Abstract. The development and clinical application of TK domain inhibitors (TKIs) provide important insights into the broader field of cancer-targeted therapies. To discuss the recent advances in the atomic level understanding of EGFR TK domain mutations, we aim at highlighting the current and future importance of these studies on malignancies where the TK domain is improperly activated. The analysis is conducted on published TK domain crystal structures deposited in the Protein Data Bank, or homology structures generated by homology modeling and AutoDock 4.2 software using the program O. Mutations in exon 19 are the most common pathogenic mutations, so the crystal structures with these mutations are analyzed and compared in detail. In addition, we demonstrate how these crystal structures of EGFR conformation with TK domain mutations and those binding with small molecule inhibitors unveil the active or inactive mechanisms. As to the increasing resistance to the TKI, we summarize the progress on overcoming this challenge. Simultaneously, we predict the structure of BIKW-2992 binding to EGFR and compare it with the validated structure of HKI-272. It is hoped that a more accurate resistance mechanism would be found. In brief, we believe that this research will provide insights into EGFR targeted therapies.

1. Introduction

The EGFR gene encompasses 118 kb of sequence on the short arm of human chromosome 7 and contains 28 exons, which encode 1186 amino acids (1). The EGFR protein can be divided into three domains: an extracellular ligand binding domain, a transmembrane domain and an intracellular domain. The intracellular domain contains a juxtamembrane (JM) segment, a kinase domain (TK) and a COOH-terminal tail (C-terminal tail) (2). As a member of tyrosine kinase receptor family, EGFRs form homo- or hetero-dimers with a variety of ligands, leading to phosphorylations of the intracellular tyrosine kinase domain. When activated, the kinases produce a cascade of responses to downstream signaling pathways including Ras/Raf/MAPK and PI3K/Akt/mTOR (3). Therefore, the EGFR plays a critical role in the regulation of cell growth, DNA synthesis and the expression of oncogenes. Therefore, in tumorigenesis of many human cancers individuals with EGFR gene mutations or aberrant relative signal pathway are found (4).

The most important region-TK domain can be divided into an N-terminal ATP-binding lobe (N-lobe) and a C-terminal substrate-binding lobe (C-lobe) (5). Whenever activated or mutated, the TK domain has striking changes in the conformation. The smaller N-lobe, largely constituted by \( \beta \)-sheet and the highly conserved \( \alpha \)-C helix, has a prominent role in conformational changes of the binding pocket. The C-lobe is larger and mostly helical. ATP is proved to be bound the deep cleft between the lobes and located near a highly conserved N-lobe structure (the phosphate binding loop, or P-loop). The kinases exist in two extreme conformational states: on (active) or off (inactive) (6). It seems that the kinases are always in the extended conformation when activated. Activation loop, or A-loop, phosphorylated...
in the active state, provides a platform for the downstream proteins. P-loop contains a conserved glycine-rich/sequence motif (GXGxG) (7). Because the ϕ is usually tyrosine or phenylalanine, the glycine residues allow the P-loop to approach the phosphates of ATP very closely, as well as make the loop very flexible to bind with small molecule inhibitors via changing backbone interactions in the absence of ATP.

In general, the mutations of the EGFR gene focus on the first four exons (18 to 21) of the TK domain gene, which code the N-lobe and part of the C-lobe (8) (Fig. 1). EGFR mutations target the part close to the ATP-binding cleft including α-C helix, the A-loop and the P-loop. The mutations can be categorized as: class (i) in-frame deletions in exon 19, class (ii) single-nucleotide substitutions that cause an amino alteration, and class (iii) in-frame duplications and/or insertions in exon 20. Class (i) mutations usually include amino acid residues Leu746 to Glu750 (L746-G750) located at the N-terminus of the kinase domain C-helix. It accounts for approximately 44% of the active TK domains mutations. In class (ii) mutations the single point mutation in exon 21 that substitutes a leucine with an arginine at codon 858 (L858R) accounts for 41% of EGFR active mutations. Class (iii) mutations of in-frame duplications or insertions of exon 20, are the remaining 5% of active mutations. There are also other active mutations with low frequency in tumors (9) (Fig. 1). Most active mutations are hypothesized to result in similar conformation changes. It is described that these mutations narrow the ATP-binding cleft so as to increase both gene activation and TKI sensitivity.

However, not all tumors accompanying activating mutations are associated with an enhanced response to TKI. Mostly the emergence of drug resistance is conferred by a second point mutation in the TK domain (10). To date, the secondary resistance mutations contain Thr790Met (T790M), Asp761Tyr (D761Y) (11), Ser 768 Ile (S768I) (12), Leu747Ser (L747S) (13) and Thr854Ala (T854A) (14). The T790M mutation is found in approximately 50% of all patients with acquired resistance to TKI (15), while other resistance mutations are relatively rare. Compared to the activating mutations, the resistance mutations are expected to have conformational changes to reduce both gene activation and TKI sensitivity. Preclinical studies are striving at this puzzle. However, the responses of many rare point mutations to TKI remain unknown.

There is a broad consensus on the mechanism of activating mutations that the stabilization of the hydrophobic regulatory spine promotes shift of the kinase towards the constitutively active kinase form, and thus have a dramatic effect on the regulation of the enzyme (16). Because different mutations lead to different sensitivities to TKI, we analyzed and compared all the available structures of EGFR, aiming at demonstrating how structural insights help our understanding of active and resistance mechanisms.

2. The activation mechanisms of TK domain

The activation of TK domain requires two procedures: stimulation of receptor catalytic activity and subsequent creation of docking sites for downstream signaling proteins (17). The homodimer or hetero-dimers of extracellular domain integrated with other ligands cause tyrosine domain autophosphorylation. There are two different types of EGFR targeted: monoclonal antibodies such as cetuximab and panitumumab (18), and tyrosine kinase inhibitors. Here we focus on the latter. We obtained 29 crystal structures for TK domain from the Protein Data Bank (PDB). All the available crystal structures of the TK domain have been published by seven groups in the PDB (19-24). To compare these structures in detail, we confirm the conclusion that the activity of protein kinases is regulated by the conformational state of the catalytic domain. In the active state, the catalytic domain conformation governs the ability to transfer a phosphate from ATP to peptide substrate, and thus controls the downstream signaling pathway (7). There are two essential factors for activation: making the correct amino acid residues oriented and keeping peptide substrate binding site available. In the active state, the A-loop is extended away from the cleft to allow peptide substrate binding, while a catalytic glutamate residue (part of the C-helix) forms an ionic interaction with a lysine residue that coordinates the α and β phosphates of ATP. When inactive, the A-loop often changes conformation dramatically to preclude the binding of peptide substrate, while the C-helix rotates away, pulling with the critical catalytic glutamate residue modified by phosphorylation and dephosphorylation events. It is also proposed that the A-loop folds into a helix similar to the inactive CDK and Src family crystal structures, which prevents C-helix rotation toward the catalytic cleft (18). This can be interpreted that they all belong to the protein kinase family.

The EGFR dimer involves the contacts not only between the two extracellular domains but also between the kinase domains. Recent studies suggest that there is a ligand-independent mechanism for TK domain activation (25). Intracellularly, the N-terminal lobe interacts with the C-terminal lobe forming an asymmetric dimer for activation. Moreover, the C-terminal half of the juxtamembrane segment latches the activated kinase domain to the activator, and the N-terminal half of this segment further potentiates dimerization, likely by forming an antiparallel helical dimer that engages the transmembrane helices of the activated receptor. The formation of the activating juxtamembrane latch is prevented by the C-terminal tail inactive kinase domain dimer. The analysis of EGFR mutations in cancer patients has shown that deletion of the C-terminal tail can drive constitutive EGFR activation (26). In addition, the TK domain can be activated by interacting with small proteins, such as Gab1 (27). This protein-protein interactions also produce resistance to target therapy.

3. The activating mechanism of somatic mutations

As mentioned above, the kinase activity of EGFR is mostly dependent on the conformation of the catalytic domain. When mutated, the shift of the helical axis narrows the ATP-binding cleft and increases both gene activation and TKI sensitivity (4). Thus, the activating mutations located in the catalytic domain cause notable alternations of the structure.

One of the most frequent mutations in exon 19 is deletion that removes residues 746-750 of the expressed protein (28). We obtained the structure of EGFR with ‘E746-A750 mutation in exon 19 by homology modeling based in the structure of the active EGFR (PDB accession code 2gs2) (Fig. 2). Yet, none of those studies would have predicted the structure. We find that the orientation of the N-terminal lobe relative to the C-terminal lobe is not very different in the two models. This is due to a rearrangement of the N-lobe in the inactive
model during the molecular dynamics calculations. The major difference lies in the conformation of the long regulatory loop (residues 709-870) (29), which is highlighted in red. In the inactive state, this loop traverses the ATP binding site, which is located in the hydrophobic pocket including Lys 721, Met 742, Cys751, Leu764 and Met 769 (20). The loop is drawn taut, and sequentially the β-sheet and C-helix have changed, which leads to a more strained system (Fig. 2B). This change probably causes a loss of interactions between secondary structure subunits so as to destabilize the C-helix. Therefore, the deletion mutations in exon 19 increase the activation or sensitivity to various inhibitors. It demonstrates that the activity depends on the stability of the C-helix. Further functional experiments will offer a more complete understanding of this activating mutation.

The A-loop, typically with 20-30 residues in length, is located in exon 21 (30). Leu858 is within the helical turn, and
Figure 3. (A and B) Crystal structures of the active EGFR with L858R mutation (PDB:2itz) and the active EGFR kinase domain with G719S (PDB:2jit) mutation, respectively.

Figure 4. Structures of the T790M mutant in complex with HKI272 (PDB:2gs6) and BIBW2992 (B). With the same aniline structure, BIBW-2992 should bind to the inactive EGFR conformation as well. HKI-272 is surrounded by hydrophobic residues in the expanded pocket, including Met766 in the C-helix, Phe856 and Met790, while BIBW-2992 does not bind to Cys797 and is located in the pocket including Met766 in the C-helix, Phe856, and Met790 (C and D). In the two chemical structures of selected EGFR inhibitors, the compounds are drawn in a consistent orientation and conformation that reflects their approximate binding mode in the EGFR kinase (E).
conducts key hydrophobic interactions with other residues in the N-lobe (Fig. 3). With the highest prevalence of single point activating mutations in the TK domain (31), L858R is located on the protein surface in the active state, while hides in a tightly packed hydrophobic pocket in the inactive (32). So it is expected that the replacement of a small hydrophobic leucine with a large polar argine will not only destabilize the inactive TK domain conformation but also stabilize the active conformation. Gly719 lies in the P-loop (Fig. 3). The conformation of the backbone of Gly719 is unique to glycine and inaccessible to other amino acids. The replacement of glycine with serine leads to this conformation significantly strained and sequentially makes P-loop adopt another conformation (33).

The changes of conformation are various in different mutations. The sensitivities of active mutations to small molecule inhibitors are not equal (34). Biochemical examinations indicated that cells harboring L858R, E746-A750 deletion, and G719S mutations are clearly more sensitive to gefitinib than wild-type cells (35). In clinical research, patients with deletion mutations gain a longer time to progression and median survival time than patients with L858R mutation (P=0.01). Deletion mutations predict better response to gefitinib than the L858R mutation (36). However, the L858R mutant cells are significantly more sensitive to gefitinib than the G719S mutants (32). Therefore, we conclude the deletion mutations in exon 19 have better sensitivity to EGFR-TK target therapy. A possible explanation is that the kinase activity is mostly dependent on the conformation of the catalytic domain. We deduced that the alteration of P-loop emerges as a prominent mechanism to regulate the kinase activity. The structures of EGFR with mutation in exon 19 are the most strained, which leads to the highest activity.

In the crystal structures of gefitinib bound to the L858R mutant, the rotation of gefitinib-alanine ring is by approximately 180 degrees (37). So it is possible that this alternate mode offers a tighter affinity of gefitinib for L858R mutant compared with wild-type EGFR. The structures of the L858R and G719S mutants with gefitinib also suggest that these mutations have improved $K_{d}/K_{\text{on}}$ (34). Whether the same results can be found in exon 19 requires further study. Only by revealing the activating mechanism of the most common mutations, can we attain better therapy efficacy in the clinic.

4. Primary resistance to EGFR TKI

There are more and more mutations leading to increased resistance. Two strategies for preventing or overcoming this resistance have been found: to identify novel agents binding to and inhibiting TK domain by a distinct, non-ATP competitive mechanism, or to inhibit ATP binding to the TK domain irreversibly (38). The early studies expected that the irreversible binding compounds might show a longer duration of activation (37). A further motivation for designing these irreversible inhibitors is thought to realize the first path in the way of being noncompetitive with ATP. The resistance mechanism of T790M mainly is debated.

As the most common secondary mutation, T790M mutation seems to occur in approximately half of all patients with TKI resistance. Since Thr790 is involved in the binding site of TKI, the substitution of threonine to methionine causes steric hindrance of TKI binding (36), but T790M mutation increases the affinity to ATP as well (39), which leads to reduced potency of any ATP-competitive agent. The irreversible EGFR inhibitors, including BIBW-2992 (Afatinib) and HKI-272 (Neratinib), can inhibit growth in gefitinib-resistant NSCLC or Ba/F3 cell lines that contain the T790M mutation (40).

HKI-272 is a 4-(arylamino) quinoline-3-carbonitrile compound and a potent inhibitor of both EGFR and Her2 kinases (41). In complex with HKI-272, the EGFR kinase adopts an inactive conformation. In the inactive state, the regulatory C-helix is displaced from its active position, which creates enlarged hydrophobic pocket to accommodate the bulky aniline substituent found in HKI-272. Like aniline quinazoline compounds, the quinoline core of HKI-272 forms a single hydrogen bond with the hinge region of the kinase. The 2-pyridyl group of HKI-272 is surrounded by hydrophobic residues in the expanded pocket, including Met676 in the C-helix, Phe856 and Met790. The nitrile substituent of HKI-272 also approaches the above gatekeeper residue.

BIBW-2992 is also a dual inhibitor of both EGFR and ErbB2 kinases (41). It shows a higher affinity for binding to EGFR in the T790M resistant mutant (42). In a phase II clinical trial, BIBW-2992 attained responses in patients with deletion mutations in exon 19, L858R, L858R/T790M, L861Q, T854A (14) and G719S/S768I mutations.

It is reported that BIBW-2992 has increased efficacy over erlotinib, gefitinib and lapatinib. Using AutoDock 4.2, we obtained the structure of T790M mutation in complex with BIBW-2992 (Fig. 4). BIBW-2992 does not bind to Cys797, but is located in the pocket including Met766 in the C-helix, Phe856, and Met790. The binding energy is $-4.36 \text{ kcal/mol}$. HKI-272 is thought to bind to the inactive conformation of EGFR because of the additional aniline substitutions (13). With the same aniline structure, BIBW-2992 should bind to EGFR in the inactive conformation. Results from many studies have demonstrated that BIBW-2992 has a prolonged duration of activation in accordance with its mode of binding.

5. Other resistance mechanisms to TKI

In addition to the secondary mutations of EGFR, other resistance mechanisms such as KRAS mutations, C-met proto-oncogene (c-Met) amplification and mutations of downstream signal pathway molecules have been found.

KRAS mutations have been demonstrated to be associated with primary resistance to TKI in many tumors (38). KRAS mutations produce impaired GTPase activity and continual activation of Ras signaling, leading to the activation of proliferative and anti-apoptotic pathways (43). The most common mutations are located in exon 12 and 13. Activation of KRAS triggers both MAPK and MAPK-independent signal transduction (44). Oncogenic KRAS can promote tumor cells proliferation through Ras/Raf/MAPK pathway. Yet activating mutations of KRAS induce the constitutive MAPK-independent signal pathway and cause reduced sensitivity to EGFR target therapy.

c-Met amplification is detected in 20% of NSCLC patients with TKI resistance (half with T790M) (45). c-Met interacted with different signal modifiers, including scaffolding adaptors, cytoskeleton connectors and structurally homologous co-receptors leading to efficient activation of downstream signal...
transduction pathways that include MAPK, Jun amino-terminal kinases (JNK) and PI3K/Akt, signal transducer and activator of transcription proteins (STAT) (46). Thus, c-Met plays a critical role in cell proliferation, survival and migration. Because c-Met amplification is identified in a proportion of patients with acquired resistance to gefitinib/erlotinib, novel c-Met/TK inhibitors could be a treatment strategy for a small fraction of patients with c-Met amplification as the main mechanism of resistance (47). Trials with c-Met inhibitors (such as XL184, ARQ 197, PF-2341066, SGX523) will be discussed later (48).

Research over the past 40 years has uncovered crucial players in the EGFR signal transduction pathway, which can be roughly divided into two categories: the pro-survival arm of the pathway comprising the PI3K/Akt/mTOR cascade, and the proliferative arm consisting of the Ras/Raf/MAPK pathway (49). The importance of these signal pathways is well known. Deregulation of the PI3K/Akt pathway is associated with resistance to TKI in SCCHN (32). This offers another therapy target to overcome the resistance.

As a whole, it is a complex apparatus including KRAS mutations, c-Met amplification and the abnormal downstream signal pathways that produce other resistance to TKI. To take combination inhibitors to this complex network may overcome various resistances.

6. Conclusion

As a member of tyrosine kinase family, EGFR plays a critical role in the regulation of cell growth, DNA synthesis and the expression of oncopgenes. We have found various abnormalities in EGFR during many pathological developments, particularly mutations of EGFR. Therefore, EGFRs have become a hot therapeutic target, which contains monoclonal antibodies and tyrosine kinase inhibitors. The emergence of the drug resistance impels further studies of genuine mechanism to this treatment. Research on the conformation of this therapeutic target will be of importance. By summarizing achievements of many scholars, we try to elaborate this puzzle.

The dominated activating mechanism is that alternations of structure of TK domain offer activity to EGFR. N-lobe, C-lobe, P-loop and A-loop, which include β-sheet and α-C helix, have a prominent role in conformational changes of the binding pocket. Normally, the catalytic domain conformation governs the ability to transfer a phosphate from ATP to peptide substrate, and thus controls the downstream signal pathway. The ATP-binding catalytic cleft lies between the N-terminal (N-lobe) and C-terminal (C-lobe) lobes. Key structures that are required for function and activation surrounding the cleft, include the α-C helix, the A-loop and the P-loop. A-loop provides a platform for the downstream proteins, and P-loop approaches the phosphates of ATP. The first four exons (18 to 21) of the TK domain gene code the N-lobe and part of the C-lobe. The mutations of the four exons consequently govern this activation. In the clinic, the deletion mutations in exon 19 and the single point mutation in exon 21 are the most common in the pathological changes. On the base of a validated crystal structure, scholars have put forward that the replacement of a small hydrophobic leucine with a large polar argine will not only destabilize the inactive TK domain conformation but stabilize the active conformation in the L858R mutation, which proves the changes of the conformation in this therapeutic target is important (6). There are no crystal structures of deletion mutations in exon 19. In order to understand this most important mutation, we predict the structure of -L746-G750 with homology modeling. The modeling shows significant changes of conformation. Thus, we infer that structural changes of EGFR TK domain are the leading role of this therapeutic target.

Much progress has been made in understanding the mechanism of TKI resistance, such as secondary mutations, gene amplification and mutations of the EGFR downstream signal pathway. We mainly debate irreversible EGFR inhibitors. We predicted the structure of T790M mutation in complex with BIBW-2992. It is expected that the structure would be beneficial to overcome the resistance.

The crystal structures have already led to much clearer understanding of EGFR activating and resistance mechanisms. Such a crystallographic study research should prove the development of more effective treatment strategies through the availability of better-targeted drugs or rational combination strategies with a higher response, a better therapeutic window and a long-time disease control. Further structural and functional studies are required to gain a more complete understanding of the EGFR TK domain. The studies how TK domain recognizes the ATP and different molecules at the molecular level are required to design inhibitors for therapeutic use.

Acknowledgments

This project was supported by grants from National Natural Science Foundation of China (81141094, 30870962, 31000323), Natural Science Foundation of Jiangsu Province (BK2011656), specialized Research Fund for Doctoral Program of Higher Education of China (20100091120023, 20100091120026) and Fundamental Research Funds for the Central Universities (1095020823).

References


