. Each row represents a sequence-order-independent motif comprising the binding pocket for a given ligand-bound structure. Each column represents the corresponding amino acid residues at the same spatial position of the binding pocket for all ligand-bound structures. The D-NULL mark means there is no corresponding residue at the specific position of the given ligand-bound structures. (d) Encoded binding site using the Fs-IFP approach. Each row represents the Fs-IFP bit-strings for one given ligand-bound structure. Each column including 7 bits represents an Fs-IFP code of one residue of the binding site by using seven types of interactions. The same column means that the residue matches the same position in 2383 superposed binding pockets. A bar is used if there is no typical interaction between the specific amino acid and the ligand in the given ligand-bound structure. The bit is marked as “1” if the specific type of interaction exists, otherwise marked as “0” if the type of interaction does not exist between the specific amino acid and the ligand.

Figure 2. Distribution of the ligand characteristics in 2383 cocrystallized kinase structures: (a) distribution of the top 20 frequently occurring ligands; (b) distribution of the molecular weight of the ligands in all cocrystallized kinase structures.

Figure 3. Distribution of Tanimoto coefficient (TC) value for the fingerprints similarity of the ligands (ligand-FP) found in HPK-SD (red), kinase pockets (kinase-FP) of the HPK-SD (blue), and the kinase–ligand interaction fingerprints (Fs-IFP) of the kinase–ligand complexes (green).

Figure 4.

Correlation of the Tanimoto coefficient (TC) for ligand FP, kinase-pocket FP, and kinase-ligand IFP similarity using a violin plot: (a, b) correlation of the TC of the ligand FP similarity and the kinase pocket FP similarity; (c, e) correlation of the TC of the ligand FP similarity and the kinase-ligand IFP similarity; (d, f) correlation of the TC of the kinase pocket FP similarity and the kinase-ligand IFP similarity. In all panels, the width of the violin box is proportional to the number of TCs. The black bars range from 25% to 75%, and the white points on the bars represent the medians. Distributions (%) of ligand-FP, kinase-FP, and Fs-IFP in the corresponding bin of similarity are shown at the upper portion of every panel.

Figure 5.

Distribution of interaction types at the kinase–ligand binding interface: (a) distribution of the contribution of six interaction types at the kinase–ligand binding interface; (b) distribution of occurrences of the six interaction types for each residue location at the binding interface. The *y*-axis is the number of interaction occurrences for the corresponding amino acids given along the *x*-axis.

Figure 6. Correlation between interactions and the amino acid conservation at each amino acid position constituting the binding interface: (a) distribution of interaction occurrence for each amino acid position; (b) the corresponding distribution in (a) mapped to the template structure; (c) conservation of amino acids at the binding site; (d) correlation between residue conservation and the distribution of interactions.

Figure 7.
Rectangular tree layout of the Fs-IFP data set for HPK-SD. Five distinct clusters were found. Cluster5 was further divided into 5 subclusters: cluster 5a–d.

Figure 8. Binding modes for each cluster 1–4 and 5a–d based on complete linkage clustering: (a–h) for cluster 1–4 and 5a–d using PDB entries 4ITJ, 2A5U, 3PXF, 1H8F, 3PRZ, 1BYG, 1ZZ2, and 3Q4C as the representative structure, respectively; (i) site locations for the binding regions for the eight clusters.

Figure 9.

Correlation between Fs-IFP and experimentally determined kinase binding profiles. The curves were obtained by using fit first-order exponential decay. (a) Correlation of the pairwise difference of the K_d values and pairwise similarity of Fs-IFPs. (b) Correlation of the pairwise difference of the PoC values and pairwise similarity of Fs-IFPs.

Table 1
Distributions of the Binding Modes for FDA-Approved Kinase-Targeted Drugs^a

approved inhibitor	approved year	PDB entry	involved clusters for the binding modes							inhibitor of type
			1	2	3	4	5a	5b	5c	5d
IMATINIB	2001	4BKJ								II
GEFITINIB	2003	4I22								I
ERLOTINIB	2004	4HJO								I
SORAFENIB	2005	4ASD								II
DASATINIB	2006	3QLG								I
SUNITINIB	2006	2Y7J								I
NILOTINIB	2007	3GP0								II
LAPATINIB	2007	1XKK								I
PAZOPANIB	2009	2X9F								I
VANDETANIB	2011	2IVU								I
RUXOLITINIB	2011	3VS7								I
CRIZOTINIB	2011	3ZBF								I
VEMURAFENIB	2011	3OG7								I
TOFACITINIB	2012	3LXN								I
BOSUTINIB	2012	4OTW								I
CABOZANTINIB	2012	3U6J								II
PONATINIB	2012	4C8B								II
REGORAFENIB	2012	3WZE								II
AXITINIB	2012	4AGC								I
IBRUTINIB	2013	4IFG								I
AFATINIB	2013	4G5J								I
TRAMETINIB	2013	3PP1								III
DABRAFENIB	2013	4CQE								I
NINTEDANIB	2014	3C7Q								I
IDELALISIB	2014	4XE0								I
CERITINIB	2014	4MKC								I
PALBOCICLIB	2015	2EUF								I

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

approved inhibitor	approved year	PDB entry	involved clusters for the binding modes							inhibitor of type
			1	2	3	4	5a	5b	5c	5d
LENVATINIB	2015	3WZD								II
COBIMETINIB	2015	4AN2								III
OSIMERTINIB	2015	4ZAU								I

^aThe “ ” means the interaction exists in the corresponding regions of the binding pocket (Figure 8i).