Published in final edited form as:
J Chem Inf Model. 2019 January 28; 59(1): 453–462. doi:10.1021/acs.jcim.8b00458.

Structural insights into characterizing binding sites in EGFR kinase mutants

Zheng Zhao[1], **Lei Xie**[2,3], and **Philip E. Bourne**[1,4,*]

[1]Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia 22904, United States of America

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

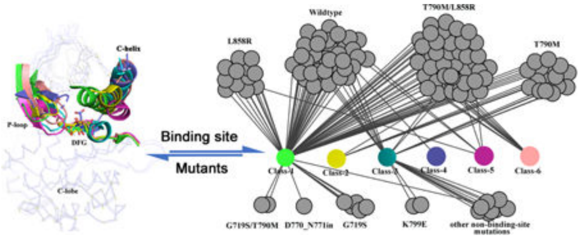
[3]The Graduate Center, The City University of New York, New York, New York 10016, United States of America

[4]Data Science Institute, University of Virginia, Charlottesville, Virginia 22904, United States of America

Abstract

Over the last two decades epidermal growth factor receptor (EGFR) kinase has become an important target to treat non-small cell lung cancer (NSCLC). Currently, three generations of EGFR kinase-targeted small molecule drugs have been FDA approved. They nominally produce a response at the start of treatment and lead to a substantial survival benefit for patients. However, long-term treatment results in acquired drug resistance and further vulnerability to NSCLC. Therefore, novel EGFR kinase inhibitors that specially overcome acquired mutations are urgently needed. To this end, we carried out a comprehensive study of different EGFR kinase mutants using a structural systems pharmacology strategy. Our analysis shows that both wild-type and mutated structures exhibit multiple conformational states that have not been observed in solved crystal structures. We show that this conformational flexibility accommodates diverse types of ligands with multiple types of binding modes. These results provide insights for designing a new-generation of EGFR kinase inhibitor that combats acquired drug-resistant mutations through a multi-conformation-based drug design strategy.

Graphical abstract



[*]**Corresponding author:** Philip E. Bourne: phone, (434) 924-6867; peb6a@virginia.edu.

Introduction

Epidermal growth factor receptor (EGFR) is one of four members of the transmembrane epidermal growth factor receptor family (EGF/ErbB receptor family). It is typically activated by an epidermal growth factor (EGF) that binds to the extracellular domain which protrudes from the cell membrane thereby regulating signaling pathways[1]. Mutation of the EGFR kinase domain leads to higher activity thereby stimulating four major downstream signaling pathways including MAPK/ERK, PLCγ/PKC, JAK/STAT and PI3K/AKT which impact transcription and cell cycle progression[2] leading to cancers and inflammatory diseases[3]. Specifically, it has been shown that diverse mutations in the EGFR kinase domain, typically L858R and del746–750, are associated with non-small cell lung cancer (NSCLC) which accounts for about 80% to 85% of all lung cancer cases[4]. Thus, EGFR kinase inhibitors that target these activation mutations are desirable for the treatment of NSCLC. Since 2009, five EGFR kinase inhibitors have been approved by the US Food and Drug Administration (FDA). The first two, Gefitinib and Erlotinib, show efficacy at the start of treatment. However, after about 12 months, drug resistance from an acquired T790M mutation arises[5]. Thus, a second-generation EGFR kinase inhibitor, Afatinib, was developed for the T790M mutation. Because of the limited therapeutic potential of Afatinib[6], soon after, a third-generation of EGFR kinase inhibitors, Osimertinib and Olmutinib, were developed. Afatinib, Osimertinib and Olmutinib are all irreversible inhibitors[7] that form a covalent bond with Cys797. However, the occurrence of an acquired C797S mutation greatly reduces the efficacy of these three covalent drugs. Therefore, to date, this acquired resistance remains a major challenge in the treatment of NSCLC[8–9]. New EGFR kinase inhibitors that overcome these acquired mutations are needed. Recently, Jia et al. discovered an allosteric EGFR kinase inhibitor (EAI045)[10], which overcomes both T790M and C797S mutations, and offers a means for treating NSCLC. Further, a few non-covalent ATP-competitive inhibitors that bind to the ATP binding site were rationally designed to overcome acquired resistance[11–15]. For example, Marcel Günther et al.[14] developed a reversible inhibitor that can inhibit the L858R/T790M/C797S triple mutant. Hwangseo Park et al.[15] also identified a few ATP-competitive inhibitors that overcome the Del746–750/T790M/C797S/ triple mutant. This next-generation allosteric or non-covalent inhibitors demonstrate promising potential and an opportunity to overcome the multiple mutations through designing elaborate inhibitors that match the mutated binding site of the EGFR kinase[16]. These findings prompted us to revisit the EGFR kinase binding site to obtain a detailed understanding of the similarity and differences between the wild-type and mutant EGFR kinases.

Currently, with many available EGFR structures, there is the data-driven opportunity to focus on the EGFR binding site using a structural systems pharmacology strategy[17–20]. We collected all released structures of the EGFR kinase domain to analyze the overall conformational space of the binding sites as well as the respective kinase-ligand binding modes. We focused on all diverse acquired mutations at the binding sites, exploring the binding site flexibility of EGFR kinase domains with the del746–750/T790M/C797S/ mutation and the L858R/T790M/C797S mutations by using μs-scale molecular dynamics (MD). MD simulation has been successfully applied to study the EGFR conformation space at a computational physiological environment[21–24]. For example, Park et. al explored the

EGFR conformational transition between the inactive and active state to determine the role of gatekeeper mutation on inhibitor selectivity using MD simulations[21]. Kannan et. al used the combined MD simulation with binding assay data to reveal a mutant specific pocket[24]. Additionally, we scrutinized the binding site features at an atomic level using the function-site interaction fingerprint approach and the volumetric analysis of surface properties. Our analyses show both wild-type and mutated structures encompass multiple conformational states. Novel conformational state that have not been observed in the solved structures can exist. The conformational flexibility in the structure accommodates diverse types of ligands with multiple types of binding modes. These results provide us with critical insights into future drug design in treating drug resistant NSCLC and other cancers[25].

Results

1. EGFR kinase conformations and mutations

The similarities and differences in EGFR kinase binding sites can be determined by comparing their X-ray structures. We clustered EGFR kinase structures in our dataset into two kinase states, active and inactive and into six classes (Figure 1a). In these classes, the side chain of the aspartic acid in the DFG motif points in three different directions: "DFG-In", "DFG-Out", and DFG-$\frac{1}{2}$In. In DFG-$\frac{1}{2}$In, the side chain of the aspartic acid points to the roof[26]. Similarly, the C-helix also exhibits significant displacements called "In" and "Out", and the state between them (labeled "$\frac{1}{2}$Out"). The number of x-ray structures and the representative structure in each class are shown in Figure 1a (a complete clustering of structures is in Table S1). Using the representative structure from each class, the major differences in the P-loop, C-helix and DFG motifs are highlighted in Figure 1b. In the active state (Class-1 and Class-2), the DFG motif is in a "DFG-In" state. The C-helix is found in the "In" state (class-1, Figure 1a-b) or has a small displacement referred to as "$\frac{1}{2}$Out " (Class-2, Figure 1a-b). In the inactive kinase (class-3–6), the DFG motif is found in three directions and the C-helix is found in the "Out" state in all structures. These distinctly different conformations in the binding site of EGFR kinases facilitates the identification of potential sites outside of the conserved ATP-binding site, providing a structural basis to accelerate structure-based drug design through the discovery of novel binding modes.

Acquired multi-drug resistance motivates the search for drugs that can target mutated structures. To begin this process, we catalog the currently available conformations of EGFR kinase domains, both wildtype and mutations, according to the scheme described above (Figure 1c). The wildtype (WT) EGFR kinases (37 structures) fall into four classes of conformation (Class-1,3,4,5). The L858R/T790 mutation (41structures) also exhibits different conformational states (Class-1,2,3,5,6). Other mutated classes such as T790M, which have a limited number of released structures (14 structures), present fewer unique conformations (Class-1,2,3). As catalogued, the WT and mutated structures can have the same conformation. For example, three types of EGFR kinase, the WT, L858R and T790M/L858R (Figure 1c) all adopt the Class-5 conformation. In summary, WT and mutated structures both present multiple conformations. From a drug design perspective, the good news is that multiple conformations suggest flexibility in the EFGR kinase domain providing opportunities to design potent and specific inhibitors by accommodating a unique conformation. The bad news is the challenge to protein-ligand docking and modeling quantitative structure-activity relationship since the ligand-induced conformation is not known a priori. Here we attempt to address this shortcoming.

We also marked two important motifs, the R- and C-spine[27–29] (Figure S1). The catalytic C-spine is assumed highly stable since it is strikingly similar in all six classes of EGFR states. However, the R-spine has different architectures. In Class-1 and 2 at the active state, the R-spine residues are linearly connected. In Class-3 and 4 at the DFG-in/C-helix-out inactive state, the R-spine is partially assembled because the C-helix displacement leads to M766 movement. In the DFG-out inactive state (Class-5 and 6), the R-spine is not assembled because the sidechain of the F856 residue flips to the other side of the DFG peptide.

2. Comparing binding sites among EGFR wildtype and mutants

We calculate changes in the size and shape of the binding pockets for different classes of conformation. The volume of the binding pockets is 1119 Å^3, 953 Å^3, 950 Å^3, 1056 Å^3 and 972 Å^3 for Class-1 to 5, respectively (Figure S2). These pockets are not significantly different, however, the volume of the binding pocket in Class-6 is significantly increased (1913 Å^3). It can thus accommodate more diverse types of small molecule ligands. The difference in the shape and size of the binding pockets is shown in Figure 2. Given that the Class-1 conformation (Figure 2a) is the most frequently observed (Figure 1a), we compared the binding cavities of the other five classes with that of Class-1. Specifically, we decomposed the binding cavity into five sub-pockets: front pocket (FP), GxGxxG motif at the P-loop (G-rich-loop), and back clefts BP-I, BP-II, and BP-III[30], and quantitatively compared them (Figure 2b-f (upper and lower). BP-I is located at the back cleft close to the adenine pocket; BP-II includes the hydrophobic pocket at the back cleft; BP-III is the sub-pocket at the back cleft comprising the DFG peptide, C-terminal end of the C-helix and the N-terminus of the catalytic loop. Table S2 lists the volume of every sub-pocket.

Firstly, compared with the back cleft of Class-1, the volume of the back cleft in Class-2 and Class-5 is smaller. Their back binding sub-pocket is not formed (Figure 2b and 2e). However, the BP-II cleft of Class-3, the BP-II cleft of Class-4 and the BP-II, III clefts of Class-6 is present (Figure 2c,2d and 2f). It is worth noting that the C-helix in Class 2–6 is in "OUT" or "$\frac{1}{2}$Out " state. Thus, the C-helix displacement does not always induce the formation of the back cleft. Secondly, in Class 1–6, the volumetric size of the FP region that binds ATP is similar but the shape is slightly different (Figure 2 and Table S2). Finally, for the sub-pocket at the G-rich loop, a large binding pocket is formed in Class-2, Class-5 and Class-6. However, there is no space to accommodate a ligand in Class-3 and Class-4 similar to Class-1 (Figure 2b-f). Interestingly, Class-6 (Figure 2f) is different from other classes; besides the regions of FP and G-rich-loop, a large sub-pocket is formed at BP-II and BP-III. Thus, the cavity of Class-6 has adequate space, not only at the ATP-competitive site, but also at the allosteric site. This conformational mode may provide the structural basis for the design of diverse type-II and/or type-III EGFR kinase inhibitors[31–32].

In summary, we compared the different binding pockets quantitatively and qualitatively. The changes in the conformations in the C-helix and DFG motif induces diverse sub-pockets within the binding cavities. It is worth noting that we only analyze the difference in the binding cavities using the rigid x-ray structures. Nevertheless, the induced sub-pockets present the diverse shape and size of the binding pockets. It suggests that the plasticity of the EGFR kinase binding pocket combined with the inducing dynamic conformational space introduced by molecular dynamics has great potential for designing inhibitors with desired selectivity profiles.

3. Dynamic conformational space of the binding sites

Starting from their X-ray structures, the μs-scale equilibration of the trajectories yields a large number of conformations for the wildtype and different mutated EGFR kinases. The trajectory analysis provides information on the similarity of conformations and frequent conformational transitions based on the total energy curves (Figure S3), the root mean square deviation of Cα atoms (Cα-RMSD) (Figure S4) and the root mean square fluctuation of Cα atoms (Cα-RMSF) (Figure S5). The main collective motions, as inferred from the first principal component of PCA, are shown in Figure S6. For all three systems, their ATP binding pockets show little variation. Around the binding site, the flexibility mainly comes from P-loop, the activation loop and C-helix. From specific detailed comparison of the binding sites, these conformations show a high degree of deviation from the initial crystal x-ray structure (PDB id: 1m17, Class-1, Figure 1a). Starting from the DFG-In/C-helix-In conformations, the resulting equilibrated structures present different conformations including both the active and inactive states (Figure 3). Among them, the DFG motif exhibits significant displacement from DFG-In to DFG-out, and the C-helix transitions from C-helix-in to C-helix-out. In the trajectory of wild-type structures (Figure 3a), all six classes of conformation can be sampled. As shown in Figure 3b-c for the mutated structures of del746–750/T790M/C797S and L858R/T790M/C797S, the conformations range from active state (Class-1 and Class-2) to inactive (Class-3 and Class-4). It is well known that conformational transition needs to overcome the free energy barrier[33]. It could be the reason that we cannot capture Class-5 and Class-6 conformations using unbiased MD simulation given our current MD time-scale. To explore a larger conformational space improved sampling technologies or longer simulation time are required[22–23] as described by Shan[22], Sutto[23], and Park[21]. Nevertheless, it is clear, even from traditional MD simulation, that multiple conformational states of the EGFR kinase in both WT and mutated systems exist. Importantly, we observe a new conformational state (labeled as Class-D1 in Figure 3c), which doesn't occur in the EGFR kinase structure dataset. This new conformation (Figure 4) has"DFG-$\frac{1}{2}$ In/C-helix-out",in which the DFG flipping and C-helix-out displacement are similar to that of Class-4. The difference is that the salt bridge between K745 and E762 occurred in Class-D1 and a hydrogen-bond interaction formed between D855 and K745. This salt bridge and the hydrogen-bond interaction stabilize this dynamic conformation (Figure 4). In the Class-D1 state, when C-helix displacement occurs, the sheets consisting of beta-strands 1–5 have a combined displacement outward (Figure 4). It will be interesting to screen for novel inhibitors using this new conformation.

We further analyzed the features of the Class-D1 binding pocket. The volume of the binding pocket is 1204 Å^3 (Figure S7), which is slightly larger than that of Class 1–5. We compared the Class-D1 binding site with the other six classes of binding sites. The main difference between Class-D1 and Class 1–5 is the BP-III region. The main difference from Class 6 is the G-rich-loop region (Figure S8). Obviously, the Class-D1 binding pocket has a large sub-pocket at the BP-III region. More specifically, we calculated the similarity between the Class-D1 binding pocket and the other six classes of binding pockets using the volumetric distance[34] (Table S3), which shows the Class-6 binding site is most similar to Class-D1. Like Class 6, the slightly larger BP-III sub-pocket is also available in the Class-D1 state. Additionally, to estimate the possible binding mode, we screened the whole kinase structurome, including 3180 kinase-ligand structures[35], to obtain the most similar binding pocket to the Class-D1 binding pocket. The top complex (Table S4) is a JAK1 kinase structure[36] (pdb id: 4e4n), where the ligand (0NL) binds to the pocket by bearing a heterocycloamine scaffold, forming hydrogen bonds with the hinge peptide and a t-butyl carbamate at the G-rich-loop region (Figure S9). We also noted that the distance between the t-butyl and the C-helix is 9.6 Å, which means this BP-III sub-pocket region is not fully utilized. This binding mode for the JAK1–0NL complex can guide new EGFR inhibitor design through binding the Class-D1 conformation. The complete structure file (PDB format) of this new conformation is available as Supporting Information.

In summary, the MD simulation reveals more conformational states than those in the released x-ray structure dataset. It is important to analyze the multiple conformational states in the context of the diverse inhibitors that bind to those different conformations. Stated another way, multiple conformations will facilitate novel inhibitor design with a specific SAR. Note, however, that the ligand could also induce a new EGFR kinase conformation with a yet to be discovered binding pocket[37] which points to the need for further structure determination of new ligand-bound complex structures, as well as further computational studies to sample a larger conformation space in the process of drug design in order to obtain the desired SAR[38].

4. EGFR kinase-ligand binding characterization

It is well known that the kinase ATP binding site is highly conserved across the whole kinome. In order to achieve the desired selectivity, it is critical to understand the atomic details of kinase-inhibitor interactions[35]. Here, using all released EGFR kinase crystal structures, we analyze and characterize the atomic details of interactions between the ligand and the residues that constitute the EGFR kinase ATP binding site (see Methods and Table S5 for specific PDB structures). Figure 5 shows the contact frequency between specific residues and the bound ligand. Seven types of interatomic interaction are defined based on geometric rules [39] but not energy (see Methods). The analysis identifies 39 amino acids that are close to the binding site (Figure 5a). The 39 residues are located predominantly in beta-

Supporting information
Figure S1–9 illustrate the C- and R-spine analysis, the volume and shape of the binding pockets, the analysis of different MD trajectories, PCA, the Class-D1 binding pocket, the different between the Class-D1 binding site and all the other six classes of EGFR binding sites, and the screened top complex. Table S1–6 describes the EGFR kinase structure dataset, the volume of every sub-pocket conformation, the similarity list, the top similar complex, the index and the pdb id of every kinase structure, and the classes of the binding mode, respectively. The Class-D1 EGFR kinase structure (**PDB**).

sheets or helixes as shown in Figure 5a and 5b. The residues with the top 5 contact frequencies are L718, V726, A743, M793 and L844. They are located at β1–3, Hinge and β6 (Figure 5b in green) and form a core binding pocket. The core binding pocket is highly hydrophobic, largely conserved and consistent with previous results[35]. The core binding pocket accommodates the adenine base of ATP or the ATP-competitive inhibitor. The high contact frequencies at the core binding pocket also suggests that most of the co-crystallized EGFR kinase inhibitors are ATP-competitive. However, by selectively binding other residues with lower contact frequency, higher selectivity can be anticipated.

To reveal the nature of binding specificity, we compare the pairwise similarity of aligned residue-ligand functional site interaction fingerprints (Fs-IFPs) (see Methods). Hierarchical clustering of the Fs-IFPs for all EGFR kinase-ligand co-crystallized structures was conducted and the binding modes grouped into six clusters that represent different binding modes (Figure 6a and Table S6). In cluster-1 (Figure 6b), the ligand is located at the position of the ATP binding site, where the core binding pocket is highlighted as aforementioned. It is not surprising that most inhibitors fall into this cluster. Typically, there are two or three hydrogen-bond interactions between the ligand and the amino acids located at the hinge[31]. Since the core binding pocket is conserved, this type of inhibitor has a multi-targeted effect across the kinome. For example, the ligand STO (1,2,3,4-tetrahydrogen staurosporine) (Figure 6b) is a derivative of the natural product staurosporine[40], which is known to have broad cross-reactivity[41]. Following the cross reactivity observed, the ligand STO has a specific kinase spectrum that can be targeted. By combining the target spectrum of a given compound, further SAR optimization becomes a valid strategy[26, 42]. By using this strategy, a close analogue of staurosporine (PKC-412, a multi-kinase inhibitor) was approved in April 2017 for treating acute myeloid leukemia (AML), myelodysplastic syndrome (MDS) and advanced systemic mastocytosis[43–45]. Similar to cluster-1, the cluster-2 interaction patterns also include the ATP-competitive binding mode (Figure 6c).

However, in this cluster the kinase is in the inactive state and belongs to the Class-6 conformation as described in Figure 1. In this case the ligand does not have any interaction with the C-helix. Rather the benzenesulfonyl fluoride ligand fragment presents characteristic interactions, namely the aromatic π-π stacking interaction with the phenylalanine side chain of the DFG motif and the electrostatic interactions with the amino acids on the G-rich loop. In cluster-3 (Figure 6d), the ligand also presents an ATP-competitive binding mode. However, different from cluster-1, the ligand has a hydrophobic fragment penetrating deeply into the hydrophobic pocket that adjoins the core binding pocket and is located at the back of the kinase binding site[30]. The hydrophobic pocket is often used to achieve the desired selectivity, but mutation of the gatekeeper residues directly affects the efficacy of this type of inhibitor[46]. Cluster-4 (Figure 6e) has a non-ATP-competitive binding mode. The ligand is located at the back of the binding site, and occupies the hydrophobic pocket, similar to cluster-3, as well as the allosteric site[47]. In this cluster, the C-helix is displaced outward to the “Out” state and opens up the allosteric site. Thus, the ligand forms interactions with the DFG motif and the C-helix. The allosteric binding mode provides high selectivity overcoming the T790M/C797S mutations[10]. In cluster-5 (Figure 6f), ligand binding presents two features. Firstly, the ligand has a phenyl group to bind the hydrophobic pocket that is similar to cluster-3 and cluster-4. Secondly, the tail of the ligand interacts with not only the sidechain of aspartic acid from the DFG motif but also the sidechain of asparagine 842 at β_6. It is notable that the position of the tail is similar to that in cluster-2, but the binding mode is totally different. Firstly, the tail of the ligand forms an interaction with the aspartic acid of DFG-in in cluster-5, but with the phenylalanine of DFG-out in cluster-2. Secondly, the tail extends towards the A-loop and forms an interaction with Asn842 in cluster-5, but extends towards the P-loop and is in contact with the G-rich loop in cluster-2. In cluster-6, the ligand is an ATP-competitive binder. The furan group of the ligand not only forms very subtle interactions with the hydrophobic pocket but also forms polar interactions with K745, E762 and the DGF motif. As aforementioned, in cluster-4, the C-helix contributes to the binding affinity of the allosteric inhibitor. However, although the inhibitors in cluster-6 are not allosteric inhibitors, the C-helix is in the "C-helix-In" state in close proximity to the oxygen atoms of the furan group thereby forming polar interactions. Thus, the furan group provides unique binding characteristics.

In summary, the binding modes of co-crystallized EGFR kinase-ligand complexes are diverse. They provide the structural basis for mediating inhibitor selectivity, important to keep in mind when designing mutant-sensitive inhibitors.

Conclusions

We have explored the impact of mutations on structural conformation of the EGFR kinase domain for all released crystal x-ray structures. From a detailed structural analysis, we identified six classes of conformation which focus on the flipping of the DFG motif and the displacement of the C-helix. There is no apparent correlation between these mutations and conformation. For example, the common T790M mutation adopted not only the typical DFG-in/C-helix-in conformation but also the DFG-out/C-helix-out. We suggest that the conformations of the EGFR kinase are flexible and can fall into distinct locally stable states. From the perspective of drug design, multi-conformation-based drug screening should be of significant value when targeting EGFR kinases.

We also explored the dynamic conformation space occupied by the EGFR kinase using μs-level MD simulation. We not only sampled the multiple different conformations from Class-1 to Class-6 but also found a new conformation mode. In the two acquired mutated kinase systems, we did not sample Class-5 and Class-6 conformations; advanced sampling technologies are needed. Less of an issue perhaps since Class-5 and Class-6 conformations that were obtained from the released ligand-bound structures (Figure 1b), as well as taking conformational flexibility into account, suggests that the EGFR kinase-ligand binding would induce a specific binding mode and EGFR kinase conformation change. Recently, Sonti et al.[38] reported a similar ligand-induced binding mechanism for Abelson tyrosine kinase (Abl). Abl belongs to the TK group, like the EGFR kinase, and has 60% sequence similarity[48]. Therefore, it is important for EGFR kinase-ligand SAR to take the induced binding mechanism into consideration. This presents a challenge to *in silico* compound screening for it requires the protein target be considered not as a static structure but as an ensemble of conformations[49].

Such a change in conformation often results in a change in the binding site. We found that the outward displacement of the C-helix does not always form the BP sub-pockets, (Figure 2b and 2e). Moreover, with these conformational changes the volume and shape of the binding cavities shows systematic plasticity (Figure 2b-f), which can be further elucidated into distinctive binding patterns through binding diverse inhibitors. The derived Fs-IFPs highlight this distinctive protein-ligand binding patterns at the atomic level. It is our hope that within each sub-cluster of fingerprints, the unique binding modes will provide useful clues in designing unique molecular fragments for a new series of EGFR kinase inhibitors.

Method

1. Dataset of EGFR kinase structures

We collected all 127 PDB structures of the EGFR kinase domain released through Dec. 2017 from the Protein Data Bank (www.rcsb.org)[50] by using the human EGFR UniProt[51] identifier, P00533. There are 20 apo and 107 holo PDB structures in the dataset. The binding sites of all structures were aligned using the SMAP software[52–54]. We clustered all structures based on the Cα-RMSD of the EGFR N-terminal lobes. The mutations closest to the binding site of each structure was extracted to determine the relationships between the conformation patterns and the mutations.

2. Differentiating the binding sites

To differentiate the different binding sites, first the given binding sites were pairwise aligned using the SMAP tool[54]. Then, each binding site and each sub-pocket was detected using the Surflex software[55–56] with default proto parameters. For each sub-pocket, a set of residues bordering the sub-pocket were chose based on Liao's description[30]. Then the volumetric size and shape of each binding site or each sub-pocket was inferred using the volumetric analysis of surface properties (VASP) software[34], in which all computations were carried out at 0.5 A resolution. Lastly, differentiating the binding sites was determined using the volumetric difference of constructive solid geometry[34]. The similarity was calculated by using the volumetric distance $v_{(x,\ y)}$,

$$v_{(x,\,y)} = 1 - \frac{v(x \cap y)}{\min(v(x),\ v(y))}$$

where, $v(i)$ represents the volume of a given binding site i in Å^3.$v(x \cap y)$ represents the volumetric intersection of $v(x)$ and $v(y)$ for x and y are binding pockets.

3. Sampling the conformational space using MD

We carried out microsecond-level molecular dynamics (MD) simulation for three systems, the wild-type and two acquired mutated types (L858R/T790M/C797S and del746–750/T790M/C719S). First, the initial structure was prepared from PDB 1m17 with the DFG-in/C-helix-in active conformation as the wild-type EGFR kinase domain. Then the corresponding mutated models were built for each system, respectively, using PDB 1m17 as the template and the software modeller[57]. Then each structure was solvated in a cubic box of water molecules with 12Å distance from the solvated molecules and neutralized by adding

Cl^- and Na^+ using an ACEMD setup script[58]. Finally, a general short-MD protocol was used for initial minimization and equilibration including 2ps minimization, 100ps for NVT, 1ns for NPT with heavy-atom constraints and 1ns for NPT without any constraints. The next stage was a 1.5 μs equilibration process without the constraint and the last 1000ns of MD trajectories were retained for further analysis.

The MD simulations were carried out using the ACEMD software[58] with the CharMM27 force field for protein and the TIP3P water model for the water molecule[59]. The electrostatic interaction was treated using PME and SHAKE constraints were applied with a 4fs integration time step[60]. The temperature was kept at 300K using a Langevin bath method and the pressure was maintained at 1 ATM using the Berendsen method. The trajectories were analyzed using the MMTSB toolset[61]. For the reaction coordinate (RC) of scatter density distribution, we used the salt-bridge distance between K745 and E762 as one RC. Because the flexibility of Cys797 is low as shown in Figure S4, we measured the two distances from C797 to D855 and F856, where the difference constitutes another RC.

4. Function-site interaction fingerprint encoding

Function-site interaction fingerprints (Fs-IFP) are an efficient means to determine functional site binding characteristics and to compare binding sites on a proteome scale as detailed in previous applications[35, 62–63]. In brief, the Fs-IFP method includes three steps. Step 1 prepares the dataset. In this paper, we collected all EGFR kinase-ligand structures as our dataset. Step 2 aligns all function sites. Here we used the SMAP software[52–54] with default parameters. Thus, we obtained all aligned residues within the binding pocket. Step 3 encodes the binding site–ligand interaction. Based on the aligned residues, we confirmed that every binding site could be described using 39 amino acids that comprised the binding pocket. Then the interactions between every amino acid and the corresponding ligand were encoded using a 7-bit array that represents 7 types of interactions: ((1)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); and (7) electrostatic interaction)[39]. Finally, the kinase–ligand interactions of each complex structure were encoded using a length of 273 bits (7 bits × 39 residues). The encoding was done using IChem software[64]. Further, the similarity of the pairwise Fs-IFPs was calculated using the Tanimoto coefficient (TC). Based on all-against-all pairwise TC similarity, the hierarchical cluster analysis was carried out using R with the single linkage method[65].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We thank Prof. Ajay N. Jain for providing the Surflex Platform version 4.1 license, Prof. Brian Y. Chen for providing VASP software and its application and Prof. Didier Rognan for providing the IChem software. We thank Cameron Mura for helpful discussions. This work was partly supported by the University of Virginia (PEB), Grant Number R01LM011986 from the National Library of Medicine of the National Institute of Health (NIH) (LX), and Grant Number R01GM122845 from the National Institute of General Medical Sciences of NIH (LX).

Reference

1. Carpenter G, Employment of the epidermal growth factor receptor in growth factor-independent signaling pathways. J. Cell Biol. 1999, 146 (4), 697–702. [PubMed: 10459005]
2. Oda K; Matsuoka Y; Funahashi A; Kitano H, A comprehensive pathway map of epidermal growth factor receptor signaling. Mol. Syst. Biol. 2005, 1, 2005.0010.
3. Normanno N; De Luca A; Bianco C; Strizzi L; Mancino M; Maiello MR; Carotenuto A; De Feo G; Caponigro F; Salomon DS, Epidermal growth factor receptor (EGFR) signaling in cancer. Gene 2006, 366 (1), 2–16. [PubMed: 16377102]
4. Sharma SV; Bell DW; Settleman J; Haber DA, Epidermal growth factor receptor mutations in lung cancer. Nat. Rev. Cancer 2007, 7 (3), 169–181. [PubMed: 17318210]
5. Kobayashi S; Boggon TJ; Dayaram T; Janne PA; Kocher O; Meyerson M; Johnson BE; Eck MJ; Tenen DG; Halmos B, EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N. Engl. J. Med. 2005, 352 (8), 786–792. [PubMed: 15728811]
6. Miller VA; Hirsh V; Cadranel J; Chen Y-M; Park K; Kim S-W; Zhou C; Su W-C; Wang M; Sun Y; Heo DS; Crino L; Tan E-H; Chao T-Y; Shahidi M; Cong XJ; Lorence RM; Yang JC-H, Afatinib versus placebo for patients with advanced, metastatic non-small-cell lung cancer after failure of erlotinib, gefitinib, or both, and one or two lines of chemotherapy (LUX-Lung 1): a phase 2b/3 randomised trial. Lancet Oncol. 2012, 13 (5), 528–538. [PubMed: 22452896]
7. Zhao Z; Bourne PE, Progress with covalent small-molecule kinase inhibitors. Drug Discovery Today 2018, 23 (3), 727–735. [PubMed: 29337202]
8. Nilsson MB; Sun H; Diao L; Tong P; Liu D; Li L; Fan Y; Poteete A; Lim SO; Howells K; Haddad V; Gomez D; Tran H; Pena GA; Sequist LV; Yang JC; Wang J; Kim ES; Herbst R; Lee JJ; Hong WK; Wistuba I; Hung MC; Sood AK; Heymach JV, Stress hormones promote EGFR inhibitor resistance in NSCLC: Implications for combinations with beta-blockers. Sci. Transl. Med. 2017, 9 (415), eaao4307. [PubMed: 29118262]
9. Herbst RS; Morgensztern D; Boshoff C, The biology and management of non-small cell lung cancer. Nature 2018, 553 (7689), 446–454. [PubMed: 29364287]
10. Jia Y; Yun CH; Park E; Ercan D; Manuia M; Juarez J; Xu C; Rhee K; Chen T; Zhang H; Palakurthi S; Jang J; Lelais G; DiDonato M; Bursulaya B; Michellys PY; Epple R; Marsilje TH; McNeill M; Lu W; Harris J; Bender S; Wong KK; Janne PA; Eck MJ, Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. Nature 2016, 534 (7605), 129–132. [PubMed: 27251290]
11. Uchibori K; Inase N; Araki M; Kamada M; Sato S; Okuno Y; Fujita N; Katayama R, Brigatinib combined with anti-EGFR antibody overcomes osimertinib resistance in EGFR-mutated non-small-cell lung cancer. Nat. Commun. 2017, 8, 14768. [PubMed: 28287083]
12. Gunther M; Juchum M; Kelter G; Fiebig H; Laufer S, Lung Cancer: EGFR Inhibitors with Low Nanomolar Activity against a Therapy-Resistant L858R/T790M/C797S Mutant. Angew. Chem. Int. Ed. 2016, 55 (36), 10890–10894.
13. Juchum M; Gunther M; Doring E; Sievers-Engler A; Lammerhofer M; Laufer S, Trisubstituted Imidazoles with a Rigidized Hinge Binding Motif Act As Single Digit nM Inhibitors of Clinically Relevant EGFR L858R/T790M and L858R/T790M/C797S Mutants: An Example of Target Hopping. J. Med. Chem. 2017, 60 (11), 4636–4656. [PubMed: 28482151]
14. Gunther M; Lategahn J; Juchum M; Doring E; Keul M; Engel J; Tumbrink HL; Rauh D; Laufer S, Trisubstituted Pyridinylimidazoles as Potent Inhibitors of the Clinically Resistant L858R/T790M/C797S EGFR Mutant: Targeting of Both Hydrophobic Regions and the Phosphate Binding Site. J. Med. Chem. 2017, 60 (13), 5613–5637. [PubMed: 28603991]
15. Park H; Jung HY; Mah S; Hong S, Discovery of EGF Receptor Inhibitors That Are Selective for the d746–750/T790M/C797S Mutant through Structure-Based de Novo Design. Angew. Chem. Int. Ed. 2017, 56 (26), 7634–7638.
16. Duong-Ly KC; Devarajan K; Liang S; Horiuchi KY; Wang Y; Ma H; Peterson JR, Kinase Inhibitor Profiling Reveals Unexpected Opportunities to Inhibit Disease-Associated Mutant Kinases. Cell Rep. 2016, 14 (4), 772–781. [PubMed: 26776524]

17. Duran-Frigola M; Mosca R; Aloy P, Structural systems pharmacology: the role of 3D structures in next-generation drug development. Chem. Biol. 2013, 20 (5), 674–684. [PubMed: 23706634]

18. 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 (5), e1003554. [PubMed: 24830652]

19. Zhao Z; Martin C; Fan R; Bourne PE; Xie L, Drug repurposing to target Ebola virus replication and virulence using structural systems pharmacology. BMC Bioinf. 2016, 17, 90.

20. Xie L; Draizen EJ; Bourne PE, Harnessing Big Data for Systems Pharmacology. Annu. Rev. Pharmacol. Toxicol. 2017, 57, 245–262. [PubMed: 27814027]

21. Park J; McDonald JJ; Petter RC; Houk KN, Molecular Dynamics Analysis of Binding of Kinase Inhibitors to WT EGFR and the T790M Mutant. J. Chem. Theory Comput. 2016, 12 (4), 2066–2078. [PubMed: 27010480]

22. Shan Y; Arkhipov A; Kim ET; Pan AC; Shaw DE, Transitions to catalytically inactive conformations in EGFR kinase. Proc. Natl. Acad. Sci. U. S. A. 2013, 110 (18), 7270–7275. [PubMed: 23576739]

23. Sutto L; Gervasio FL, Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. Proc. Natl. Acad. Sci. U. S. A. 2013, 110 (26), 10616–10621. [PubMed: 23754386]

24. Kannan S; Venkatachalam G; Lim HH; Surana U; Verma C, Conformational landscape of the epidermal growth factor receptor kinase reveals a mutant specific allosteric pocket. Chem. Sci. 2018, 9 (23), 5212–5222. [PubMed: 29997876]

25. Chong CR; Janne PA, The quest to overcome resistance to EGFR-targeted therapies in cancer. Nat. Med. 2013, 19 (11), 1389–1400. [PubMed: 24202392]

26. 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 (2), 146–159. [PubMed: 23438744]

27. Kornev AP; Taylor SS; Ten Eyck LF, A helix scaffold for the assembly of active protein kinases. Proc. Natl. Acad. Sci. U. S. A. 2008, 105 (38), 14377–14382. [PubMed: 18787129]

28. Kornev AP; Haste NM; Taylor SS; Eyck LF, Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. Proc. Natl. Acad. Sci. U. S. A. 2006, 103 (47), 17783–17788. [PubMed: 17095602]

29. James KA; Verkhivker GM, Structure-based network analysis of activation mechanisms in the ErbB family of receptor tyrosine kinases: the regulatory spine residues are global mediators of structural stability and allosteric interactions. PLoS One 2014, 9 (11), e113488. [PubMed: 25427151]

30. Liao JJ, Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. J. Med. Chem. 2007, 50 (3), 409–424. [PubMed: 17266192]

31. 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 (6), 1230–1241. [PubMed: 24730530]

32. Muller S; Chaikuad A; Gray NS; Knapp S, The ins and outs of selective kinase inhibitor development. Nat. Chem. Biol. 2015, 11 (11), 818–821. [PubMed: 26485069]

33. Miyashita O; Onuchic JN; Wolynes PG, Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. Proc. Natl. Acad. Sci. U. S. A. 2003, 100 (22), 12570–12575. [PubMed: 14566052]

34. Chen BY; Honig B, VASP: a volumetric analysis of surface properties yields insights into protein-ligand binding specificity. PLoS Comput. Biol. 2010, 6 (8), e1000881. [PubMed: 20814581]

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. 2016, 59 (9), 4326–4341. [PubMed: 26929980]

36. Kulagowski JJ; Blair W; Bull RJ; Chang C; Deshmukh G; Dyke HJ; Eigenbrot C; Ghilardi N; Gibbons P; Harrison TK; Hewitt PR; Liimatta M; Hurley CA; Johnson A; Johnson T; Kenny JR; Bir Kohli P; Maxey RJ; Mendonca R; Mortara K; Murray J; Narukulla R; Shia S; Steffek M; Ubhayakar S; Ultsch M; van Abbema A; Ward SI; Waszkowycz B; Zak M, Identification of

imidazo-pyrrolopyridines as novel and potent JAK1 inhibitors. J. Med. Chem. 2012, 55 (12), 5901–5921. [PubMed: 22591402]

37. Chaikuad A; Tacconi EM; Zimmer J; Liang Y; Gray NS; Tarsounas M; Knapp S, A unique inhibitor binding site in ERK1/2 is associated with slow binding kinetics. Nat. Chem. Biol. 2014, 10 (10), 853–860. [PubMed: 25195011]

38. Sonti R; Hertel-Hering I; Lamontanara AJ; Hantschel O; Grzesiek S, ATP Site Ligands Determine the Assembly State of the Abelson Kinase Regulatory Core via the Activation Loop Conformation. J. Am. Chem. Soc. 2018, 140 (5), 1863–1869. [PubMed: 29319304]

39. Marcou G; Rognan D, Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. J. Chem. Inf. Model. 2007, 47 (1), 195–207. [PubMed: 17238265]

40. Omura S; Iwai Y; Hirano A; Nakagawa A; Awaya J; Tsuchiya H; Takahashi Y; Asuma R, A new alkaloid AM-2282 of Streptomyces origin taxonomy, fermentation, isolation and preliminary characterization. J. Antibiot. 1977, 30 (4), 275–282. [PubMed: 863788]

41. Miduturu CV; Deng X; Kwiatkowski N; Yang W; Brault L; Filippakopoulos P; Chung E; Yang Q; Schwaller J; Knapp S; King RW; Lee JD; Herrgard S; Zarrinkar P; Gray NS, High-throughput kinase profiling: a more efficient approach toward the discovery of new kinase inhibitors. Chem. Biol. 2011, 18 (7), 868–879. [PubMed: 21802008]

42. Zhang J; Yang PL; Gray NS, Targeting cancer with small molecule kinase inhibitors. Nat. Rev. Cancer 2009, 9 (1), 28–39. [PubMed: 19104514]

43. Weisberg E; Boulton C; Kelly LM; Manley P; Fabbro D; Meyer T; Gilliland DG; Griffin JD, Inhibition of mutant FLT3 receptors in leukemia cells by the small molecule tyrosine kinase inhibitor PKC412. Cancer Cell 2002, 1 (5), 433–443. [PubMed: 12124173]

44. Zarrinkar PP; Gunawardane RN; Cramer MD; Gardner MF; Brigham D; Belli B; Karaman MW; Pratz KW; Pallares G; Chao Q; Sprankle KG; Patel HK; Levis M; Armstrong RC; James J; Bhagwat SS, AC220 is a uniquely potent and selective inhibitor of FLT3 for the treatment of acute myeloid leukemia (AML). Blood 2009, 114 (14), 2984–2992. [PubMed: 19654408]

45. Manley PW; Weisberg E; Sattler M; Griffin JD, Midostaurin, a Natural Product-Derived Kinase Inhibitor Recently Approved for the Treatment of Hematological MalignanciesPublished as part of the Biochemistry series “Biochemistry to Bedside”. Biochemistry 2018, 57 (5), 477–478. [PubMed: 29188995]

46. Yun CH; Mengwasser KE; Toms AV; Woo MS; Greulich H; Wong KK; Meyerson M; Eck MJ, The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc. Natl. Acad. Sci. U. S. A. 2008, 105 (6), 2070–2075. [PubMed: 18227510]

47. Pargellis C; Tong L; Churchill L; Cirillo PF; Gilmore T; Graham AG; Grob PM; Hickey ER; Moss N; Pav S; Regan J, Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. Nat. Struct. Biol. 2002, 9 (4), 268–272. [PubMed: 11896401]

48. Altschul SF; Madden TL; Schaffer AA; Zhang J; Zhang Z; Miller W; Lipman DJ, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997, 25 (17), 3389–3402. [PubMed: 9254694]

49. Ferrari AM; Wei BQ; Costantino L; Shoichet BK, Soft docking and multiple receptor conformations in virtual screening. J. Med. Chem. 2004, 47 (21), 5076–5084. [PubMed: 15456251]

50. Berman HM; Westbrook J; Feng Z; Gilliland G; Bhat TN; Weissig H; Shindyalov IN; Bourne PE, The Protein Data Bank. Nucleic Acids Res. 2000, 28 (1), 235–242. [PubMed: 10592235]

51. The UniProt C, UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017, 45 (D1), D158–D169. [PubMed: 27899622]

52. 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 (12), i305–i312. [PubMed: 19478004]

53. 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 (14), 5441–5446. [PubMed: 18385384]

54. 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.

55. Jain AN, Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J. Med. Chem. 2003, 46 (4), 499–511. [PubMed: 12570372]

56. Spitzer R; Jain AN, Surflex-Dock: Docking benchmarks and real-world application. J. Comput.-Aided Mol. Des. 2012, 26 (6), 687–699. [PubMed: 22569590]

57. Eswar N; Webb B; Marti-Renom MA; Madhusudhan MS; Eramian D; Shen MY; Pieper U; Sali A, Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinf. 2006, Chapter 5, Unit-5.6.

58. Harvey MJ; Giupponi G; Fabritiis GD, ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. J. Chem. Theory Comput. 2009, 5 (6), 1632–1639. [PubMed: 26609855]

59. Vanommeslaeghe K; Hatcher E; Acharya C; Kundu S; Zhong S; Shim J; Darian E; Guvench O; Lopes P; Vorobyov I; Mackerell AD, Jr., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem. 2010, 31 (4), 671–690. [PubMed: 19575467]

60. Brooks BR; Brooks III CL; Mackerell AD, Jr.; Nilsson L; Petrella RJ; Roux B; Won Y; Archontis G; Bartels C; Boresch S; Caflisch A; Caves L; Cui Q; Dinner AR; Feig M; Fischer S; Gao J; Hodoscek M; Im W; Kuczera K; Lazaridis T; Ma J; Ovchinnikov V; Paci E; Pastor RW; Post CB; Pu JZ; Schaefer M; Tidor B; Venable RM; Woodcock HL; Wu X; Yang W; York DM; Karplus M, CHARMM: the biomolecular simulation program. J. Comput. Chem. 2009, 30 (10), 1545–1614. [PubMed: 19444816]

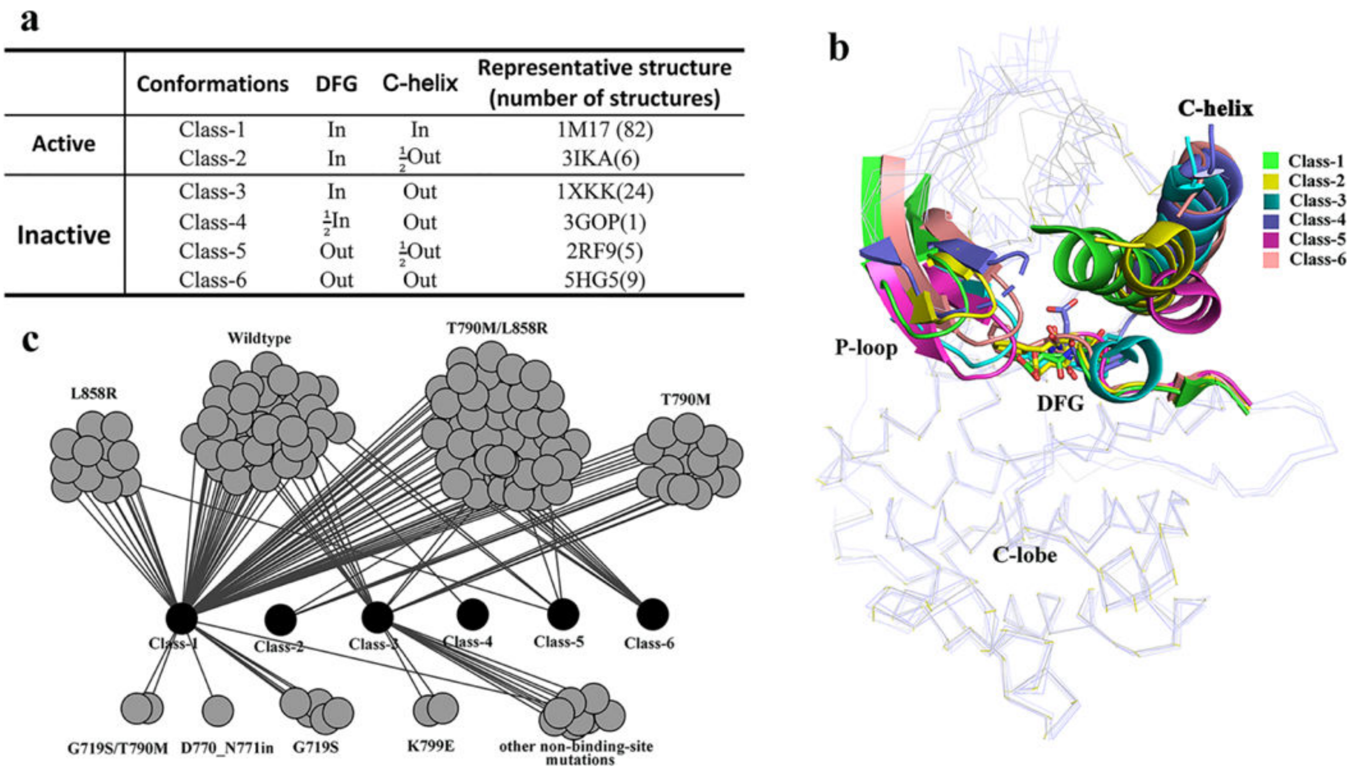
61. Feig M; Karanicolas J; Brooks III CL, MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J. Mol. Graphics Model. 2004, 22 (5), 377–395.

62. 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. [PubMed: 28628649]

63. Zhao Z; Liu Q; Bliven S; Xie L; Bourne PE, Determining Cysteines Available for Covalent Inhibition Across the Human Kinome. J. Med. Chem. 2017, 60 (7), 2879–2889. [PubMed: 28326775]

64. Desaphy J; Raimbaud E; Ducrot P; Rognan D, Encoding protein-ligand interaction patterns in fingerprints and graphs. J. Chem. Inf. Model. 2013, 53 (3), 623–637. [PubMed: 23432543]

65. Galili T, dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 2015, 31 (22), 3718–3720. [PubMed: 26209431]



	Conformations	DFG	C-helix	Representative structure (number of structures)
Active	Class-1	In	In	1M17 (82)
	Class-2	In	$\frac{1}{2}$Out	3IKA(6)
Inactive	Class-3	In	Out	1XKK(24)
	Class-4	$\frac{1}{2}$In	Out	3GOP(1)
	Class-5	Out	$\frac{1}{2}$Out	2RF9(5)
	Class-6	Out	Out	5HG5(9)

Figure 1.

EGFR kinase structure dataset with mutations. (a) Classes of EGFR kinase domain conformations. (b) Comparison of different classes of EGFR kinase with representative conformations. (c) Network of currently released structure conformations and mutations.

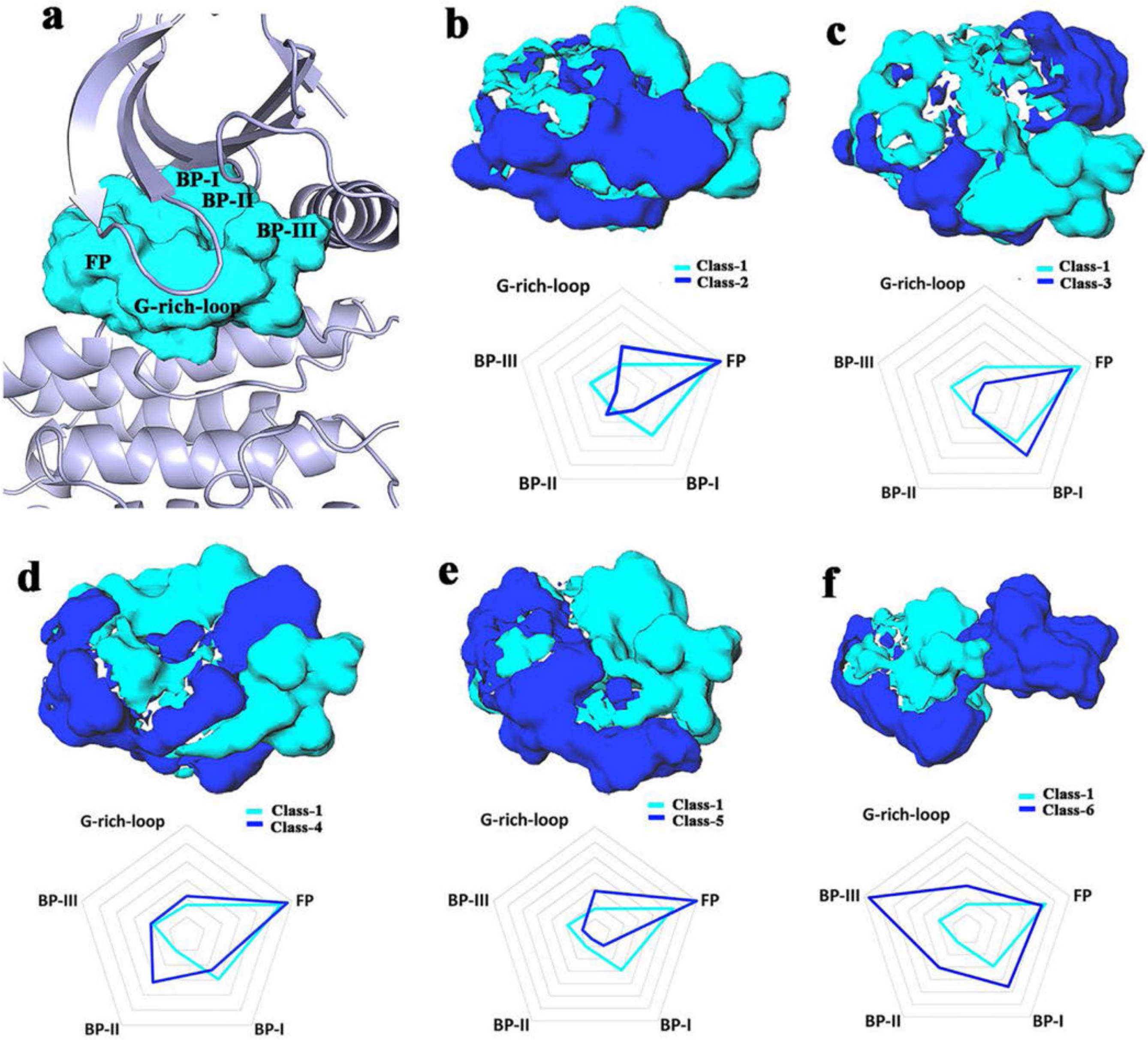


Figure 2.

Differences in binding pockets. The cavity of Class-1 is showed in cyan and the cavity of each other class in blue, respectively. (a) The Class-1 binding pocket shown in PDB 1m17 as a reference showing every region. (b-f) Significant pairwise differences (upper); quantitative difference of every pairwise binding pocket illustrated using a radar chart and a contour interval of 100 Å^3 (lower).

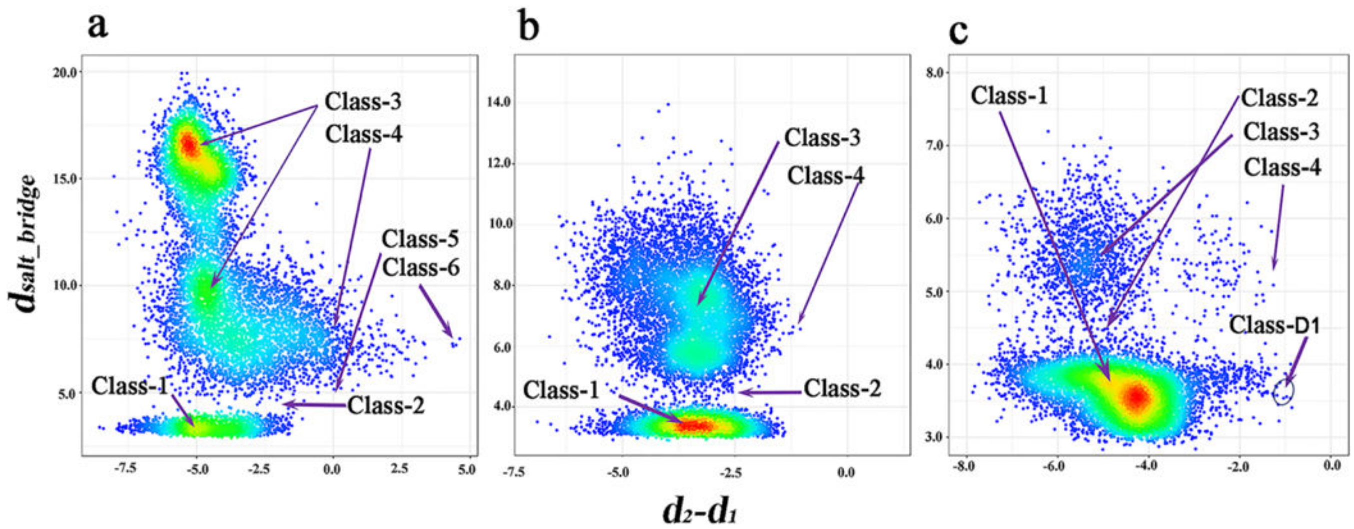


Figure 3.

Scatter density plot of the distributions of conformation space for the different systems (a) wildtype, (b) del746–750/T790M/C797S and (c) L858R/T790M/C797S. The y axis represents the distance of K745 Nζ : E762 Cδ and the x axis represents the flexibility of the DFG peptide. d_2-d_1=d_2[C797 Cα : D855 Cγ]−d_1[C797 Cα : F856 Cζ]. High-density and low-density regions are colored in red and blue, respectively.

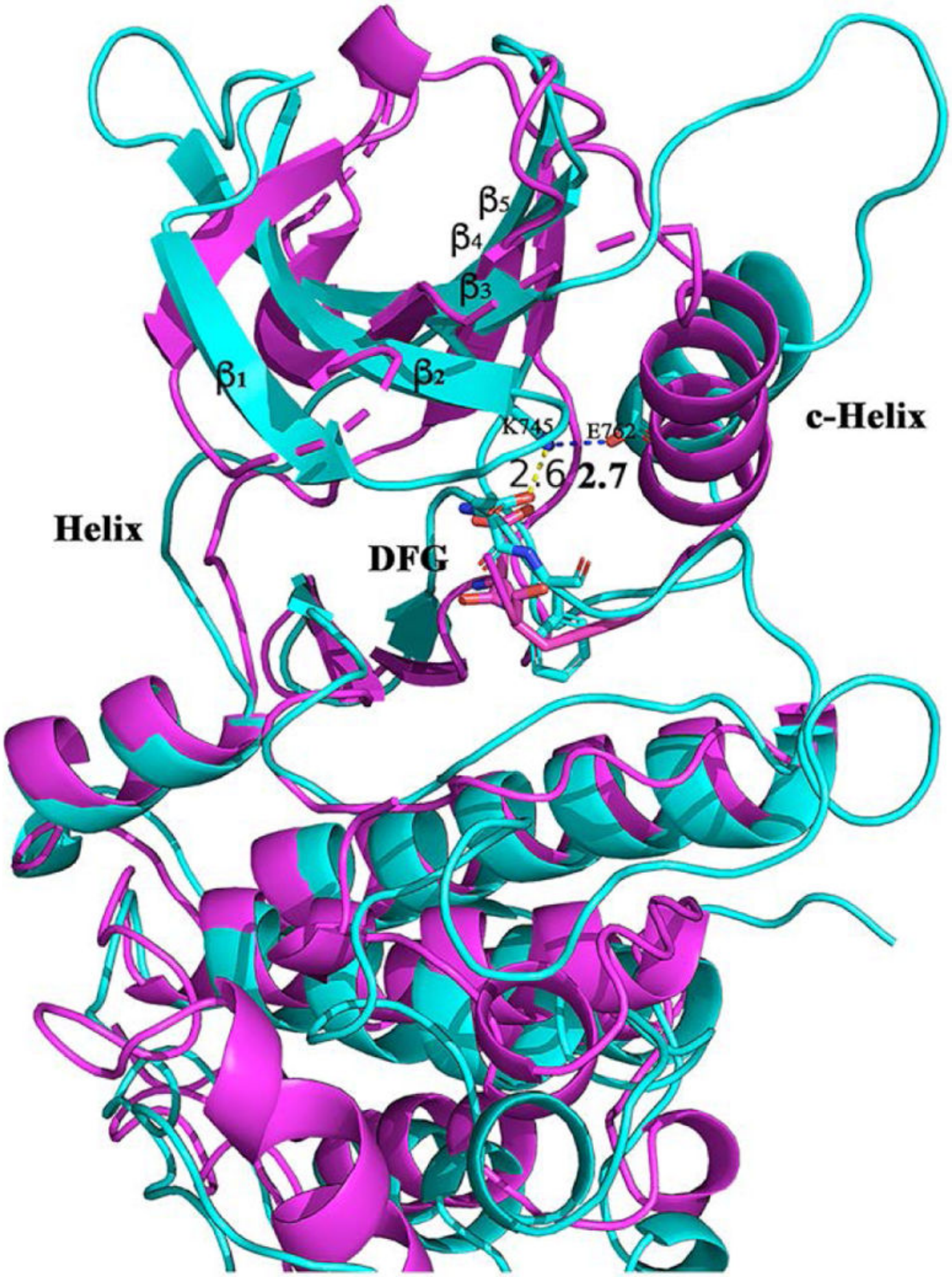


Figure 4.
The Class-D1 of conformation (cyan color) comparing with that of Class-4 in purple color.

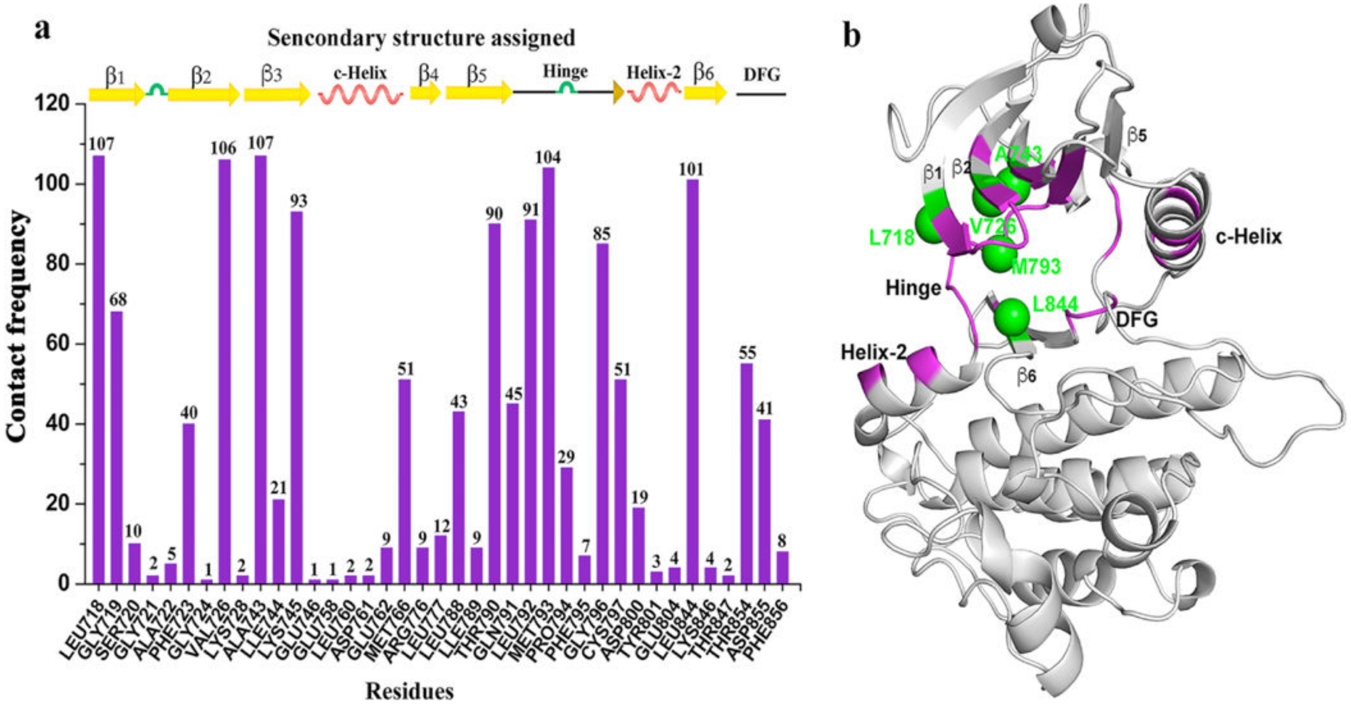


Figure 5.

Interaction details between the amino acids that constitute the binding site and the corresponding ligand. (a) the contact frequency of each amino acid providing the binding interaction and the corresponding locations. (b) distribution of interacting amino acids in the 3D binding site (purple) and the top 5 residues with the highest contact times (green).

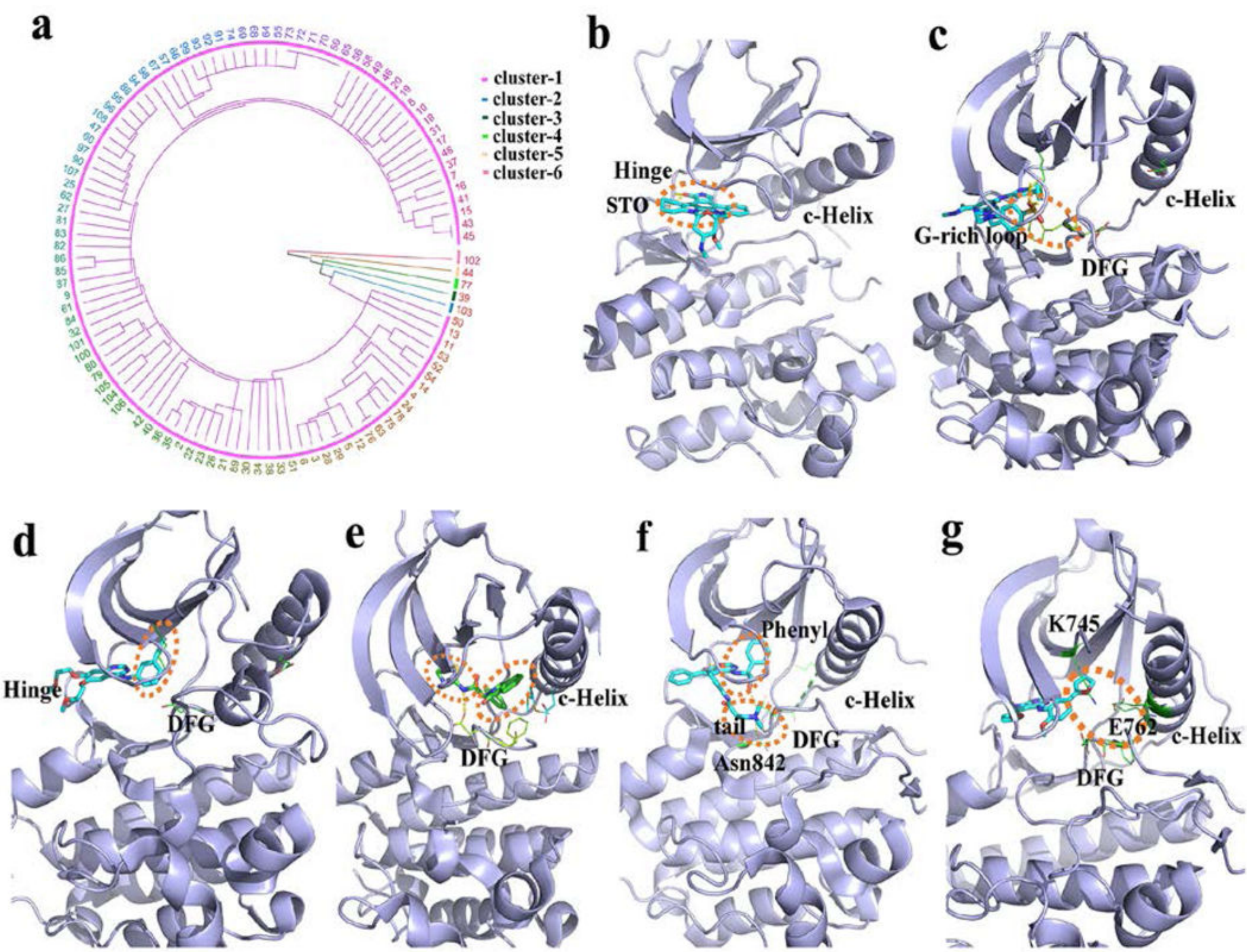


Figure 6.

EGRF ligand-binding clusters and features. (a) hierarchical cluster analysis. The number represents the index of every structure (Table S5). (b-g) the clustered binding modes (clusters 1–6) are shown using the PDB structures 2itu, 5u8l, 4hjo, 5d41, 4jq8 and 4jeb, respectively. The dash-line circles (orange) highlight the binding characteristics of the core binding pocket in: (b) the location of binding G-rich loop and DFG motif; (c) the hydrophobic pocket; (d) the hydrophobic pocket and the allosteric site; (e) the hydrophobic pocket and the location of binding DFG and Asn842; (f) and the combined binding position with the interactions from the hydrophobic pocket and the allosteric site (g).