Harnessing systematic protein-ligand interaction fingerprints for drug discovery

Zheng Zhao* and Philip E. Bourne*

School of Data Science and Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia 22904, United States of America

*Corresponding author

Email:zz7r@virginia.edu (ZZ) and peb6a@virginia.edu (PEB);

Tel: 1-(434)-982-2600 (ZZ) and 1-(434)-982-6867 (PEB)

Keywords

Interaction fingerprints; proteome-wide drug discovery; mechanism of drug resistance; antivirus drug discovery; structural systems biology; polypharmacology; allosteric inhibitors; covalent inhibitors

Research Highlights

- Review of protein-ligand interaction fingerprints for drug discovery
- Exploring polypharmacology
- Revealing drug-resistance mechanisms
- Design of kinase allosteric inhibitors and covalent inhibitors

Abstract

Determining protein-ligand interaction characteristics and mechanisms is critical in the drug discovery process. Here we review recent progress and successful applications of a systematic protein-ligand interaction fingerprint (IFP) approach for investigating proteome-wide protein-ligand interactions for drug development. Specifically, we review the use of this IFP approach for revealing polypharmacology across the whole kinome, predicting promising targets from which to design allosteric inhibitors and covalent kinase inhibitors, uncovering the binding mechanisms of drugs of interest, and demonstrating resistant mechanisms of specific drugs. Together, we demonstrate that the IFP strategy is efficient and practical for drug design research and development in the current era of big data.

Introduction

Drug discovery is a time-consuming, costly, and complicated process. For this reason, in silico drug design has been a routine component used to decrease the cost and shorten the time period to drug launch [1,2]. Success supports the advantage and efficiency of computational drug discovery strategies. For example, Table 1 lists 26 FDA-approved drugs, which were designed using computer-aided drug discovery approaches. [3-5] Most recently, advances in biomedical data science including the advent of machine/deep learning applications have made computational pharmacology even more important in facilitating drug discovery [6,7].

In the early stage of novel drug development, one of the most important tasks is to discover promising lead molecules able to withstand the subsequent rigor of clinical trials[8]. Computer-aided drug discovery can economically narrow the choices from a myriad of compound screening databases and reveal the relevant binding characteristics, used to explore libraries of drug-like compounds [9]. Computer-aided drug screening can be loosely divided into three classes: receptor-based, ligand-based, and protein-ligand-interaction-based methods [10,11]. A receptor-based method uses the 3D structure of receptors, where the structural flexibility, the size and shape of the binding pocket, and the properties of binding affinity are utilized to facilitate drug discovery. The ligand-based method is an approach that relies on the knowledge of ligand molecules and is used in cases where receptor 3D information is not available[12,13]. Protein-ligand-interaction-based methods combine the two [14,15]. For example, in the process of computer-aided virtual screening, properties of ligand similarity [16] and molecular drug-like rules [17] such as molecular weight and the number of hydrogen bonds, are often applied to find lead candidates. With the advent of structure-based biomedical data science, a large number of protein structures and bioactive molecules have been compiled [18-21], providing a substantial basis for the development of protein-ligand interaction-

based approaches, such as pharmacophore-based and interaction fingerprint-based methods [22-24]. A pharmacophore-based scheme trains a receptor-ligand interaction model as the representative based on a set of bioactive ligands interacting with a given receptor [24]. Given the representative binding pharmacophores, the virtual screening can be carried out for any candidate compound. By comparison, the scheme based on the protein-ligand interaction fingerprint (IFP) [22] is to encode the details of protein-ligand interaction into a binary string using a set of predefined interaction types (van der Waals, π–π stacking, hydrogen bond, electrostatic interaction, and so on) and criteria [22]. Using this encoded binary string, the binding features, given a ligand-bound complex, can be captured in great detail and easily manipulated. Since Deng et al. (2004) presented the structural interaction fingerprint (S-IFP) method [22], protein-ligand IFP strategies have been improved and applied to virtual screening [25-28], post-processing of dock poses [29,30], scoring functions [31-33], and so on [23,34]. However, these applications have been restricted to use on the same receptor or a cluster of highly homologous protein structures [25].

To explore proteome-scale polypharmacology for personalized drug discovery, Zhao et al. provided a scheme, Fs-IFP [35], combining structural systems pharmacology [36] and a structural interaction fingerprint strategy [22]. For Fs-IFP, first, the binding pockets of receptors are determined for the whole proteome. Second, every known complex for each binding pocket is encoded into a binary string using the interaction fingerprint strategy[22]. Finally, comparable interaction fingerprints are obtained by extracting the aligned "pocketome" and the corresponding fingerprint strings. Utilizing this Fs-IFP scheme, the authors have extended the structural interaction fingerprint strategy to multiple protein families, regardless of whether they are highly homologous or distant from each other, thus uncovering characteristics and nuances of binding. We introduce the specifics of the IFP and Fs-IFP protocols and provide applications of the method to rational

drug discovery. Finally, we discuss the prospective applications of this strategy by merging biomedical data science and machine/deep learning models.

An overview of IFP methods

Different descriptor schemes have been used to characterize protein-ligand interactions[37-39]. For example, an element-based descriptor scheme[38,40], in which the protein-ligand interaction is described using a combination of the direct interacting atom pairs from protein and ligand separately based on the element types, e.g., C-N or C-O. Thus, every position of the fingerprint in this scheme represents a paired element type to describe the given protein-ligand interaction. Atom types can be described more specifically based on the protein environments. For example, the SYBYL scheme[38] classifies the atoms into distinct subtypes based on their chemical atom properties. For example, there are 5 subtypes of carbon atom types: C1 (sp carbon), C2 (sp2 carbon), C3 (sp3 carbon), Car (aromatic carbon), and Ccat (carbocation), which leads to 25 types of different C-C interaction descriptors instead of a single type.

One popular descriptor scheme is based on structural interaction fingerprints (IFPs)[22,29]. In 2004, Deng et al. proposed an IFP scheme, known as S-IFP, to structurally characterize protein-ligand interactions [22]. The atom types were defined using the SMARTs definiation[41]. The S-IFP approach encodes protein-ligand interactions in a 1D binary string. Specifically, each residue, comprising the binding pocket, is encoded into a 7-digit binary substring (Figure 1a) using pre-defined geometric criteria. The 7-digit binary substring describes the contributions of each amino acid to the protein-ligand interaction and represents 7 types of interaction, whether existing or not, including (1) contact with the ligand, (2) involving main-chain atoms, (3) involving side-chain atoms, (4) polar interactions, (5) nonpolar interactions, (6) hydrogen bond interactions (amino acid

as acceptor), and (7) hydrogen bond interactions (amino acid as donor). For example, 1000010, where "1" indicates the interaction exists, and "0" indicates no interaction. Following the S-IFP method, a few variations have been developed such as r-IFP [42], and w-IFP [43]. The r-IFP method also encodes the r-group or core-fragment of the ligand bearing the interactions with the specific amino acid of the target [42]. The w-IFP method [43] encodes the interactions into IFPs with features of relative importance [43]. Based on S-IFP, two additional interactions (aromatic and charged) were added in Stefan et al.'s method [44] creating a 9-digit binary substring describing the interactions between each residue and the ligand [44].

Rognan et al. encoded the protein-ligand interactions into a 1D binary IFP string with an array of 11-bit substrings (Figure 1b), which describe how each amino acid interacts with the ligand. Each amino acid in the binding pocket is encoded into one 11-bit substring corresponding to 11 types of interaction: i.e. (i) hydrophobic interaction, (ii) aromatic interaction (face-to-face), (iii) aromatic interaction (edge-to-face), (iv) hydrogen interaction (protein atom as acceptor), (v) hydrogen interaction (protein atom as donor), (vi) ionic interaction (protein atom with positive charge), (vii) ionic interaction (protein atom with negative charge), (viii) weak hydrogen interaction (protein atom as acceptor), (ix) weak hydrogen interaction (protein atom as donor), (x) π-cation interaction, and (xi) metal ionic interaction with the ligand. A variety of software is available for encoding the ligand-binding interaction fingerprints, such as PyPLIF [30] and IChem [45]. PyPLIF encodes the given protein-ligand complex into an IFP string using 7 types of interaction (i-vii), however, IChem can encode the IFPs for every residue using 11 types of interaction (i-xi).

Another class of advanced IFPs are the extended connectivity fingerprints (ECFPs)[46], which encode all of the local interactions between the ligand and the close protein atoms within the binding site[47-50]. Each ligand-protein atom pair is considered an interaction if the interatomic

distance is less than a predefined threshold, for example, 4.5 Å[48]. Every atom/bond type is defined within the ECFP and used as an identifier[46]. Currently, the ECFP-based IFP scheme has been applied to virtual screening and binding affinity predictions[47,48,50]. Compared to the early S-IFP method, the ECFP-based IFPs are protein fragment-based and encoded by iteratively hashing the interacting atom pairs into integers and then folding the virtual strings,[47,50] whereas S-IFP-based IFPs are residue-based and encoded using the pre-defined 7-type geometric rules[29].

Recently, the Fs-IFP method was developed for proteome-wide drug discovery [35]. The motivation being profiling polypharmacology is necessary not only to reveal primary targets but also off-targets [51,52]. Fs-IFP extends the IFP approach to the whole structural proteome by combining the IFP approach with sequence order-independent binding-site alignment to ultimately determine the polypharmacology of select inhibitors. The Fs-IFP method consists of four steps (**Figure 1c**). Step 1 is to construct the structural dataset including all the ligand-bound protein complexes, all of which can be downloaded from the Protein Data Bank (PDB) [18] or a PDB-binding database [19]. Step 2 is to align all of the binding sites using a sequence order-independent binding-site alignment method, for example, SMAP [53], which provides a sequence-order-independent secondary-structure alignment and outputs a corresponding matched-residues matrix. Using the binding site of the study target as a template, all other binding sites across the structural proteome can be aligned and presented as an alignment matrix of amino acids. In this residue matrix, every row of amino acids comprising the binding sites is extracted for every complex. Every column represents the aligned residues occurring in the same position within all the binding sites. Step 3 encodes all of the protein-ligand interaction types in every complex into a 1D array of bit strings using off-the-shelf tools such as IChem [45]. Specifically, each residue in the binding site of each complex is encoded into a 7-bit substring [29]. The 7-bit substring represents 7 types of interaction

(i-vii) as described above between an amino acid and the corresponding ligand within the binding site. Thus, the 1D array of bit strings for every complex is composed of a series of 7-bit substrings [30,54]. Step 4 combines the aligned residue matrix of Step 2 with the 1D array of bit strings of every complex from Step 3. In other words, every amino acid in the residue matrix is replaced using the corresponding 7-bit interaction fingerprint for every binding site.

Overview of drug discovery using the Fs-IFP approach

Revealing patterns of ligand binding across the proteome

Polypharmacology makes it challenging to achieve the desired bioactivity and selectivity in targeted drug discovery [55,56]. The Fs-IFP approach provides a practical means of exploring polypharmacology across the proteome. For example, we explored the polypharmacology and binding patterns of kinase inhibitors across the whole kinome [57] using the Fs-IFP approach. The human kinase family comprises more than 500 kinases many being drug targets for treating different diseases. It has been challenging to design specific kinase inhibitors because all kinases have a common ATP binding pocket [58]. Thus exploring the polypharmacology of kinase inhibitors is mandatory in any assay seeking to discover new kinase inhibitors[52]. In the study in question we first collected 2383 complex structures from 208 kinases to use as a kinase dataset [35]. Then, a set of comparable Fs-IFP-encoded interaction fingerprints from all kinase-ligand complexes were obtained following the Fs-IFP protocol (i.e., binding-site alignments, encoding the IFP of every complex, and systematic Fs-IFP fingerprints, Figure 1c). Subsequetly, the binding characteristics of diverse inhibitors in the ATP binding pocket and/or its vicinity could be classified across the whole human kinome. The kinase binding patterns could be clustered into 5 classes with corresponding binding features and positions (**Figure** 2a).

The largest group of kinase inhibitors are of type Cluster5, which occupy the ATP binding pocket (Figure 2b). This type of ATP-competitive inhibitor forms conserved interactions with the residues distributed in the Hinge region, β3, and β7 in the vicinity of the adenine moiety of ATP (**Figure** 2b). To achieve greater binding affinity than ATP, kinase inhibitors often occupy the Cluster5 region but also extend into other proximal clusters as shown in Figure 2a (Cluster1-4) [35]. For example, Imatinib [59], a Type-II tyrosine kinase inhibitor used to treat several cancers, not only occupies the ATP-binding region (Cluster5) but also extends into the hydrophobic area close to the gatekeeper (Cluster1) and the region between the roof of the β3 and DFG tripeptide (Cluster2, Figure 2a), the allosteric area (Cluster3, Figure 2a), and the area between the activation loop and C-Helix (Cluster4, Figure 2a). Similarly, the Type-I kinase drug Lapatinib (Figure 2c), an EGFR/HER2 inhibitor to treat breast cancer and other solid tumors[60], mainly binds to the ATP binding site (Cluster5) but also extends into the hydrophobic region (Cluster1, Figure 2c). In contrast, the back cleft of the binding pocket close to the ATP binding site is used to design allosteric inhibitors. For example, the Type-III kinase drug Cobimetinib[61] (Figure 2d), a MEK inhibitor to treat melanoma, binds mainly to the back cleft of the binding pockets which includes the main allosteric region (Cluster3) and the vicinal areas (Cluster1 and 2). Further, the piperidine group of Cobimetinib extends into the distal region of the ATP binding site (Cluster5).

Another example involved antivirus drug discovery revealing the binding modes of 47 RNA viruses[62]; pertinent during a COVID-19-induced pandemic. We focused on RNA-dependent RNA polymerase found in 47 distinct RNA viruses, including SARS-CoV-2, obtaining the binding sites for further drug screening[62]. Using the same Fs-IFP approach, we obtained the features of the binding pockets for the 47 RNA viruses, classifying them into four classes. Virtual screening was then undertaken [63].

Uncovering specific interaction features for designing precise inhibitors

Revealing specific intermolecular atomic interactions is an obvious advantage of interaction fingerprints [54]. For example, when an electrostatic interaction occurs between an amino acid and an inhibitor within a binding site we can determine more detailed information about which atoms are involved in the interaction and how the interaction is formed within the binding site. In practice, Zhao et al. utilized the Fs-IFP approach to determine which cysteines are available across the human kinome to facilitate the discovery of covalent kinase inhibitors, knowing that covalent kinase inhibitors substantially improve the binding affinity and selectivity across the whole kinome [64,65].

To do so, the authors first collected 1599 complex structures belonging to 169 kinases that have at least one cysteine residue located within the binding sites. The authors then analyzed the interaction details between all these cysteines and their corresponding ligands. The analysis revealed that cysteines exhibit two kinds of interaction: hydrogen-bond interactions (21.4%) from the backbone atoms N and O and hydrophobic interactions (78.6%) from the other atoms: C, CA, CB, and SG (**Figure** 3a-b). Combined with the calculation of the potential energy surface between the cysteine thiol group and the warhead using *ab initio* DFT force fields, the authors demonstrated the orientation and reactivity of the thiol group for every cysteine. The authors verified the top 5 easily-available regions: the roof region of β3, P-loop area, front-pocket area, the catalytic loop (called: Catalytic-2), and a position near the DFG peptide (called DFG-3) across the human kinome (**Figure** 3c). These insights into cysteine-related covalent reactivity enable the design and discovery of prospective covalent kinase inhibitors.

As of April 2022, more than 70 small-molecule kinase drugs have been approved by the FDA including eight covalent drugs[66,67]; 5 target EGFR and 3 target BTK (**Table** 2). All of these drugs

form a covalent interaction with their respective targets in the front-pocket (Figure 3c). For example, Osimertinib[68], an EGFR kinase inhibitor to treat patients with advanced T790M-mutation-positive NSCLC[69,70], forms two hydrogen bonds with Met793 in the Hinge region and has an electrophilic acrylamide group that forms an irreversible covalent interaction with Cys797 in the front pocket of the ATP binding site (Figure 3d), leading to irreversible inhibition of the T790M-mutant EGFR with a nanomolar-level IC50[71]. Given the abundance of cysteines around the binding sites of kinase domains[65], covalent kinase inhibitor development is promising and will likely attract further attention[55].

Developing novel inhibitors across the proteome

Knowledge-based drug discovery is another application illustrating the potential of the Fs-IFP approach. Here we used the Fs-IFP approach to discover allosteric kinase inhibitors [72]. Currently, different types of kinase inhibitors have been reported and occupy different areas of the binding pockets[66,73]. Type-I kinase inhibitors, also called ATP-competitive kinase inhibitors, occupy ATP binding areas located in the front cleft of the kinase binding site, specifically the space between the Hinge region, hydrophobic pocket, P-loop, and DFG peptide (Figure 4a)[58]. Type-II kinase inhibitors occupy the front cleft where ATP binds, but also extend into the nearby allosteric region (Figure 4b). The allosteric region is not so conserved as the ATP-binding pocket,[55,74], thus, making exploiting kinase allosteric regions is one promising strategy for achieving desirable selectivity across the whole kinome[75]. Allosteric kinase inhibitors targeting MEK[72], P38α[76], BRAF[77], and EGFR[78] have been reported with most focus on mitogen-activated protein kinase (MEK). Indeed, as of now, all of the approved allosteric drugs are MEK-targeted[73]. Zhao et al. sought to explore the MEK-allosteric-inhibitor binding features intent on designing allosteric kinase inhibitors against other kinases besides MEK.

First, the authors studied the binding characteristics of all MEK-allosteric-inhibitor complexes from 29 PDB structures and their associated conformations from 1.2μs molecular dynamics simulation[72]. Then, binding characteristics were described using the Fs-IFP approach combined with pharmacophore modeling. The authors confirmed that MEK allosteric inhibitors always form two conserved interactions with residues S212 and K97 (**Figure** 4c). Moreover, the authors found that all conformations from the MEK MD trajectories contained a conserved short helix within the activation loop (**Figure** 4c). The PDB structures of P38α and BRAF (PDB ids 2yix, 4pp7, and 4wo5) also contained a short helix within the same segment of the activation loop[72]. Based on these binding characteristics, the authors predicted which kinase targets hold promise for allosteric inhibitor design (**Figure** 4d). From those predictions, of the top 15 kinase targets, 10 including MEK belong to the STE group, 3 (MAST 1-3) from the AGC group, 1 (JAK3) from the TK group, and 1 (NRBP1) from the Other group, indicating promising targets in the search for Type-III allosteric kinase inhibitors.

Demonstrating drug-resistant mechanisms

Targeted drug therapy is playing an increasingly important role in cancer treatment. For instance, lung cancer is one of the most common cancers worldwide, with 80-85% being non-small cell lung cancer (NSCLC) with 3-5% of those patients having gene fusions of anaplastic lymphoma kinase (ALK). Thus, targeting ALK is important in treating NSCLC[79]. However, due to acquired genetic mutations, drug resistance weakens the efficacy of anti-cancer drugs. Thus, revealing the mechanisms of drug resistance is crucial in facilitating the development of next-generation anticancer drugs. Here, we used the Fs-IFP scheme combined with the investigation of binding free energy surfaces to determine the drug's mechanism of action[80], facilitating the development of next-generation anti-cancer drugs.

Specifically, Crizotinib is a first-generation ALK drug to treat NSCLC, but the gatekeeper mutation (L1196) of ALK impacts the treatment of NSCLC in the clinic. We first aligned the Crizotinib-binding ALK PDB structures before and after L1196M mutation. The alignment of complexed structures shows that Crizotinib is a Type-I ALK inhibitor and that binding modes are similar before and after L1196M mutation (Figure 5a). Therefore, we explored the drug binding mechanism before and after residue mutations by calculating binding free energy surfaces. Subsequently, the atom-level Fs-IFPs were obtained for every conformation to investigate the differences in drug binding before and after mutation. Upon Fs-IFP analysis, the mechanism of L1196M-induced drug resistance was revealed. Crizotinib has significantly weaker interactions with A1123, A1125, and G1226 of the P-loop after L1196M mutation (**Figure** 5c). Thus, drug resistance of Crizotinib is associated with the change of residue interactions within a segment of the P-loop. Subsequently, using the same Fs-IFP method, we explored the binding mechanism of Ceritinib, which is a second-generation drug designed to overcome the L1196M mutation (**Figure** 5b). Ceritinib is also a Type-I ALK inhibitor and we found that Ceritinib overcomes the L1196M mutation by strengthening the corresponding interactions with the P-loop residues (**Figure** 5d). Given this example, the systematic application of the Fs-IFP method combined with umbrella sampling and free energy calculations is a promising approach for revealing the mechanisms underpinning drug resistance.

Summary and discussion

The Fs-IFP approach shows promise in drug design and discovery[35,62,64,72,80,81]. Specifically, the approach encodes the interaction features of any given protein-ligand complex into a bit string, facilitating large-scale data analyses. Furthermore, the binding-site alignments are based on a

sequence-order-independent structure comparison method, which allows us to explore similar targets but with different sequences across the proteome[82]. The comparable binary IFPs, based on the matched residues within the binding sites, offer a convenient means to analyze binding modes and train machine/deep learning models[32,38].

Drug resistance is a major limitation in which the efficacy of targeted drug therapy is significantly attenuated in clinical trials and beyond. Uncovering the different protein-ligand interactions of wild-type and mutants at the atomic level provides a pragmatic strategy for determining drug binding and/or drug-resistant mechanisms. Using interaction fingerprints before and after the conferral of drug resistance is an effective approach for developing next-generation anti-resistant drugs through a combination of the Fs-IFP approach combined with other approaches, such as free energy surface calculations.

Current drug discovery requires analysis of complicated drug-target-disease interaction networks and is thus omics in scale. The Fs-IFP approach provides omics level data through aligned interaction fingerprints revealing details of target-drug interaction networks. As illustrated here, for kinase drug discovery, selectivity across the whole kinome remains challenging even though huge advances have been made leading to more than 70 FDA-approved kinase-targeted drugs [73]. The next challenge would seem to be combining comparable fingerprints via the Fs-IFP approach with machine/deep learning models to predict polypharmacology. In so doing, we anticipate that early-stage drug discovery will be faster and more effective.

It is worth noting that the S-IFP-related approaches, including Fs-IFP, rely on pre-existing protein-ligand complex resources. The quality and quantity of available protein-ligand complexes directly effects the analysis and application of structure-activity relationships for given targets. With the development of structural biology technologies, such as Cryo-EM[83] and AlphaFold2[84],

more protein structures will be available to support the application of the IFP method. A further limitation of S-IFP methods is how to accurately detect and encode the protein-ligand interactions. Currently, detecting whether or not protein-ligand interaction patterns exist is based on pre-defined geometric rules. The pre-defined geometric criteria limit means some interaction types are not counted, such as metal interactions. Pre-defining more interaction patterns is needed. Recently an extended connectivity fingerprint (ECFP) has been applied to encode all atom/bond interaction types based on every pair of interatomic interactions between ligand and protein substructures[47-51]. The ECFP-based IFPs are costly to apply to determine polypharmacology across the structural proteome due to their enormous fingerprint sizes[47]. Another geometric limitation of S-IFP-related methods is the geometric rules to pre-define the hydrogen-bond interaction, electrostatic interaction, and so on.[29] For example, using a Euclidean distance ≤ 4.0 Å as the threshold to detect intermolecular ionic interactions[29]. Simplistic, yet hard cut-off rules may ignore some marginal interactions. This is offset to some extent by undertaking molecular dynamics (MD) simulation to provide flexible ligand-binding features[80]. By collecting all the interaction fingerprints of every conformation along with the MD trajectories, the IFP is more robust since minor interactions around the cut-off boundary will be detected due to the flexible protein-ligand interactions[80].

Acknowledgment

We thank Haiyan Liu (USTC), Lei Xie (CUNY), Qingsong Liu (HMFL), Li Xie, Spencer Bliven, Chunxiao Bi, Eli J. Draizen, Cameron Mura, and Philippe Youkharibache for many helpful discussions over many years, and Didier Rognan (CNRS) and Ajay N. Jain (BioPharmics LLC) for software help. This work was partly supported by the NIH intramural program and subsequently the University of Virginia (PEB).

Reference

1. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. Mar 2010;9(3):203-214.

2. Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. *Chem Rev*. Sep 25 2019;119(18):10520-10594.

3. Talele TT, Khedkar SA, Rigby AC. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Curr Top Med Chem*. 2010;10(1):127-141.

4. Bisht N, Singh BK. Role of computer-aided drug design in drug development and drug discovery. *Int J Pharm Sci Res*. 2018;9(4):1405-1415.

5. Athanasiou C, Cournia Z. From Computers to Bedside: Computational Chemistry Contributing to FDA Approval. *Biomolecular Simulations in Structure-Based Drug Discovery*. 2018:163-203. *Methods and Principles in Medicinal Chemistry*.

6. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. Jun 2019;18(6):463-477.

7. Zhao Z, E. Bourne P. Using the Structural Kinome to Systematize Kinase Drug Discovery. *Protein Kinases - Promising Targets for Anticancer Drug Research*. 1st ed. IntechOpen; 2021, DOI: 10.5772/intechopen.100109. *Biochemistry*.

8. Bajorath J. Computer-aided drug discovery. *F1000Res*. 2015;4

9. Sisay MT, Peltason L, Bajorath J. Structural interpretation of activity cliffs revealed by systematic analysis of structure-activity relationships in analog series. *J Chem Inf Model*. Oct 2009;49(10):2179-2189.

10. Taft CA, Da Silva VB, Da Silva CH. Current topics in computer-aided drug design. *J Pharm Sci*. Mar 2008;97(3):1089-1098.

11. Macalino SJ, Gosu V, Hong S, Choi S. Role of computer-aided drug design in modern drug discovery. *Arch Pharm Res*. Sep 2015;38(9):1686-1701.

12. Johnson MA, Maggiora GM. *Concepts and applications of molecular similarity*. 6th ed. New York: Wiley; 1990.

13. Vuorinen A, Engeli R, Meyer A, Bachmann F, Griesser UJ, Schuster D, et al. Ligand-based pharmacophore modeling and virtual screening for the discovery of novel 17beta-hydroxysteroid dehydrogenase 2 inhibitors. *J Med Chem*. Jul 24 2014;57(14):5995-6007.

14. Tan L, Geppert H, Sisay MT, Gutschow M, Bajorath J. Integrating structure- and ligand-based virtual screening: comparison of individual, parallel, and fused molecular docking and similarity search calculations on multiple targets. *ChemMedChem*. Oct 2008;3(10):1566-1571.

15. Vazquez J, Lopez M, Gibert E, Herrero E, Luque FJ. Merging Ligand-Based and Structure-Based Methods in Drug Discovery: An Overview of Combined Virtual Screening Approaches. *Molecules*. Oct 15 2020;25(20)

16. Nikolova N, Jaworska J. Approaches to Measure Chemical Similarity - a Review. *QSAR Comb Sci*. 2003;22(9-10):1006-1026.

17. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 2001;46(1-3):3-26.

18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. Jan 1 2000;28(1):235-242.

19. Wang R, Fang X, Lu Y, Wang S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J Med Chem*. Jun 3 2004;47(12):2977-2980.

20. 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*. Jan 23 2014;57(2):249-277.

21. Desaphy J, Bret G, Rognan D, Kellenberger E. sc-PDB: a 3D-database of ligandable binding sites--10 years on. *Nucleic Acids Res*. Jan 2015;43(Database issue):D399-404.

22. Deng Z, Chuaqui C, Singh J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J Med Chem*. Jan 15 2004;47(2):337-344.

23. Tan L, Batista J, Bajorath J. Computational methodologies for compound database searching that utilize experimental protein-ligand interaction information. *Chem Biol Drug Des*. Sep 1 2010;76(3):191-200.

24. Voet A, Qing X, Lee XY, De Raeymaecker J, Tame J, Zhang K, et al. Pharmacophore modeling: advances, limitations, and current utility in drug discovery. *J Recept Ligand Channel Res*. Nov 11 2014;7:81-92.

25. Chuaqui C, Deng Z, Singh J. Interaction profiles of protein kinase-inhibitor complexes and their application to virtual screening. *J Med Chem*. Jan 13 2005;48(1):121-133.

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*. Jan 15 2014;24(2):580-585.

27. 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*. Mar 2009;49(3):582-592.

28. Perez-Nueno VI, Rabal O, Borrell JI, Teixido J. APIF: a new interaction fingerprint based on atom pairs and its application to virtual screening. *J Chem Inf Model*. May 2009;49(5):1245-1260.

29. Marcou G, Rognan D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J Chem Inf Model*. Jan-Feb 2007;47(1):195-207.

30. Radifar M, Yuniarti N, Istyastono EP. PyPLIF: python-based protein-ligand fnteraction fingerprinting. *Bioinformatics*. 2013;9(6):325-328.

31. 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*. Mar-Apr 2006;46(2):686-698.

32. Liu J, Su M, Liu Z, Li J, Li Y, Wang R. Enhance the performance of current scoring functions with the aid of 3D protein-ligand interaction fingerprints. *BMC Bioinformatics*. Jul 18 2017;18(1):343.

33. Wang DD, Chan MT, Yan H. Structure-based protein-ligand interaction fingerprints for binding affinity prediction. *Comput Struct Biotechnol J*. 2021;19:6291-6300.

34. 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*. Apr 22 2013;53(4):763-772.

35. Zhao Z, Xie L, Xie L, Bourne PE. Delineation of Polypharmacology across the Human Structural Kinome Using a Functional Site Interaction Fingerprint Approach. *J Med Chem*. May 12 2016;59(9):4326-4341.

36. Xie L, Ge X, Tan H, Xie L, Zhang Y, Hart T, et al. Towards structural systems pharmacology to study complex diseases and personalized medicine. *PLoS Comput Biol*. May 2014;10(5):e1003554.

37. Sotriffer CA, Sanschagrin P, Matter H, Klebe G. SFCscore: scoring functions for affinity prediction of protein-ligand complexes. *Proteins*. Nov 1 2008;73(2):395-419.

38. Ballester PJ, Schreyer A, Blundell TL. Does a more precise chemical description of protein-ligand complexes lead to more accurate prediction of binding affinity? *J Chem Inf Model*. Mar 24 2014;54(3):944-955.

39. Liu J, Wang R. Classification of current scoring functions. *J Chem Inf Model*. Mar 23 2015;55(3):475-482.

40. Ballester PJ, Mitchell JB. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics*. May 1 2010;26(9):1169-1175.

41. Schreyer A, Blundell T. CREDO: a protein-ligand interaction database for drug discovery. *Chem Biol Drug Des*. Feb 2009;73(2):157-167.

42. Deng Z, Chuaqui C, Singh J. Knowledge-based design of target-focused libraries using protein-ligand interaction constraints. *J Med Chem*. Jan 26 2006;49(2):490-500.

43. Nandigam RK, Kim S, Singh J, Chuaqui C. Position specific interaction dependent scoring technique for virtual screening based on weighted protein--ligand interaction fingerprint profiles. *J Chem Inf Model*. May 2009;49(5):1185-1192.

44. Mordalski S, Kosciolek T, Kristiansen K, Sylte I, Bojarski AJ. Protein binding site analysis by means of structural interaction fingerprint patterns. *Bioorg Med Chem Lett*. Nov 15 2011;21(22):6816-6819.

45. Da Silva F, Desaphy J, Rognan D. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein-Ligand Interactions. *ChemMedChem*. Mar 20 2018;13(6):507-510.

46. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model*. May 24 2010;50(5):742-754.

47. Wojcikowski M, Kukielka M, Stepniewska-Dziubinska MM, Siedlecki P. Development of a protein-ligand extended connectivity (PLEC) fingerprint and its application for binding affinity predictions. *Bioinformatics*. Apr 15 2019;35(8):1334-1341.

48. Da C, Kireev D. Structural protein-ligand interaction fingerprints (SPLIF) for structure-based virtual screening: method and benchmark study. *J Chem Inf Model*. Sep 22 2014;54(9):2555-2561.

49. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: a benchmark for molecular machine learning. *Chem Sci*. Jan 14 2018;9(2):513-530.

50. Wang DD, Xie H, Yan H. Proteo-chemometrics interaction fingerprints of protein-ligand complexes predict binding affinity. *Bioinformatics*. Feb 26 2021;37(17):2570-2579.

51. Rastelli G, Pinzi L. Computational polypharmacology comes of age. *Front Pharmacol*. 2015;6:157.

52. Laufer S, Bajorath J, Gehringer M, Gray N, Frye S, Lindsley CW. Publication Criteria and Requirements for Studies on Protein Kinase Inhibitors horizontal line What Is Expected? *J Med Chem*. May 5 2022:DOI: 10.1021/acs.jmedchem.2c00623.

53. 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 Bioinform*. 2007;8 Suppl 4:S9.

54. Desaphy J, Raimbaud E, Ducrot P, Rognan D. Encoding protein-ligand interaction patterns in fingerprints and graphs. *J Chem Inf Model*. Mar 25 2013;53(3):623-637.

55. Attwood MM, Fabbro D, Sokolov AV, Knapp S, Schioth HB. Trends in kinase drug discovery: targets, indications and inhibitor design. *Nat Rev Drug Discov*. Nov 2021;20(11):839-861.

56. Ferguson FM, Gray NS. Kinase inhibitors: the road ahead. *Nat Rev Drug Discov*. May 2018;17(5):353-377.

57. Klaeger S, Heinzlmeir S, Wilhelm M, Polzer H, Vick B, Koenig PA, et al. The target landscape of clinical kinase drugs. *Science*. Dec 1 2017;358(6367)

58. Liao JJ. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J Med Chem*. Feb 8 2007;50(3):409-424.

59. Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach DR, Miller WT, et al. Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res*. Aug 1 2002;62(15):4236-4243.

60. 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 3:10-15.

61. Takahashi RH, Choo EF, Ma S, Wong S, Halladay J, Deng Y, et al. Absorption, Metabolism, Excretion, and the Contribution of Intestinal Metabolism to the Oral Disposition of [14C]Cobimetinib, a MEK Inhibitor, in Humans. *Drug Metab Dispos*. Jan 2016;44(1):28-39.

62. Zhao Z, Bourne PE. Structural Insights into the Binding Modes of Viral RNA-Dependent RNA Polymerases Using a Function-Site Interaction Fingerprint Method for RNA Virus Drug Discovery. *J Proteome Res*. Nov 6 2020;19(11):4698-4705.

63. Stevens LJ, Pruijssers AJ, Lee HW, Gordon CJ, Tchesnokov EP, Gribble J, et al. Mutations in the SARS-CoV-2 RNA dependent RNA polymerase confer resistance to remdesivir by distinct mechanisms. *Sci Transl Med*. Apr 28 2022:eabo0718.

64. Zhao Z, Liu Q, Bliven S, Xie L, Bourne PE. Determining Cysteines Available for Covalent Inhibition Across the Human Kinome. *J Med Chem*. Apr 13 2017;60(7):2879-2889.

65. Liu Q, Sabnis Y, Zhao Z, Zhang T, Buhrlage SJ, Jones LH, et al. Developing irreversible inhibitors of the protein kinase cysteinome. *Chem Biol*. Feb 21 2013;20(2):146-159.

66. Ayala-Aguilera CC, Valero T, Lorente-Macias A, Baillache DJ, Croke S, Unciti-Broceta A. Small Molecule Kinase Inhibitor Drugs (1995-2021): Medical Indication, Pharmacology, and Synthesis. *J Med Chem*. Jan 27 2022;65(2):1047-1131.

67. Roskoski R, Jr. Properties of FDA-approved small molecule protein kinase inhibitors: A 2022 update. *Pharmacol Res*. Jan 2022;175:106037.

68. Cross DA, Ashton SE, Ghiorghiu S, Eberlein C, Nebhan CA, Spitzler PJ, et al. AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discov*. Sep 2014;4(9):1046-1061.

69. Lamb YN, Scott LJ. Osimertinib: A Review in T790M-Positive Advanced Non-Small Cell Lung Cancer. *Target Oncol*. Aug 2017;12(4):555-562.

70. Soria JC, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, et al. Osimertinib in Untreated EGFR-Mutated Advanced Non-Small-Cell Lung Cancer. *N Engl J Med*. Jan 11 2018;378(2):113-125.

71. Yosaatmadja Y, Silva S, Dickson JM, Patterson AV, Smaill JB, Flanagan JU, et al. Binding mode of the breakthrough inhibitor AZD9291 to epidermal growth factor receptor revealed. *J Struct Biol*. Dec 2015;192(3):539-544.

72. Zhao Z, Xie L, Bourne PE. Insights into the binding mode of MEK type-III inhibitors. A step towards discovering and designing allosteric kinase inhibitors across the human kinome. *PLoS One*. 2017;12(6):e0179936.

73. Cohen P, Cross D, Janne PA. Kinase drug discovery 20 years after imatinib: progress and future directions. *Nat Rev Drug Discov*. Jul 2021;20(7):551-569.

74. Wu P, Clausen MH, Nielsen TE. Allosteric small-molecule kinase inhibitors. *Pharmacol Ther*. Dec 2015;156:59-68.

75. Lu X, Smaill JB, Ding K. New Promise and Opportunities for Allosteric Kinase Inhibitors. *Angew Chem Int Ed Engl*. Aug 10 2020;59(33):13764-13776.

76. Adachi-Yamada T, Nakamura M, Irie K, Tomoyasu Y, Sano Y, Mori E, et al. p38 mitogen-activated protein kinase can be involved in transforming growth factor beta superfamily signal transduction in Drosophila wing morphogenesis. *Mol Cell Biol*. Mar 1999;19(3):2322-2329.

77. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science*. Dec 6 2002;298(5600):1912-1934.

78. Jia Y, Yun CH, Park E, Ercan D, Manuia M, Juarez J, et al. Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. *Nature*. Jun 2 2016;534(7605):129-132.

79. Song X, Zhong H, Qu X, Yang L, Jiang B. Two novel strategies to overcome the resistance to ALK tyrosine kinase inhibitor drugs: Macrocyclic inhibitors and proteolysis-targeting chimeras. *MedComm (2020)*. Sep 2021;2(3):341-350.

80. Zhao Z, Bourne PE. Revealing Acquired Resistance Mechanisms of Kinase-Targeted Drugs Using an on-the-Fly, Function-Site Interaction Fingerprint Approach. *J Chem Theory Comput*. May 12 2020;16(5):3152-3161.

81. Zhao Z, Xie L, Bourne PE. Structural Insights into Characterizing Binding Sites in Epidermal Growth Factor Receptor Kinase Mutants. *J Chem Inf Model*. Jan 28 2019;59(1):453-462.

82. 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*. Apr 8 2008;105(14):5441-5446.

83. Renaud JP, Chari A, Ciferri C, Liu WT, Remigy HW, Stark H, et al. Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat Rev Drug Discov*. Jul 2018;17(7):471-492.

84. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. Aug 2021;596(7873):583-589.

Legend:

Figure 1. (a) 7 types of interactions encoded for every binding pocket residue using the S-IFP approach. (b) 11 types of interactions encoded in the IFP approach from the Rognan group. "1" indicates the interaction exists and "0" indicates no interaction with the residue. (c) A flowchart of the Fs-IFP approach.

Figure 2. (a). The binding clusters mapped to the tyrosine kinase Imatinib (green) complex (PDB id: 4CSV) (b). Residues with conserved interactions with bioactive ligands. The spheres 1-5 correspond to residues in the Hinge region, β3, and β7 (PDB id: 4AN2). (c). The binding clusters of the Type-I kinase Lapatinib (red) complex (PDB id: 3BBT). (d) The binding clusters of Type-III kinase drug Cobimetinib (red) complex (PDB id: 7JUS).

Figure 3. (a) The six non-hydrogen atoms of cysteine. (b). The contribution of each atom to ligand interactions. (c) The locations of top 5 easily-available cysteines (PDB template: 3BYU). (d). The covalent binding mode of Osimertinib (PDB id: 6JXT).

Figure 4. (a) The binding mode of Type-I kinase inhibitors (PDB id: 4w9x, a JAK-Baricitinib complex structure). (b) The binding mode of Type-II kinase inhibitors (PDB id: 4CSV, an SRC-ABL-Imatinib complex). (c) The binding modes of allosteric inhibitors targeting MEK (PDB id: 4AN2). (d). The top 15 predicted kinases targets suitable for allosteric inhibitor design using the TREEspot software (www.discoverx.com).

Figure 5. (a). The similar binding characteristics of Crizotinib in the ALK binding site before and after the L1196M mutation (PDB ids: 2XP2 and 2YFX). (b) Ceritinib-bound ALK complex (PDB ID: 3MKC). (c) The changes of Fs-IFPs in the Crizotinib-bound state before and after L1196M mutation in the ALK-Crizotinib system. (d) The changes of Fs-IFPs in the Ceritinib-bound state before and after L1196M mutation in the ALK-Ceritinib system. An asterisk indicates a significant difference (p-value <0.001).

Figure 1

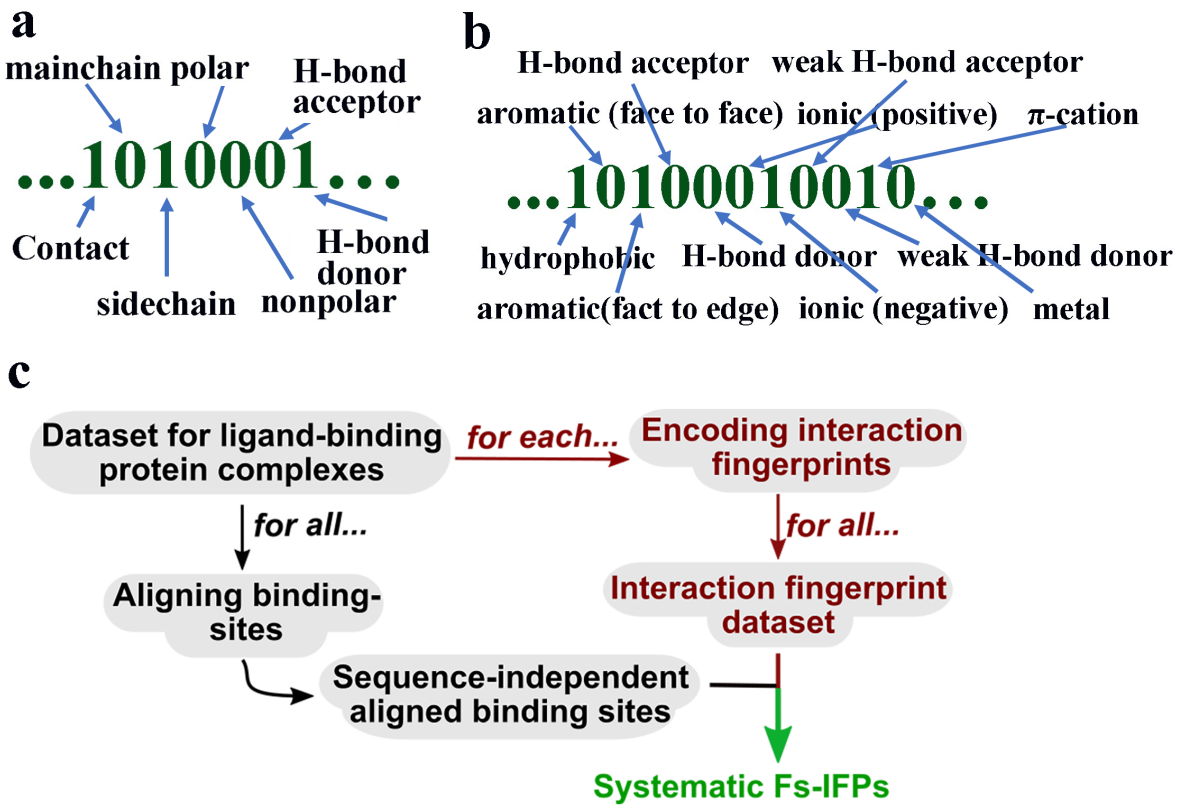
a
mainchain polar
H-bond acceptor
...1010001...
Contact
sidechain
nonpolar
H-bond donor
b
H-bond acceptor
weak H-bond acceptor
aromatic (face to face)
ionic (positive)
π-cation
...10100010010...
hydrophobic
H-bond donor
weak H-bond donor
aromatic(fact to edge)
ionic (negative)
metal
c
Dataset for ligand-binding protein complexes
for each...
Encoding interaction fingerprints
for all...
for all...
Aligning binding-sites
Interaction fingerprint dataset
Sequence-independent aligned binding sites
Systematic Fs-IFPs

Figure 2.

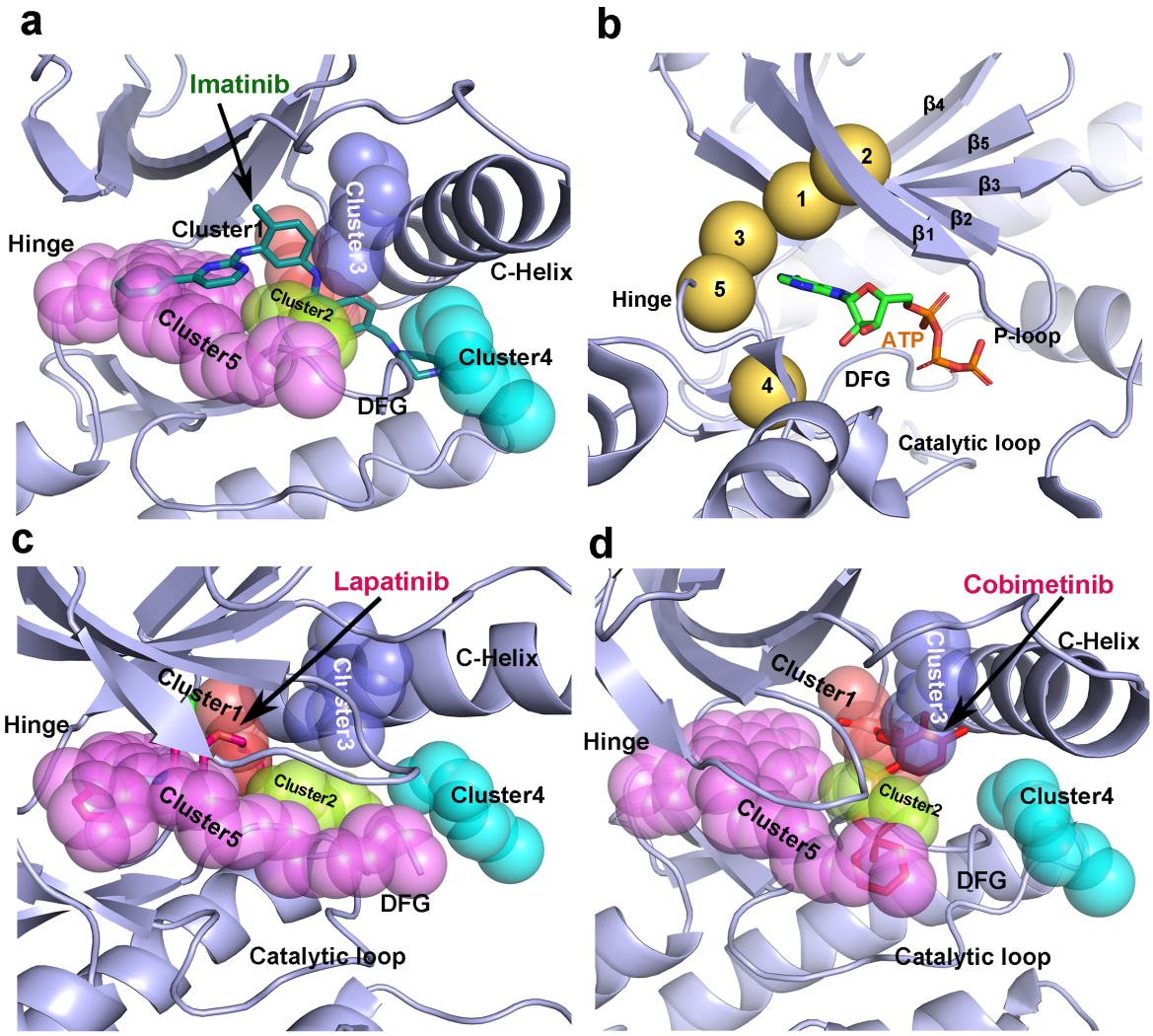
a
Imatinib
Cluster1
Cluster3
Hinge
C-Helix
Cluster2
Cluster5
Cluster4
DFG
b
β4
β5
β3
β2
β1
2
1
3
5
4
Hinge
ATP
P-loop
DFG
Catalytic loop
c
Lapatinib
C-Helix
Cluster1
Cluster3
Hinge
Cluster2
Cluster4
Cluster5
DFG
Catalytic loop
d
Cobimetinib
C-Helix
Cluster1
Cluster3
Hinge
Cluster2
Cluster4
Cluster5
DFG
Catalytic loop

Figure 3.

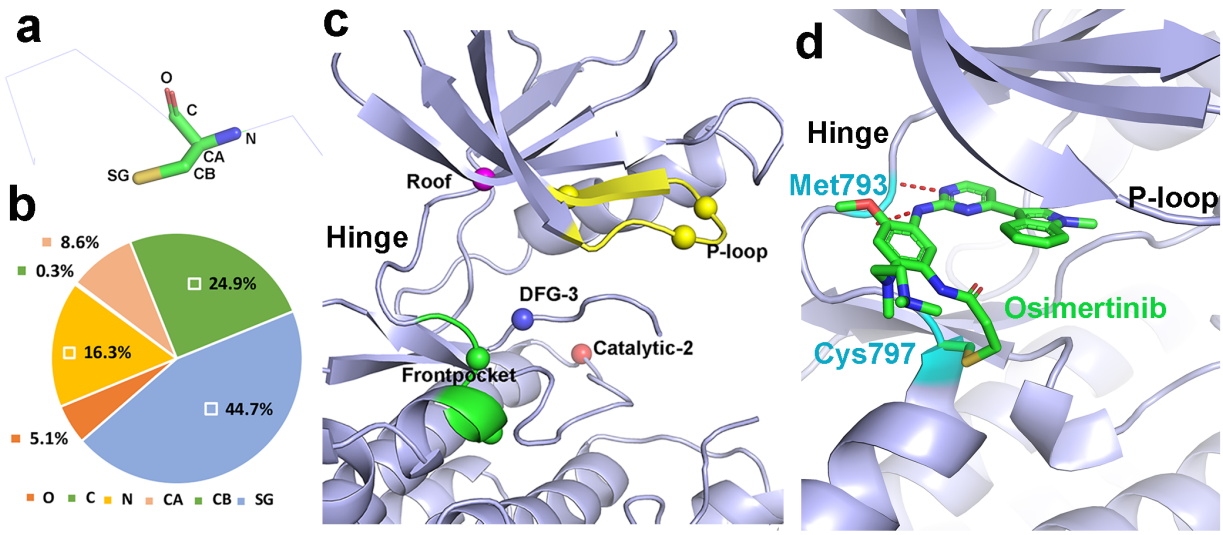
a
O
C
N
CA
CB
SG
b
8.6%
0.3%
24.9%
16.3%
44.7%
5.1%
O
C
N
CA
CB
SG
c
Roof
Hinge
P-loop
DFG-3
Catalytic-2
Frontpocket
d
Hinge
Met793
P-loop
Osimertinib
Cys797

Figure 4.

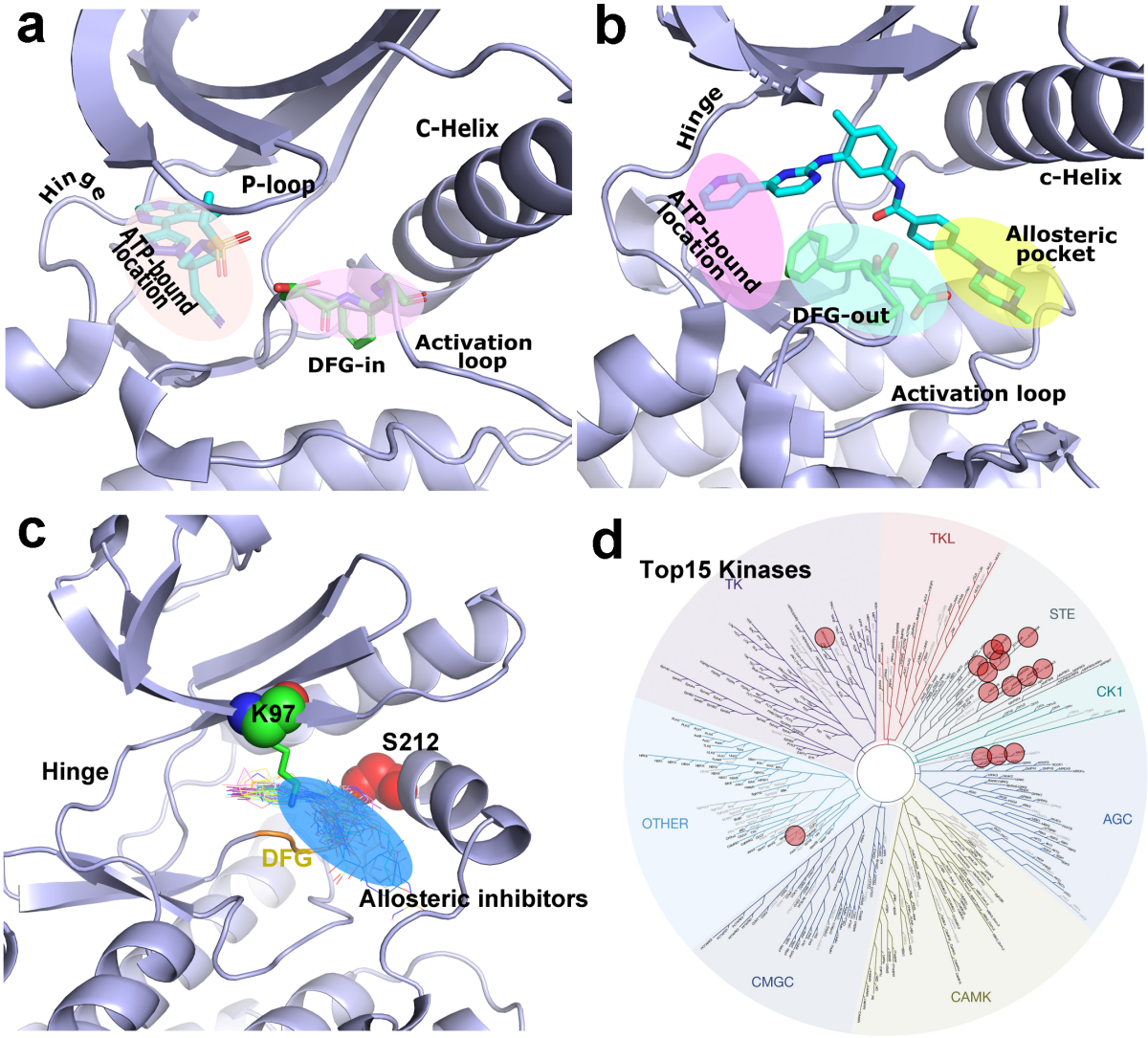
a
Hinge
P-loop
C-Helix
ATP-bound location
DFG-in
Activation loop
b
Hinge
c-Helix
ATP-bound location
Allosteric pocket
DFG-out
Activation loop
c
K97
S212
Hinge
DFG
Allosteric inhibitors
d
Top15 Kinases
TKL
TK
STE
CK1
AGC
OTHER
CMGC
CAMK

Figure 5.

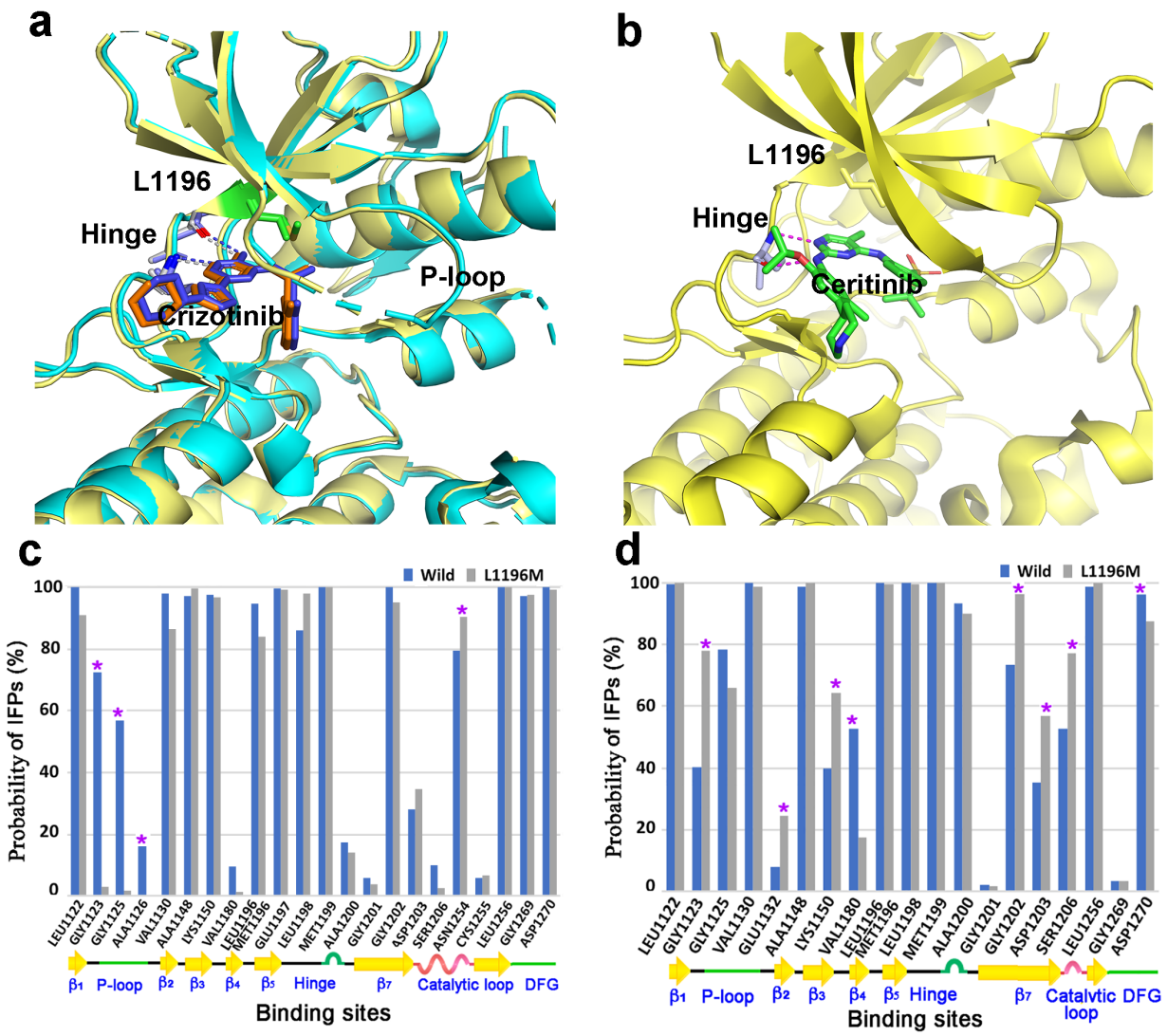
a
L1196
Hinge
P-loop
Crizotinib
b
L1196
Hinge
Ceritinib
c
Wild
L1196M
Probability of IFPs (%)
100
80
60
40
20
0
LEU1122 GLY1123 GLY1125 ALA1126 VAL1130 ALA1148 LYS1150 VAL1180 LEU1196 MET1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 SER1206 ASN1254 CYS1255 LEU1256 GLY1269 ASP1270
β1 P-loop β2 β3 β4 β5 Hinge β7 Catalytic loop DFG
Binding sites
d
Wild
L1196M
Probability of IFPs (%)
100
80
60
40
20
0
LEU1122 GLY1123 GLY1125 VAL1130 GLU1132 ALA1148 LYS1150 VAL1180 LEU1196 MET1196 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 SER1206 LEU1256 GLY1269 ASP1270
β1 P-loop β2 β3 β4 β5 Hinge β7 Catalytic loop DFG
Binding sites

Table 1. The approved drugs by using computer-aided drug discovery strategies.

DRUGS	DISEASES	TARGETS
ALISKIREN	Hypertension	Renin
AMPRENAVIR	AIDS	HIV-1 protease
BETRIXABAN	Cardiovascular disease	Serine protease factor Xa
BRIGATINIB	Non-small cell lung cancer	Anaplastic lymphoma kinas
BOCEPREVIR	Hepatitis C virus	Proteases
CAPTOPRIL	Hypertension or high BP	Angiotensin-converting enzyme
CRIZOTINIB	Non-small cell lung cancer	Anaplastic lymphoma kinas
DORZOLAMIDE	Glaucoma	Carbonic anhydrase II
ENFUVIRTIDE	HIV-1 infection	HIV protein
GRAZOPREVIR	Hepatitis C virus	NS3/4a protease
INDINAVIR	AIDS	Proteases of HIV1 and HIV 2
LOSARTAN	Hypertension	Angiotensin II antagonist
NELFINAVIR	AIDS	HIV-1 protease
NILOTINIB	Chronic myeloid leukemia	Bcr-Abl tyrosine kinase
NORFLOXACIN	Urinary tract infections and prostatitis	DNA gyrase
OSELTAMIVIR	Influenza	Neuraminidase
RALTEGRAVIR	AIDS	HIV integrase
RITONAVIR	AIDS	Proteases of HIV1 and HIV 2
RUCAPARIB	Advanced ovarian cancer	Poly(ADP-ribose) polymerase
SAQUINAVIR	AIDS	Proteases of HIV1 and HIV 2
SUNITINIB	Gastrointestinal stromal tumor	Vascular endothelial growth factor receptor
TIROFIBAN	Blood clots	Fibrinogen
VABORBACTAM	Gram-negative bacteria	β-lactamase
VALSARTAN	High blood pressure, heart failure, and diabetic kidney disease	Angiotensin receptor II antagonist
ZANAMIVIR	Influenza A and influenza B	Neuraminidase
ZOLMITRIPTA	Migraine	Serotonin receptor agonist

Table 2. The FDA-approved covalent kinase inhibitors until April 2022.

Drugs	Approved Date	Primary targets	Nucleophilic residues	Locations of nucleophilic residues	PDB IDs
Afatinib	2013/07	EGFR	Cys797	Frontpocket	4G5J
Ibrutinib	2013/11	BTK	Cys481	Frontpocket	5P9J
Osimertinib	2015/11	EGFR	Cys797	Frontpocket	6JXT
Acalabrutinib	2017/10	BTK	Cys481	Frontpocket	-
Neratinib	2017/06	EGFR	Cys797	Frontpocket	2JIV
Dacomitinib	2018/09	EGFR	Cys797	Frontpocket	4I24
Zanubrutinib	2019/11	BTK	Cys481	Frontpocket	6J6M
Mobocertinib	2021/09	EGFR	Cys797	Frontpocket	-