

## **INTEGRATING GENETIC AND STRUCTURAL DATA ON HUMAN PROTEIN KINOME IN NETWORK-BASED MODELING OF KINASE SENSITIVITIES AND RESISTANCE TO TARGETED AND PERSONALIZED ANTICANCER DRUGS**

GENNADY M. VERKHIVKER†

*Department of Computational Biosciences, Schmid College of Science & Technology, Chapman University, One University Drive, Orange CA 92866, USA*

*Department of Pharmacology, University of California San Diego, 9500 Gilman Drive, San Diego CA 92093, USA*

*Email: verkhivk@chapman.edu*

The human protein kinome presents one of the largest protein families that orchestrate functional processes in complex cellular networks, and when perturbed, can cause various cancers. The abundance and diversity of genetic, structural, and biochemical data underlies the complexity of mechanisms by which targeted and personalized drugs can combat mutational profiles in protein kinases. Coupled with the evolution of system biology approaches, genomic and proteomic technologies are rapidly identifying and characterizing novel resistance mechanisms with the goal to inform rationale design of personalized kinase drugs. Integration of experimental and computational approaches can help to bring these data into a unified conceptual framework and develop robust models for predicting the clinical drug resistance. In the current study, we employ a battery of synergistic computational approaches that integrate genetic, evolutionary, biochemical, and structural data to characterize the effect of cancer mutations in protein kinases. We provide a detailed structural classification and analysis of genetic signatures associated with oncogenic mutations. By integrating genetic and structural data, we employ network modeling to dissect mechanisms of kinase drug sensitivities to oncogenic EGFR mutations. Using biophysical simulations and analysis of protein structure networks, we show that conformational-specific drug binding of Lapatinib may elicit resistant mutations in the EGFR kinase that are linked with the ligand-mediated changes in the residue interaction networks and global network properties of key residues that are responsible for structural stability of specific functional states. A strong network dependency on high centrality residues in the conformation-specific Lapatinib-EGFR complex may explain vulnerability of drug binding to a broad spectrum of mutations and the emergence of drug resistance. Our study offers a systems-based perspective on drug design by unravelling complex relationships between robustness of targeted kinase genes and binding specificity of targeted kinase drugs. We discuss how these approaches can exploit advances in chemical biology and network science to develop novel strategies for rationally tailored and robust personalized drug therapies.

---

† This work is partly supported by funding from Chapman University.

## 1. Background

The era of significant scientific breakthroughs and technological advancements in genetics and biology has brought to clinical settings personalized health care that has the capacity to detect the onset of disease at its earliest stages and preempt the progression of disease. The comprehensive cancer genome characterization efforts have refined our understanding of specified genes responsible for development and progression of tumors<sup>1</sup>. Several malignancies are associated with the mutation or increased expression of protein kinases, including lung, breast, stomach, colorectal, head and neck, and pancreatic carcinomas and glioblastoma<sup>2</sup>. Tumor sequencing efforts have identified a rich source of naturally occurring mutations with many being simple single nucleotide polymorphisms (SNPs) in protein kinases. A subset of these SNPs occurs in the coding regions (cSNPs) of kinases and result in a change in the encoded amino acid sequence (nonsynonymous coding SNP; nscSNPs). Genome studies have revealed the importance of “driver” somatic alterations that activate crucial oncoproteins such as *EGFR*, *BCR-ABL*, and other kinase genes. Mutations in these protein kinases are often implicated in many cancers and exemplify the phenomenon of ‘oncogene addiction,’ according to which the effects of driver genomic alterations are pivotal for tumor proliferation and have a selective advantage for the formation of the tumor during somatic cell replication<sup>3</sup>. Oncogene dependencies induced by genetic alterations in *BCR-ABL*, *KIT*, *EGFR* and other kinase genes are well known and have provided decisive clinical proof of principle for the genomics-informed drug discovery of kinase drugs<sup>4</sup>. Although tumor dependencies driven by dominant oncogenes could respond to targeted therapies, clinical responses to single agents are often followed by the development of drug resistance. The tumor dependency concept is especially relevant to understand mechanisms of acquired resistance, where resistant mutations, seemingly developed due to drug treatment, may instead represent evolutionary selection of cell subpopulations which harbor preexistent somatic mutant variants which confers a primary resistance to these cells and provides them with a selective advantage. The spectrum of lung cancer EGFR mutations can induce oncogenic transformation by leading to constitutive kinase activity of EGFR and confer markedly different sensitivity to EGFR inhibitors<sup>5</sup>. The most common reported mutations are the deletion of five exon-19 residues and the exon-21 substitution L858R in the catalytic domain of EGFR<sup>6</sup>. Together, these mutations correspond to more than 90% of the activating EGFR mutations observed in non-small-cell lung cancer (NSCLC). While T790M has only a modest effect on EGFR function, a tandem of T790M and L858R mutations can result in a dramatic enhancement of EGFR activity. More than 200 activating and drug resistance EGFR mutations with different clinical responses to tyrosine kinase inhibitors have been reported<sup>7</sup> and molecular mechanisms of mutation-induced kinase activation have been extensively discussed<sup>8</sup>.

Gefitinib and Erlotinib are orally effective protein-kinase targeted inhibitors that are used in the treatment of ERBB1/EGFR-mutant lung cancer. Afatinib is another EGFR-

targeted kinase drug approved by the FDA for the first-line treatment of patients with metastatic NSCLC whose tumors have *EGFR* exon 19 deletions or exon 21 (L858R) substitutions. Lapatinib, a small molecule tyrosine kinase inhibitor of both EGFR and HER2/ErbB2 is now also approved for advanced HER2-amplified breast cancer<sup>9</sup>. Structural and biochemical studies have characterized the inhibition of intrinsic catalytic activity of EGFR and HER2/ErbB2 variants by Lapatinib using a diverse array of enzymatic and cell-based assays<sup>10,11</sup>. Cell-based EGFR resistance mutation screens have demonstrated that Lapatinib produced the broadest mutation spectra of any of the EGFR-targeted drugs tested in *in vitro* system, with a number of Lapatinib-specific resistant mutations clustered around the selectivity pocket and the EGFR-A-loop<sup>12</sup>. The association between EGFR mutations and differential drug sensitivity suggested that genetic EGFR alterations and corresponding changes in structural and interaction profiles of the EGFR kinase domain render tumors sensitive to selective inhibitors. Oncogenic kinases can adopt different mechanisms to alleviate negative regulatory processes associated with their intrinsic conformational instability. One of them is the recruitment of unstable kinase forms to the Hsp90 system that protects abnormally activated kinases in cancer cells<sup>13</sup>. HSP90 stabilizes viral kinases and various mutated oncogenes, including oncogenic EGFR mutants that are dependent on the chaperoning function through direct interactions to maintain their stability<sup>14</sup>. HSP90 inhibition reduces mutant EGFR levels and activity, suggesting a viable EGFR inhibition strategy. Crystallographic studies<sup>15</sup> have supported this mechanism by showing that the catalytic domains of the EGFR-L858R and EGFR-L858R/T790M oncogenic mutants can adopt flexible inactive conformations that may facilitate conformational release from the autoinhibitory state. This may be exploited by the Hsp90 chaperone to bind the unstable mutant conformations and promote an accumulation of a constitutively active form. According to the newly emerging paradigm, kinase inhibitors may exert their primary effect by “arresting” the kinase domain in the specific inactive form, thereby depriving the Hsp90 system from access to unstable conformational states and preventing uncontrollable accumulation of the active form<sup>16</sup>.

The abundance and diversity of genetic, structural, and biochemical data underlies the complexity of mechanisms by which targeted and personalized kinase agents can combat mutational profiles in EGFR kinase. We employ a battery of synergistic computational approaches that integrate genetic, biochemical, and structural data to characterize the effect of cancer mutations in protein kinases. We show that binding specificity and drug resistance of EGFR drugs may be linked with the global network properties of key residues that are responsible for structural stability of specific targeted conformations. The results of this study offer a network-based perspective on drug design of targeted and personalized kinase drugs, showing how the efficiency and robustness of the interaction networks may be associated with kinase binding preferences and emergence of resistant mutations.

## 2. Methods

### 2.1. *Data mining*

Protein kinase sequences were obtained from Kinbase (<http://kinase.com/kinbase/>). Common SNPs were retrieved from PupaSNP and dbSNP using the Ensembl data mining tool, Biomart ([http://www.ensembl.org/Homo\\_sapiens/martview](http://www.ensembl.org/Homo_sapiens/martview)). The disease causing SNPs were retrieved from OMIM, KinMutBase, and HGMD databases. We used all kinase gene entries referenced in NCBI and SwissProt database, and 7955 unique SNP entries corresponding to these kinase genes as they are referenced in NCBI. These unique SNP entries include 3722 synonymous, 3985 missense, 75 nonsense and 173 frameshift mutations. We have also gathered 780 OMIM variant entries from NCBI and 3542 SwissProt variant entries. Cancer mutations were retrieved from OMIM and COSMIC databases. Motif-based alignments of kinase sequences to the catalytic core were first generated by implementation of the Gibbs motif sampling method. This method identifies characteristic motifs for each individual subdomain of the kinase catalytic core, which are then used to generate high-confidence motif-based Markov chain Monte Carlo multiple alignments based on these motifs<sup>17</sup>. The nsSNPs were then mapped to the kinase catalytic domain in accordance with this alignment. Cancer driver predictions were performed by using the SVM approach as described in the earlier work<sup>18</sup>.

### 2.2. *Somatic mutation distributions and driver mutation hotspots in protein kinome*

Functionally important subdomains of the kinase catalytic core were examined to determine the distribution of nsSNPs and identify structurally conserved hotspots of functionally important mutations. The number of SNPs in each of the subdomains was calculated from the structure-informed multiple sequence alignment. The expected probability  $E(p)$  of a SNP occurring in a kinase subdomain region was calculated separately for each SNP type. In brief, the average length of each region was calculated as the weighted average of the region length in each kinase considered, where weights correspond to the total number of SNPs occurring within each kinase. The probability of a SNP occurring within a particular region purely by chance was computed as its weighted average length over the sum of every region's weighted average length. The probability (p-value) of the observed total number ( $x$ ) of SNPs occurring within each region, where  $n$  is the total number of SNPs considered, was calculated using the general binomial distribution. The average length of each subdomain was calculated as the weighted average of the region length in each kinase considered, where weights correspond to the total number of SNPs occurring within each kinase. The probability of a SNP occurring within a particular region purely by chance was computed as its weighted average length divided by the sum of every region's weighted average length. The probability (p-value) of the observed total number of SNPs occurring within each region was then calculated using the general binomial distribution. Cancer

mutant predictions and analysis were performed as described in previous studies<sup>21</sup>. A support vector machine (SVM) was trained upon common SNPs (presumed neutral) and congenital disease causing SNPs characterized by a variety of sequence, structural, and phylogenetic parameters. The threshold taken for calling a SNP a driver is 0.49 for catalytic domain mutations, and 0.53 for all other mutations.

### 2.3. Network modeling of residue interaction networks in protein kinases

Molecular dynamics (MD) simulations were carried out using NAMD 2.6 with the CHARMM27 force field<sup>19</sup>. The binding free energies and computational alanine scanning of kinase-drug complexes were done using MM-GBSA approach<sup>20</sup>. A graph-based representation of proteins was used in the protein structure network analysis, where residues were considered as nodes and edges correspond to the nonbonding residue-residue interactions. The pair of residues with the interaction strength  $I_{ij}$  greater than a user-defined cut-off  $I_{\min}$  are connected by edges and produce a protein structure network graph for a given interaction strength  $I_{\min}$ . The strength of interaction between two amino acid side chains is

$$I_{ij} = \frac{n_{ij}}{\sqrt{(N_i \times N_j)}} \times 100 \quad (1)$$

where  $n_{ij}$  is number of distinct atom pairs between the side chains of amino acid residues  $i$  and  $j$  that lie within a distance of 4.5 Å.  $N_i$  and  $N_j$  are the normalization factors for residues  $i$  and  $j$  respectively<sup>21</sup>. We considered any pair of residues to be connected if  $I_{\min}$  was greater than 3.0%. A weighted network representation of the protein structure is adopted that includes non-covalent connectivity of side chains and residue cross-correlation fluctuation matrix<sup>22</sup>. In this model, the weight  $w_{ij}$  of an edge between nodes  $i$  and  $j$  is measured as  $w_{ij} = -\log(|C_{ij}|)$  where  $C_{ij}$  is the element of the covariance matrix measuring the cross-correlation residue fluctuations obtained from MD simulations. The shortest paths between two residues are determined using the Floyd-Warshall algorithm. We computed the residue-based betweenness which is defined as the sum of the fraction of shortest paths between all pairs of residues that pass through residue  $i$ :

$$C_b(n_i) = \sum_{j < k} \frac{g_{jk}(i)}{g_{jk}} \quad (2)$$

where  $g_{jk}$  denotes the number of shortest geodesics paths connecting  $j$  and  $k$ , and  $g_{jk}(i)$  is the number of shortest paths between residues  $j$  and  $k$  passing through the node  $n_i$ .

### 3. Results

#### 3.1. *Structural and functional signatures of cancer mutations in protein kinases*

Genetic variations in protein kinase genes are widely spread across both phylogenetic and structural space, and only a subset of all SNPs could be directly mapped to the kinase catalytic domain (Figure 1A). We constructed the distribution of various SNPs categories that could be mapped onto the 12 functional subdomains (SDs) of the kinase catalytic core (Figure 1B). Structural mapping of sSNPs resulted in a uniform coverage of kinase subdomains, showing only a weak preference towards SD II which has no obvious functional role in kinase regulation. The distribution of nsSNPs pointed to the preferential bias towards specific functional regions. Functionally important P-loop (SD I), hinge region (SD V), catalytic loop (SD VIB), and A-loop (SD VII) along with the P+1 loop region (SD VIII) are more densely populated. The catalytic domain of protein kinases harbors a large number of SNPs falling into three major categories: common and neutral SNPs; inherited disease causing germline SNPs; and cancer causing SNPs. By compiling and mapping a total of 355 common SNPs, 428 inherited disease causing SNPs, and 541 cancer associated SNPs we found a statistically significant enrichment of different categories of SNPs in specific regions of the catalytic domain (Figure 1C). Common nsSNPs are randomly distributed within the catalytic core, only sparsely populating functional segments of the catalytic core, such as the catalytic or A-loops, whereas these nsSNPs more densely occupy evolutionary unconserved regions of the C-terminal tail. The disease-causing nsSNPs primarily mapped to the regions involved in regulation and substrate binding, such as the APE-loop and the P+1 region, as well as the catalytic loop (Figure 1C). Cancer-associated nsSNPs tend to target regions directly involved in the catalytic activity that are mainly localized in the P-loop, A-loop and catalytic loop. The distribution of kinase nsSNPs across functional kinase subdomains suggested that the kinase regions that are enriched in different types of SNPs are markedly different and have only a minimal overlap. The distribution revealed a preference for cancer-causing nsSNPs to populate primarily the A-loop (SDVII) and the P-loop (SD I). The functionally important for substrate and protein binding P+1 loop are enriched largely in disease-associated mutations, but not cancer-causing mutations. These results indicated that disease-associated mutations could primarily affect the kinase regions involved in functional regulation, allosteric interactions and substrate binding<sup>23</sup>.

Kinome-wide analysis of sequence and structure-based signatures of cancer mutations revealed that a significant number of cancer mutations could fall at structurally equivalent positions within the catalytic core. These structurally conserved mutations tend to cluster into specific mutational hotspots which may be shared by multiple kinase genes. We classified cancer mutation hotspots which had been identified as a frequent target of tumorigenic activating mutations. Cancer mutation hotspots in protein kinases are largely localized within the P-loop, hinge region, and A-loop (Figure 1).

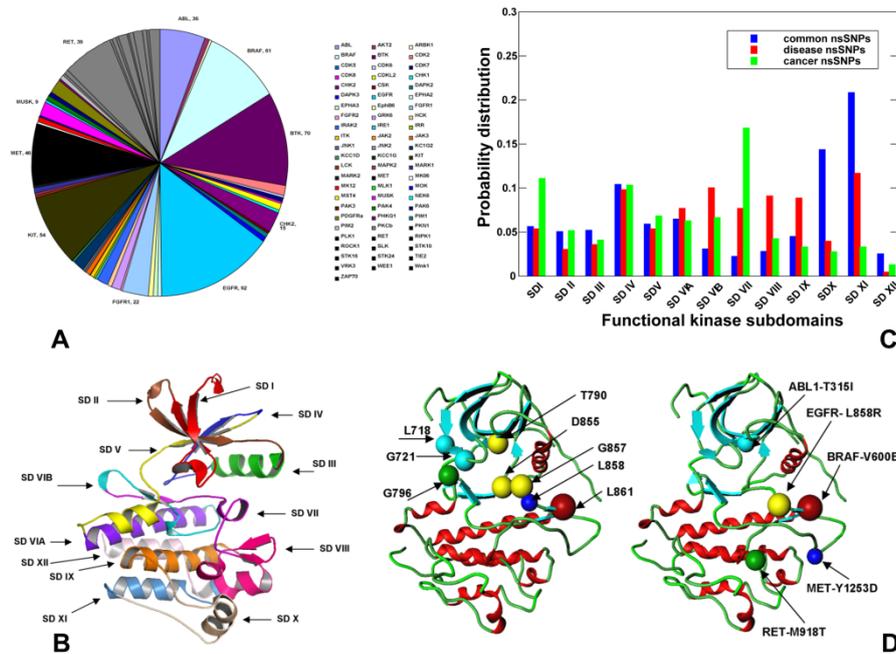


Figure 1. The distribution of nsSNPs in the catalytic core (A,C). The catalytic domain was subdivided into 12 subdomains (B) with some subdomains corresponding to functional regions : SD I (P-loop); SDIII( $\alpha$ C-helix); SDV (hinge region); SDVIB (catalytic loop); SDVII (A-loop) ; SDVIII (P+I loop). (B)Structural mapping is shown for common nsSNPs , disease-causing nsSNPs , and cancer-causing nsSNPs. (D) Structural localization of driver mutations is mapped onto the crystal structure of the active EGFR (pdb entry 2J6M). Structural annotation of cancer driver mutations is arranged according to their oncogenic potential. The higher the oncogenic potential of the cancer drive, the larger the ball denoting structural position of the respective mutation.

### 3.2 Structural bioinformatics analysis of oncogenic kinase mutants: distinct structural signatures of Hsp90-dependent kinase clients are associated with oncogenic potential

Oncogenic kinase mutants may rely on the Hsp90 dependence for the maintenance of stability and accumulation of the constitutively active form. In particular, Hsp90 function is essential to maintain high-level expression of mutant EGFR in lung cancer cells<sup>14</sup>. We performed kinome-wide structural bioinformatics analysis of chaperone-regulated kinases (Figure 2). The proteomics-based client annotation (Figure 2A) was compared against structure-based mapping of the Hsp90-Cdc37 kinase clients (Figure 2b). Structural coupling of the catalytic DFG motif and the regulatory  $\alpha$ C-helix is recognized as central in controlling kinase activity and dynamic equilibrium between the inactive (DFG-out/ $\alpha$ C-helix-in), the Cdk/Src-like inactive (DFG-in/ $\alpha$ C-helix-out) and the active kinase forms (DFG-in/ $\alpha$ C-helix-in). Although many of the Hsp90 kinase clients can occupy evolutionary different branches

of the human kinome, we found they share a common Cdk/Src- type structural arrangement of their inactive functional states. The Cdk/Src-like inactive structures shared by the Hsp90 kinase clients are unified by a common structural determinant whereby the regulatory  $\alpha$ C-helix is moved to a  $\alpha$ C-out conformation and forms autoinhibitory clamp with the A-loop, thus preventing the formation of the catalytically competent active kinase.

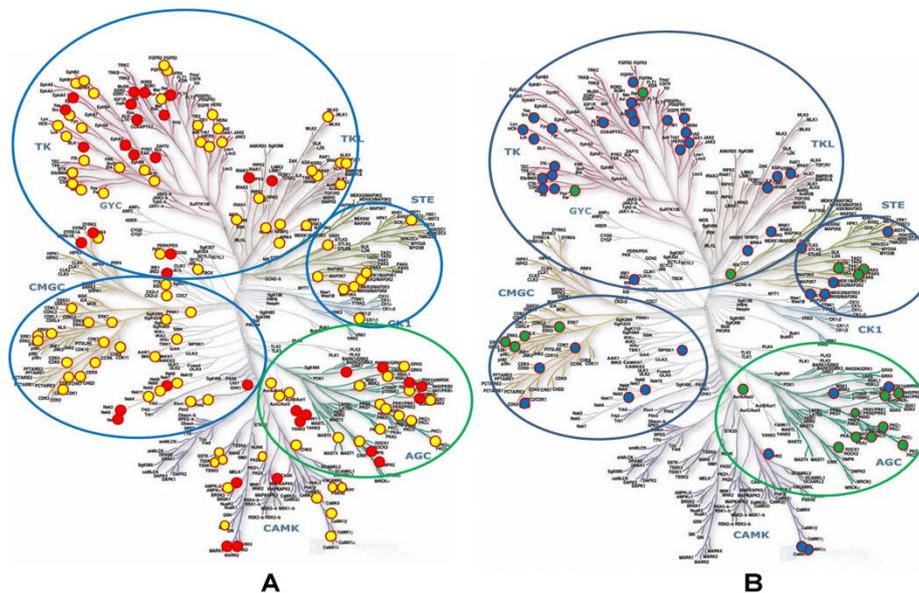


Figure 2. The distribution of the Hsp90-dependent protein kinase clients in the human kinome. (A) Kinome mapping of Hsp90-Cdc37 clients discovered in proteomic-based studies<sup>16</sup> is depicted. The kinases that are found to be downregulated by Hsp90 inhibition in the experimental profiling are shown in yellow (confirmed kinase clients) and red (novel kinase clients from proteomics studies<sup>16</sup>). (B) Structure-based mapping of the Hsp90-Cdc37 kinase clients. The Cdk/Src kinase clients are marked in blue filled spheres. A high density of the Cdk/Src clients in the TK, TKL, STE, CAMK, and CMGC groups of the human kinome tree is highlighted by blue circles. The second category of kinase clients is characterized by active structures stabilized through allosteric interactions (green spheres).

According to our analysis, oncogenic kinase mutations in the conserved hotspots (A-loop), may perturb the constraints keeping the  $\alpha$ C-helix-out in the rigid inactive position, and allow the A-loop to assume an extended active conformation (A-loop open) that is seen in the crystal structures of the EGFR-L858R and EGFR-L858R/T790M mutants<sup>15</sup>. These Cdk/Src-like active conformations that can be adopted by oncogenic mutants are far more flexible and unstable. As a result, they may be sequestered by the Hsp90 to promote uncontrollable transformation and accumulation of the constitutively active state for kinase cancer mutants.

### ***3.3 Integrating genetic and structural data on oncogenic EGFR mutations: modeling of thermodynamic and networking signatures of targeted drug binding***

By using MD simulations and MM-GBSA binding free energy simulations, we evaluated the thermodynamic effect of oncogenic EGFR mutations on different conformational states of EGFR (Figure 3A). Our results showed that oncogenic mutations L747P, L747S, L858R and L861Q can destabilize the rigid autoinhibitory structure that is thermodynamically favorable in the wild-type EGFR<sup>24</sup>. Strikingly, oncogenic mutations L747P/S, L858R and L861Q seemed to favor a highly flexible Cdk/Src –active conformation and marginally destabilize the active conformation. As a result, EGFR mutations with a high oncogenic potential may destabilize the dormant autoinhibitory structure. These mutations may induce fast equilibrium between flexible Cdk/Src-like active conformation and active structure that could lead to uncontrollable activity, which is a “deadly” signature of cancer mutations. The major Lapatinib-resistant mutations with the high oncogenic potential occurred in the residues that do not directly contact ligand. L747 is located at a loop adjacent to  $\alpha$ C-helix; V765 and V769 are at or near the C-terminal portion of  $\alpha$ C-helix, and T790 is at the gatekeeper position in the ATP binding site. Of the remainder, N857 is located in helix D, T854 forms the base of the ATP binding site, L858 and H870 are in the A-loop (Figure 3). To determine the thermodynamic contribution of the EGFR residues to Lapatinib binding and identify energetic hot spots susceptible to mutations, we performed free energy simulations and computational alanine scanning (Figure 3B). First, we found that only some Lapatinib-interacting residues corresponded to cancer mutation hotspots, suggesting that escaping binding interactions with the drug via mutations may not be a primary mechanism that drives emergence of Lapatinib-resistant mutations. The energetic hot spots of Lapatinib binding that corresponded to cancer mutation drug-resistant EGFR sites included L718, L777, L788, T790 (gate-keeper), and T854 residues. However, the EGFR mutations of high oncogenic potential that can render differential sensitivity to Lapatinib such as L747, L858, and L861 make fairly moderate contributions to binding energetics that could not explain high resistance. These results suggested that the mechanism of Lapatinib-induced somatic mutations may rather be associated with the intrinsic stability of the Cdk/Src inactive EGFR structure that binds Lapatinib<sup>10-12</sup>. Several hypotheses have suggested that the mechanism of Lapatinib-induced somatic mutations is linked with a conformation-specific mode of Lapatinib binding to an inactive EGFR structure<sup>11,12</sup> as drug resistant cancer mutations may stabilize the constitutively active EGFR form and thus interfere with the drug binding. To test this mechanism, we evaluated organization of the residue interaction networks and structural stability of EGFR states. The stability of the inactive EGFR conformation targeted by Lapatinib is mediated by interaction networks formed by high centrality residues F723 (P-loop),  $\alpha$ C-helix (V765, M766, and V769), the  $\alpha$ C- $\beta$ 4-loop (L774), the HRD motif (H835, D837), DFG motif (D855, F856) and L858 (A-loop) (Figure 3C, Table 1). The central result of the network analysis showed that although some somatic mutations may emerge in residues

with medium centrality, Lapatinib-resistant cancer mutations can be developed in high centrality sites that determine interaction network of the specific EGFR form (Table 1). Due to their central position in the structural network, mutations of V765 and V769 ( $\alpha$ C-helix) and L858 (A-loop) can severely compromise the integrity of the interaction network by weakening or dissolving the central autoinhibitory lock between the P-loop/A-loop interactions holding the  $\alpha$ C-helix in the inactive position. Targeted mutations of these high centrality sites could disrupt allosteric coupling between functional regions, leading to the weakening and fragmentation of the residue interaction network. A strong network dependency on high centrality residues may explain a broad spectrum of Lapatinib-resistant mutations that are located away from the inhibitor, near the  $\alpha$ C-helix and in the A-loop. Hence, residue centrality may be used as a metric for assessing severity of drug resistance mutations and differentiating between highly resistant and moderately resistant positions.

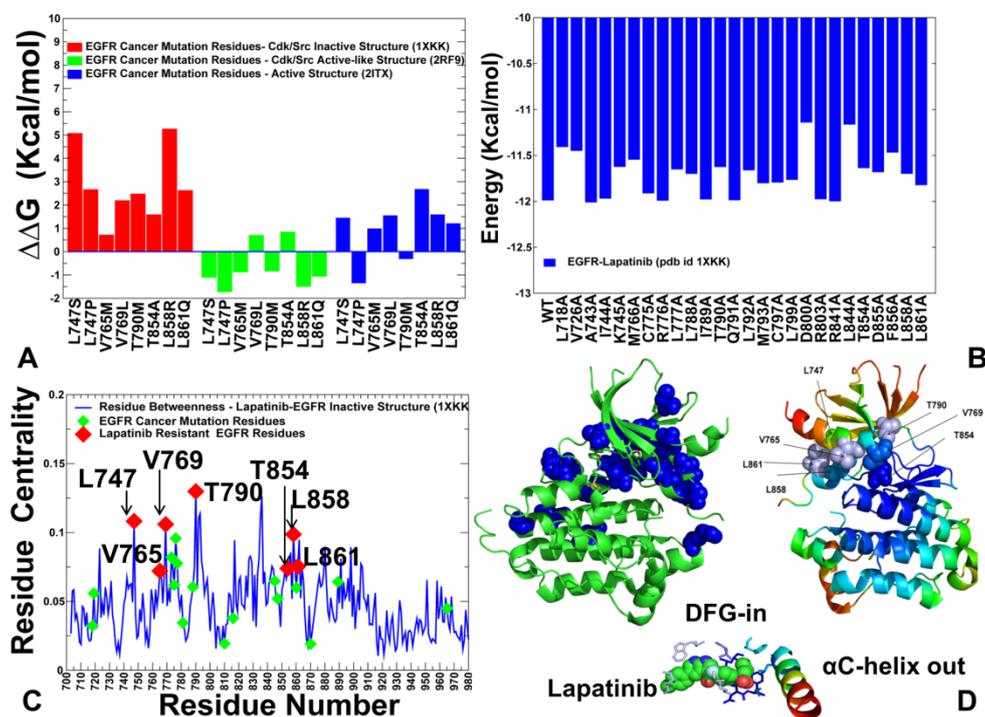


Figure 3. Structure-based network modeling of EGFR cancer mutations and drug binding. (A) Free energy changes caused by oncogenic mutations in different conformational states of EGFR. (B) Computational alanine scanning of binding site residues in the Lapatinib-EGFR complex (pdb id 1XKK). (C) The residue centrality profile of Lapatinib-EGFR complex (in blue). EGFR mutations are shown in green diamonds and Lapatinib-resistant oncogenic mutations are shown in red diamonds. (D) Structural mapping of EGFR cancer mutations (blue spheres) on the crystal structure of Lapatinib-EGFR complex (green ribbons). Mapping of Lapatinib-resistant mutations (indicated by arrows) on the crystal structure of Lapatinib-EGFR complex colored according to structural stability.

Table 1: Structure-based network analysis of the EGFR kinase domain and Lapatinib-EGFR complex. Structural region and network centrality of functional EGFR residues targeted by cancer mutations and drug resistant mutations are reported.

Residue	Residue#	Betweenness	Mutation	Exon	Kinase/segment/spine
Leu	718	0.03230	L718P	Exon 18	$\beta$ 1 strand
Gly	719	0.05586	G719A/C/R/S	Exon 19	Gly-rich P-loop
Leu	747	0.10818	L747S/P	Exon 19	$\beta$ 3- $\alpha$ C loop
Val	765	0.07211	V765M	Exon 20	$\alpha$ C-helix
Val	769	0.10593	V769L	Exon 20	$\alpha$ C-helix
His	773	0.08204	H773L	Exon 20	$\alpha$ C- $\beta$ 4 loop
Cys	775	0.06188	C775F/R/Y	Exon 20	$\alpha$ C- $\beta$ 4 loop
Arg	776	0.09576	R776S/C/H/P/L	Exon 20	$\alpha$ C- $\beta$ 4 loop
Leu	777	0.07761	L777Q/P/M	Exon 20	$\alpha$ C- $\beta$ 4 loop(R-spine)
Cys	781	0.03448	C781F	Exon 20	$\beta$ 4 strand
Leu	788	0.06048	L788V/I/F	Exon 20	$\beta$ 5 strand
Thr	790	0.12979	T790M/A	Exon 20	$\beta$ 5 strand
Gly	810	0.01939	G810S/D/A	Exon 20	$\alpha$ D- $\alpha$ E loop
Asn	816	0.03781	N816K	Exon 20	$\alpha$ E-helix
Val	845	0.06473	V845M/A/L	Exon 21	$\beta$ 7strand (C-spine)
Thr	847	0.05199	T847I/A/K	Exon 21	$\beta$ 7strand
Thr	854	0.07392	T854A/I/A	Exon 21	$\beta$ 7strand
Leu	858	0.10864	L858R/Q/K/V/M	Exon 21	Short helix/A-loop
Lys	860	0.05991	K860T/E/I	Exon21	Short helix/A-loop
Leu	861	0.07540	L861Q/R/E/F/K/P	Exon 21	Short helix/A-loop
His	870	0.01914	H870R/N/Y	Exon 21	A-loop
Arg	889	0.06413	R889S	Exon 22	A-loop
Ile	965	0.04496	I965S/N	Exon 23	$\alpha$ I-helix

Our study suggests that binding of selective and personalized kinase agents can be linked with the robustness of the residue networks in kinase structures. We have found that selective EGFR inhibitors with preferential binding to specific inactive conformations, such as Lapatinib, could be vulnerable to a broad spectrum of resistant mutations pointing to a “dark side” of targeted agents that reflects the inherent conflict between the efficiency and robustness of kinase drugs. The association of network properties with kinase regulation and drug binding suggests that residue interaction networks may be reorganized and specifically tailored through therapeutic agents targeting high centrality residue nodes. Integration of genetic, biochemical and structural data in the unified framework of protein structure networks and systems biology may help to understand and rationally exploit the complex relationships between robustness of targeted genes and binding specificity of personalized drugs.

## References

1. E.D. Pleasance, R.K. Cheetham, P.J. Stephens, D.J. McBride et al., *Nature* **463**, 191 (2010).
2. Z.Kan, B.S. Jaiswal, J. Stinson, V. Janakiraman V et al., *Nature* **466**, 869 (2010).
3. I.B. Weinstein IB. *Science* **297**, 63 (2002).
4. W. Pao, V.A. Miller, K.A. Politi, G.J. Riely et al., *PLoS Med.* **2**, 12 (2005).
5. H. Shigematsu, T. Takahashi, M. Nomura, K. Majmudar et al., *Cancer Res.* **65**, 1642 (2005).
6. C.H. Yun, K.E. Mengwasser, A.V. Toms AV, M.S. Woo, et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2070 (2008).
7. E. Massarelli, F.M. Johnson, H.S. Erickson, Wistuba II, and V. Papadimitrakopoulou, *Lung Cancer* **80**, 235 (2013).
8. M.J. Eck and C.H. Yun, *Biochim. Biophys. Acta* **1804**, 559 (2010).
9. G.E. Konecny, M.D. Pegram, N. Venkatesan N, R. Finn et al., *Cancer Res.* **66**, 1630 (2006).
10. E.R. Wood, A.T. Truesdale, O.B. McDonald, D.Yuan et al., *Cancer Res.* **64**, 6652 (2004).
11. T.M. Gilmer, L. Cable, K. Alligood, D. Rusnak et al., *Cancer Res.* **68**, 571 (2008).
12. E. Avizienyte, R.A. Ward and A.P. Garner, *Biochem. J.* **415**, 197 (2008).
13. M. Taipale, I. Krykbaeva, M. Koeva, C. Kayatekin et al., *Cell* **150**, 987 (2012).
14. T. Shimamura, D. Li, H. Ji, H.J. Haringsma et al., *Cancer Res.* **68**, 5827 (2008).
15. K.S. Gajiwala, J. Feng, R. Ferre, K. Ryan et al., *Structure* **21**, 209 (2013).
16. S. Polier, R.S. Samant, P.A. Clarke, P. Workman et al., *Nat. Chem. Biol.* **9**, 307 (2013).
17. A.F. Neuwald and J.S. Liu, *BMC Bioinformatics* **5**, 157 (2004).
18. A. Torkamani and N.J. Schork *Cancer Res.* **68**, 1675 (2008).
19. J.C. Phillips, R. Braun, W. Wang, J. Gumbart et al., *J. Comput. Chem.* **26**, 1781 (2005).
20. P.A. Kollman, I. Massova, C. Reyes, B. Kuhn et al., *Acc. Chem. Res.* **33**, 889 (2000).
21. K.V. Brinda and S. Vishveshwara, *Biophys. J.* **89**, 4159 (2005).
22. A. Sethi, J. Eargle, A.A. Black and Z. Luthey-Schulten, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6620 (2009).
23. A. Dixit, L. Yi, R. Gowthaman, A. Torkamani et al., *PLoS One* **4**, e7485 (2009).
24. X. Zhang, J. Gureasko, K. Shen, P.A. Cole and J. Kuriyan, *Cell* **125**, 1137 (2006).