

Article

Uncovering the Molecular Basis for the Better Gefitinib Sensitivity of EGFR with Complex Mutations over Single Rare Mutation: Insights from Molecular Simulations

Miaomiao Li ¹ , Mengrong Li ¹, Yanjie Xie ¹  and Jingjing Guo ^{1,2,*} 

¹ College of Life Sciences, Nanjing Agricultural University, Nanjing 210095, China; 2021216040@stu.njau.edu.cn (M.L.); 2019216037@njau.edu.cn (M.L.); yjxie@njau.edu.cn (Y.X.)

² Engineering Research Centre of Applied Technology on Machine Translation and Artificial Intelligence, Faculty of Applied Science, Macao Polytechnic University, Macao 999078, China

* Correspondence: guojj@njau.edu.cn

Abstract: Epidermal growth factor receptor (EGFR) is an intensively focused target for anti-tumor compounds used in non-small cell lung cancer (NSCLC) therapy. Compared to the classical activating mutations, there are still many uncommon EGFR mutations associated with poorer responses to EGFR inhibitors. A detailed understanding of the molecular basis for multiple EGFR mutants exhibiting diverse responses to inhibitors is of critical importance for related drug development. Herein, we explored the molecular determinants contributing to the distinct responses of EGFR with a single rare mutation (G719S) or combined mutations (G719S/L858R and G719S/L861Q) to Gefitinib (IRE). Our results indicated that interactions, formed within the tetrad of residues S768 (in the α C-helix), D770 (in the α C- β 4 loop), Y827 (in the α E-helix), and R831 (in the catalytic loop), play an important role in the stability of α C-helix and the maintenance of K745–E762 salt bridge in the absence of IRE, which are weakened in the EGFR^{G719S} system and enhanced in the EGFR^{G719S/L858R} system upon IRE binding. Besides, the introduced hydrogen bonds by the co-occurring mutation partner also contribute to the stability of α C-helix. The work done for inhibitor dissociation suggests that IRE exhibits a stronger binding affinity to EGFR^{G719S/L858R} mutant. Together, these findings provide a deeper understanding of minor mutations, which is essential for drug development targeting EGFR with less common mutations.

Keywords: EGFR rare mutants; EGFR complex mutations; molecular dynamics simulation; steered molecular dynamics simulation; Gefitinib (IRE)



Citation: Li, M.; Li, M.; Xie, Y.; Guo, J. Uncovering the Molecular Basis for the Better Gefitinib Sensitivity of EGFR with Complex Mutations over Single Rare Mutation: Insights from Molecular Simulations. *Molecules* **2022**, *27*, 3844. <https://doi.org/10.3390/molecules27123844>

Academic Editor: Jan Brezovsky

Received: 23 May 2022

Accepted: 11 June 2022

Published: 15 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Many human cancers' occurrence is associated with the abnormal activation of tyrosine kinase [1,2]. Abundant genomic alterations in epidermal growth factor receptor (EGFR) have been revealed, including point mutations, deletions, and insertions within exon 18–21 of the EGFR gene [3–5]. The mutational constitutive activation of the kinase domain of EGFR is linked to non-small cell lung cancer (NSCLC), [6–8] accounting for 80% of lung cancer [9]. About 85% of observed EGFR mutations are common mutations occurring frequently in NSCLC patients, while the remaining 15% cover various rare mutations [3,10]. To treat lung cancer, characterized as the leading cause of morbidity and mortality, some targeted drugs targeting EGFR have been developed and approved, such as Gefitinib (IRE), Erlotinib, and Afatinib [11–14]. Gefitinib has a novel antitumor effect on NSCLC and has been documented to be useful in patients previously treated with chemotherapy regimens [15], or in selected groups of patients as first-line therapy, with effects of tumor regression and symptom amelioration [16].

Among rare EGFR mutations, G719X (including G719S, G719A, G719C, and G719D) substitutions in the phosphate-binding “P-loop” within the N-lobe, is one of the most

commonly observed mutations, and represent roughly 1.5–3% of all EGFR mutations in NSCLC patients [10]. G719X mutations could occur as an independent EGFR one or together with other additional point mutations (rare or classical) as compound EGFR mutations [17]. Complex mutations have been reported to account for about 5–15% of EGFR mutations. [18–21] Effective response to IRE is associated with EGFR mutations, as reported in various studies. Chiu et al. [22] assessed the effectiveness of IRE in patients with rare G719X mutation, common L858R mutation, and rare–rare EGFR mutation combinations. The clinical study focused on uncommon EGFR mutations revealed that although patients with a single G719X mutation were sensitive to IRE treatment, they were markedly less sensitive compared to those with the classical L858R mutation. Furthermore, the results of rare and rare combined mutations are heterogeneous as patients with G719X + L861Q exhibit more impressive outcomes than those with G719X + S768I. Hence, it is likely that the sensitivity of NSCLC patients to EGFR inhibitors is influenced by the specific co-occurring mutation partner within rare and rare combinations [22]. Beau-Faller et al. and Cheng et al. also uncovered a better response to IRE for EGFR with complex mutations than a single rare mutation [17,23]. EGFR featured with complex mutations exhibits a stronger sensitivity to IRE, especially under a rare mutation combined with a classical mutation excluding the T790M mutation [23,24].

NSCLC-diagnosed patients exhibit diverse outcomes upon IRE treatment based on clinical study, yet molecular determinants behind the different performances are not fully understood. Many studies have focused on common mutations such as L858R and T790M, [25,26] but less is known for identifying rare mutations at the molecular level. In view of the high incidence of lung cancer, there are a lot of NSCLC-diagnosed patients found to harbor rare EGFR mutations [27]. Therefore, a detailed understanding of minor mutations is favorable to assess the current treatment options and develop new effective therapeutic strategies [28].

For this purpose, we carried out a systematic computational study to provide insights into different performances of IRE on multiple EGFR mutants. In the present work, conventional molecular dynamic (CMD) simulations and steered molecular dynamic (SMD) simulations have been carried out to explore the characteristics of EGFR mutants. A total of six systems were considered, including APO and IRE-bound EGFR mutants: rare (G719S), rare + rare (G719S/L861Q), and rare + classical (G719S/L858R). Our analyses showed that in the absence of IRE, cancer mutations favor the active α C-in conformation, and the stability of α C-helix is mediated by hydrogen bond interactions formed within a tetrad of residues S768 (in the α C-helix), D770 (in the α C- β 4 loop), Y827 (in the α E-helix), and R831 (in the catalytic loop). Upon IRE binding, the favorable interactions were disrupted in the EGFR^{G719S} system, whereas enhanced in the EGFR^{G719S/L858R} system. Besides, the free energy profile suggests that IRE has a stronger binding affinity to EGFR^{G719S/L858R} mutant. These results provide us with serviceable insights on the better IRE sensitivity of the EGFR compound mutant over the single rare one, which is useful for developing effective therapeutic strategies.

2. Results and Discussion

2.1. EGFR Minor Mutations Favor the Active α C-in Conformation

EGFR is a transmembrane receptor tyrosine kinase, of which the kinase domain comprises two lobes (Figure 1A), namely the N-terminal and the C-terminal lobes. The catalytic cleft formed between the two lobes consists of two distinct binding pockets. The one in the front is the ATP-binding site, while the back one is a non-ATP-contact region. The general kinase ATP-binding pocket could be divided into four regions: the adenine region, the ribose region, the phosphate region, and the hydrophobic pocket II. In the active conformation, the non-ATP-contact region is primarily the hydrophobic pocket I, which also is identified as the affinity and selectivity pocket, and hence is important for inhibitor binding [29]. Apart from the orthosteric site occupied by the majority of inhibitor IRE, the benzene ring part is extended to the hydrophobic pocket I (Figure 1B). Moreover, for

the active EGFR conformation, the α C-helix is positioned in an “in” conformation and the K745–E762 salt bridge is present.

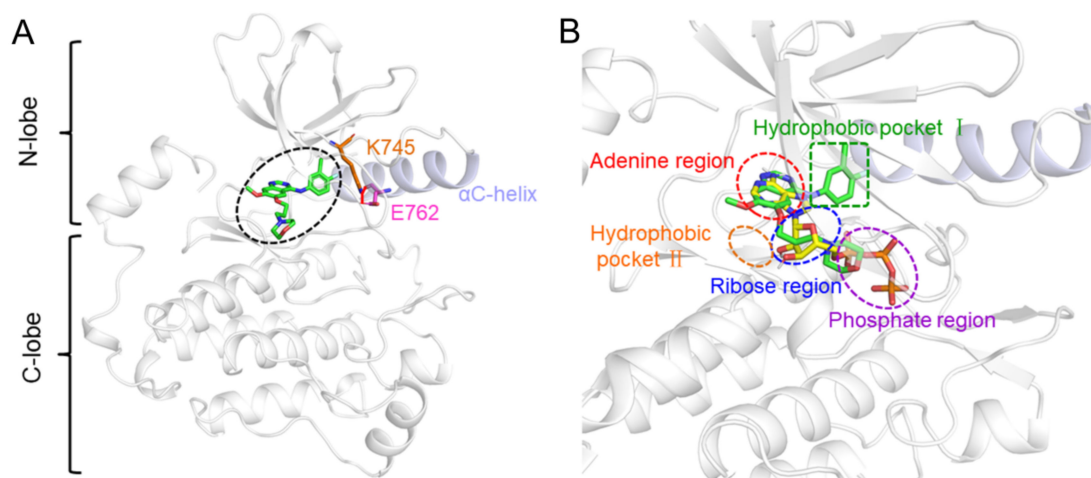


Figure 1. The structure of EGFR is complex with IRE in the absence (A) and presence (B) of ATP. The α C-helix is marked in light blue. The residues E762 and K745 forming a salt bridge are shown as orange and magenta sticks, respectively. The small molecules IRE and ATP are shown as green and yellow sticks, respectively. The IRE binding site is highlighted by a black circle in (A), and the five regions consisting of the catalytic cleft are highlighted by circles with different colors in (B).

The root mean square deviation (RMSD) of the backbone atoms is generally considered to be an indicator of stability. Firstly, the RMSDs of the whole EGFR kinase domain in distinct forms were calculated according to their initial positions to describe the conformational stability of protein. As presented in Figure S1, the RMSDs of most simulations are stable and smaller than 3 Å on the whole, indicating the high stability of EGFR systems during the course of simulations; therefore, the studied systems performed three parallel runs of 800 nanoseconds is feasible [30]. The equilibrium periods of the trajectories were determined to be between 600 and 800 ns in unbound and IRE-bound EGFRs (Figure S1); therefore, the next analyses were based on the equilibrium portion of the trajectories. Then, to intuitively compare the stability of functional region α C-helix, we also performed a distribution calculation based on the RMSD values of α C-helix backbone atoms in the absence of IRE. It has been well established that the active α C-in conformation is characterized by the placement of the α C-helix to maintain the catalytically important salt bridge formed between K745 and E762 (referred to hereafter as KE salt bridge) [31]. Therefore, the distance between the charged groups of K745 and E762 was also calculated.

From the RMSD distribution figure, it is found that the α C-helix exhibits lower flexibility with the average RMSD values below 2 Å among the three APO systems (Figure 2A). Additionally, from the distance calculation, it is obtained that the average K745–E762 distances are 3.05 ± 0.83 Å for the G719S-APO system, 3.35 ± 1.12 Å for the G719S/L858R-APO one, and 2.79 ± 0.27 Å for the G719S/L861Q-APO one, respectively. The results indicate that compared with the single rare mutation, rare–rare combined mutations result in the K745–E762 distance value shifting to a relatively smaller area, while rare-classical combinations induce a broader distribution area of the K745–E762 distance values (Figure 2B–D). On the whole, such changes observed in the mutant systems indicate that EGFR mutants tend to sample the active α C-in conformation with the presence of the KE salt bridge.

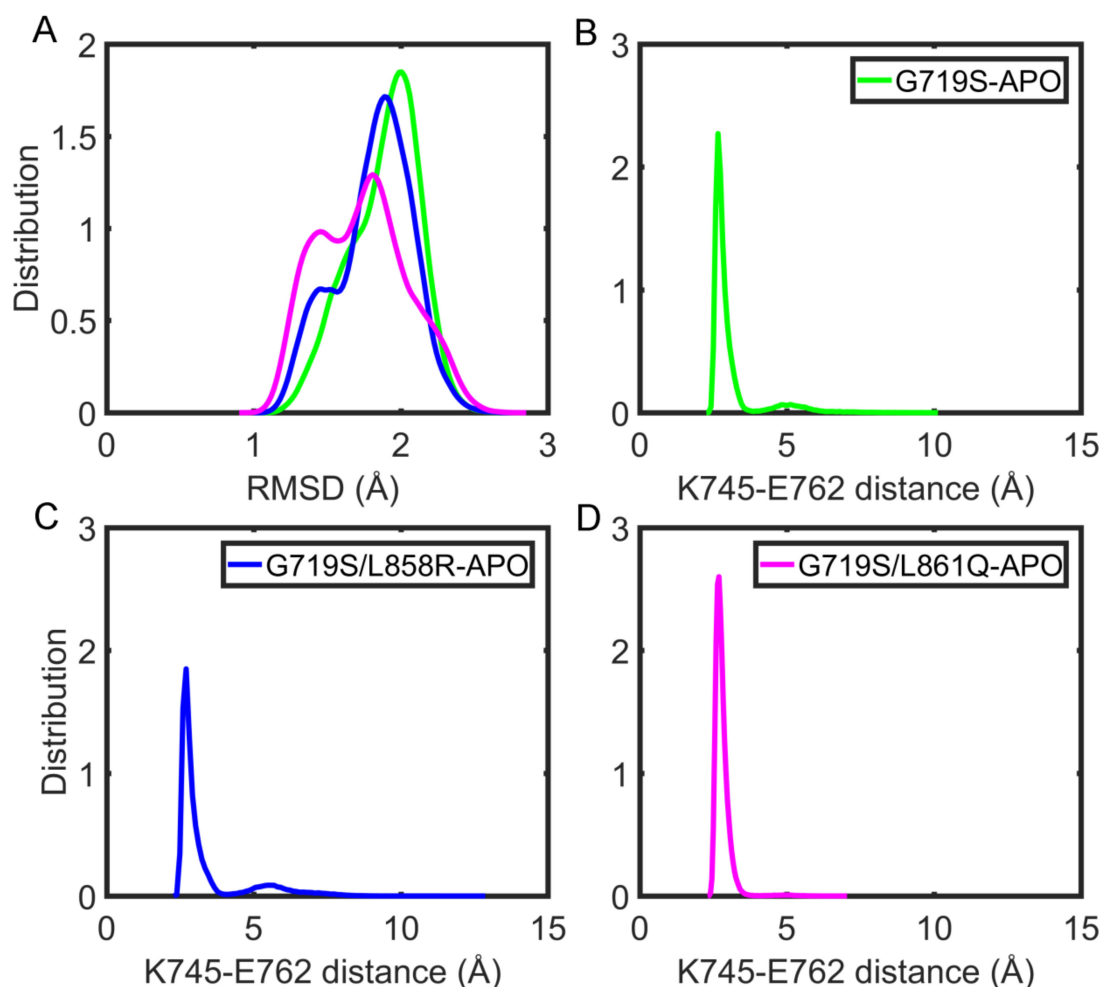


Figure 2. The stability of the α C-helix and KE salt bridge during the simulations for APO systems. (A) The RMSD values of the α C-helix backbone atoms in the last 200 ns MD trajectories using the initial structure as the reference. (B–D) The probability distribution profile of the distance between atoms K745/NZ and E762/(OE1, OE2) during the last 200 ns is based on all three replicates.

2.2. Identification of Favorable Interactions Contributing to the Stability of the α C-Helix

In order to investigate the dynamic behaviors of EGFR mutants, the root-mean-square fluctuation (RMSF) of $C\alpha$ atoms of protein residues and the cross-correlations of $C\alpha$ atomic displacements were calculated based on the last 200-ns MD trajectories considering all three replicates together. The important regions of EGFR were highlighted in different colors shown in (Figure S2) to visualize the spatial position of the regions mentioned next on the protein. As RMSF results indicated (Figure S3), the flexibility of the P-loop was reduced upon complex formation. This structure acts as a flexible clamp that covers and anchors the non-transferable phosphates of ATP for the EGFR kinase domain [32,33]. When the activate site was occupied by the inhibitor IRE, the reduced flexibility of the P-loop may contribute to its stable binding. It was also noted that the A-loop exhibited the highest flexibility in the G719S-IRE system compared with the other two IRE-bound EGFRs with complex mutations, which might be associated with different IRE sensitivity of EGFR in distinct forms.

Through constructing a covariance matrix, the correlations of each pair of $C\alpha$ atoms are intuitively clear. As presented in the covariance matrix maps (Figure 3), an apparent correlated motion is observed in the region spanning from K754 to L862, including the α C-helix, the α C- β 4 loop, the hinge region, the α E-helix, the catalytic loop, and the N-terminal of the A-loop regions. These regions demonstrated by a black square box exhibit are positively correlated. Among these regions, the α C- β 4 loop, the α E-helix, the catalytic

loop, and the A-loop (N-terminal) regions are adjacent either to the α C-helix or to the hinge region which regulates the relative motions between the N lobe and the C lobe of kinase domain as well as the conformation of the binding site (Figure S2). The tetrad of residues S768 (in the α C-helix), D770 (in the α C- β 4 loop), Y827 (in the α E-helix), and R831 (in the catalytic loop) make hydrogen bond interactions which, as a consequence, contribute to maintaining the α C-in conformation (Figure S4A).

In addition, hydrogen bonds, formed between the α C-helix and the β 4- β 5 loop or the A-loop, are also in favor of the stability of α C-in conformation as well as the KE salt bridge. The residue N756 in the α C-helix forms hydrogen bonds with S784, T785, and V786 on the β 4- β 5 loop (Figure S4B), and the residue E758 also located in the α C-helix interacts with R858, K860, and K875 on the A-loop (Figure S4B–D). Overall, the α C-in conformation was maintained during the simulations as a result of these formed favorable interactions.

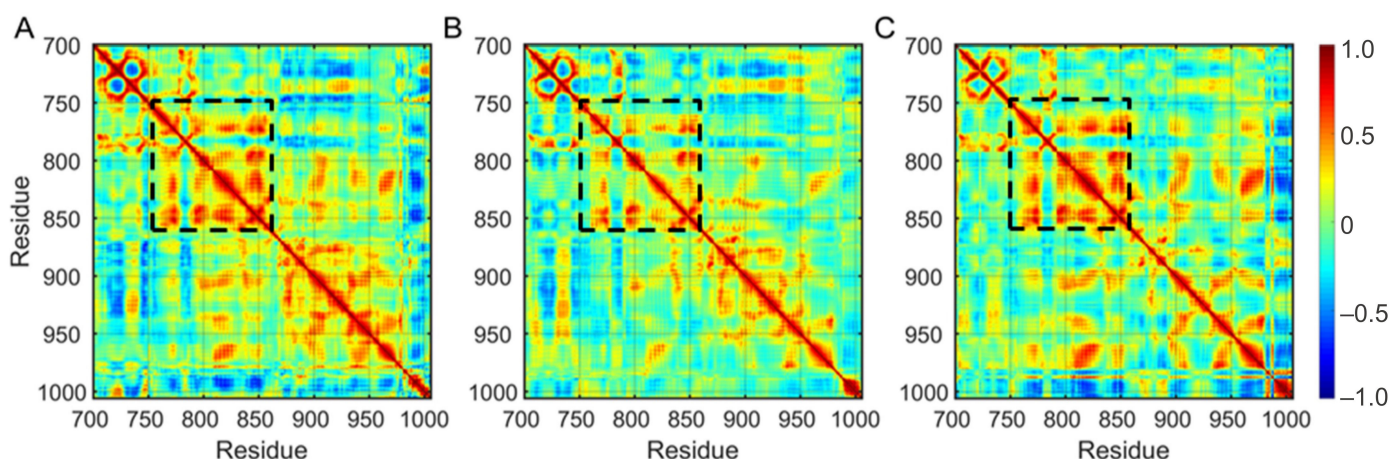


Figure 3. The covariance matrix maps of EGFR in the APO forms: (A) G719S-APO, (B) G719S/L858R-APO, and (C) G719S/L861Q-APO. The positively correlated regions existing in all APO systems are demonstrated by a black square box.

2.3. IRE Binding Alters the Conformational Dynamics of EGFR Kinase Domain

To characterize the conformational changes of the EGFR kinase domain upon IRE binding, principal component analysis was carried out on the last 200 ns MD trajectories considering three replicates together. The results are shown in the free energy surface using the first two principal components (named PC1 and PC2) (Figure 4). It is noted that for the EGFR^{G719S} mutant in APO form, the conformational space is wider along the PC2 direction, which is considered to be more flexible without the occupation of an inhibitor at the binding pocket (Figure 4A). Moreover, the G719S/L861Q-APO system is more stable demonstrated by the least conformational space (Figure 4C). All EGFR mutants in IRE-bound form discriminate from the APO one by PC1 (Figure 4A–C). Besides, the discrimination between the APO form and IRE-bound form is also according to PC2 for EGFR^{G719S} and EGFR^{G719S/L858R} mutants.

The flexibility of the EGFR kinase domain is reflected in a variation of the inter-lobe rearrangement. PC1 mainly captures the twisting motion of the N-lobe (Supplementary Movie S1), and PC2 illustrates an open-closed movement of the N-lobe as well as the rearrangement of the A-loop (Supplementary Movie S2), both resulting in the conformational change of binding site. Along the PC1 direction, the N-lobe as a whole presents an outward twisting motion in the IRE-bound form compared with in the APO one; thereby, as a consequence, the α C-helix has an outward displacement in the complex systems to enlarge the non-ATP contact region to fit the benzene ring part of IRE. Along the PC2 direction, the N-lobe exhibits a downward movement in the IRE-bound form compared with in the APO one, which makes the binding pocket more stable in the presence of IRE. Generally speaking, the inhibitor binding translates into protein conformational changes, which mainly affect the functionally important regions of the target protein. The binding pocket

undergoes dynamical and conformational changes to better accommodate the inhibitor in the complex systems.

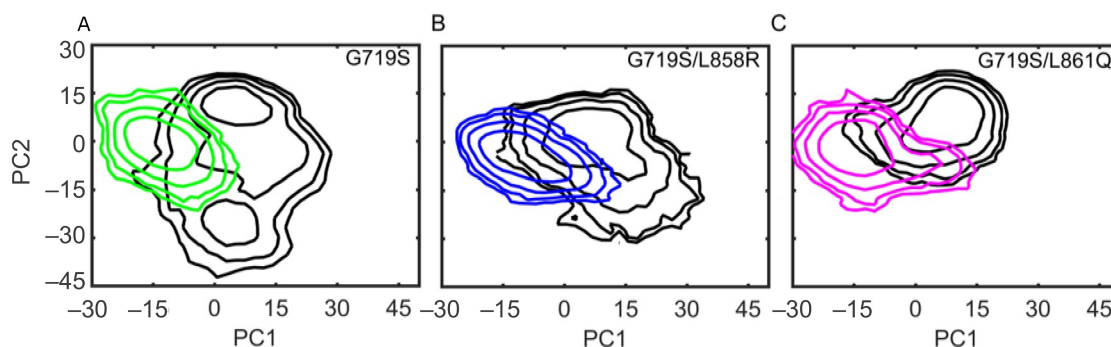


Figure 4. Conformational dynamics of EGFR in the APO and IRE-bound forms. The first two principal components (PCs) are shown in free energy surfaces. All APO forms are colored in black (A–C). Complex systems of G719S-IRE (A), G719S/L858R-IRE (B), and G719S/L861Q-IRE (C) are colored in green, blue, and magenta, respectively.

2.4. IRE Binding Destabilizes the α C-Helix and KE Salt Bridge in the G719S System

The α C-helix serves as an important functional region of EGFR, of which RMSD is also calculated to character its conformational stability in the IRE-bound form. As shown in Figure 5, the α C-helix is relatively more flexible in the IRE-bound G719S system with the largest RMSD value of 2.02 ± 0.27 Å (Figure 5A). For the G719S/L858R-IRE system, the binding of IRE reduces the flexibility of the α C-helix, demonstrated by the average RMSD value of 1.77 ± 0.22 Å in the IRE-bound form and 1.82 ± 0.26 Å in the APO one (Figure 5B). For the G719S/L861Q-IRE system, IRE binding also contributes to less flexibility of the α C-helix with a smaller RMSD value distribution area compared with that in the APO form (Figure 5C).

The KE salt bridge serves as a proxy for characterizing the α C-in conformation, of which stability is associated with that of the α C-helix, thus the distance calculation is performed for the two charged residues to reflect the effect of IRE binding. From Figure 5D, it is found that in the IRE-bound G719S system, the KE salt bridge is disrupted, as demonstrated by the larger distance distribution with an average of 9.35 ± 4.84 Å. For the other two EGFR mutants in the IRE-bound form, the KE salt bridge is well maintained, with an average distance of 2.81 ± 0.26 Å in the EGFR^{G719S/L858R} system and 2.84 ± 0.32 Å in the EGFR^{G719S/L861Q} one (Figure 5E,F). Overall, the G719S/L858R combined mutations enhanced the sensitivity for EGFR toward IRE based on the smallest RMSD value of α C-helix and KE distance value.

Shan et al. suggested the presence of a third conformation of EGFR kinase characterized by a broken KE salt bridge and partially unfolded α C-helix, which was an α C-out conformation that differed from the Src/CDK2-like (α C-out) inactive conformation [25]. It is worth mentioning that IRE is an ATP-competitive and reversible inhibitor, which binding to EGFR featured with an active α C-in conformation. For the G719S-IRE system, the K745–E762 distance change indicates that the α C-in conformation is less sampled during the simulation, and the α C-helix conformation might sample the above-mentioned third conformation, which might suggest less sensitivity of EGFR^{G719S} mutant to inhibitor IRE.

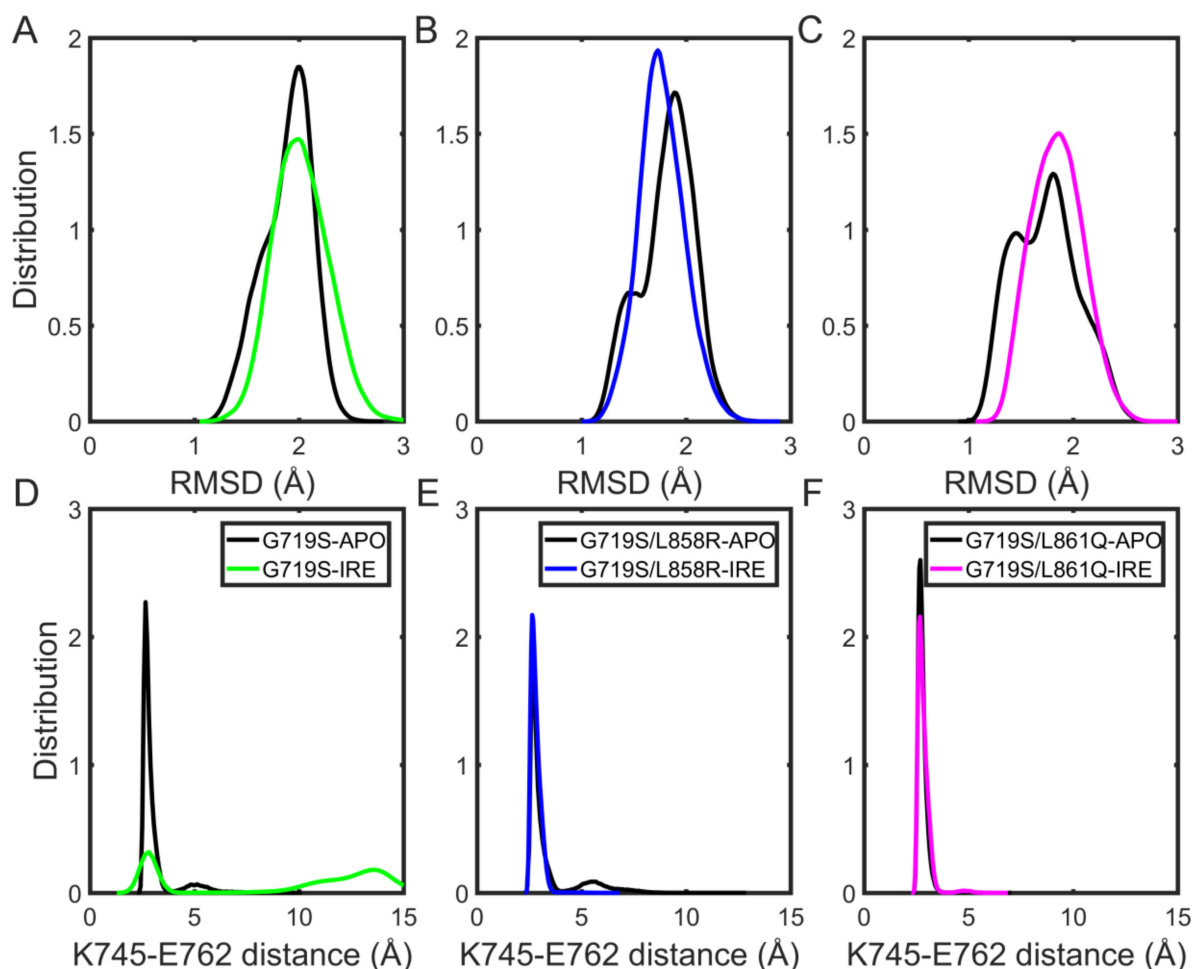


Figure 5. The stability of the α C-helix and KE salt bridge upon IRE binding during the simulations. The RMSD of the α C-helix backbone atoms in the absence and presence of IRE: (A) G719S mutants; (B) G719S/L858R mutants; (C) G719S/L861Q mutants. The probability distribution profiles of the distance between atoms K745/NZ and E762/(OE1, OE2) in the absence and presence of IRE: (D) G719S mutants with the APO form colored in black and with the complex form colored in green; (E) G719S/L858R mutants with the APO form colored in black and with the complex form colored in blue; (F) G719S/L861Q mutants with the APO form colored in black and with the complex form colored in magenta.

By analyzing the hydrogen bond interactions mentioned above, it might provide a plausible explanation for the different observations for the α C-helix and KE salt bridge among the complex systems. Comparing the hydrogen bond occupancy between the APO systems and the complex ones (Figure 6), it is found that for the IRE-bound G719S system, the hydrogen bond formed between R831 and D770 disappears, and the occupancy of the Y827–D770 hydrogen bond is significantly reduced. For the G719S/L858R-IRE system, it is observed the formation of the R831–D770 hydrogen bond, and the prolonged existence of the Y827–D770 hydrogen-bond interaction. There is less change of hydrogen bond occupancy for the EGFR^{G719S/L861Q} mutant upon IRE binding. These hydrogen bond interactions could exert a favorable effect on the α C-in positioning. The broken and reduced interactions in the G719S-IRE system might give rise to less stability of the α C-helix.

Furthermore, the hydrogen bonds introduced by the co-occurring mutated residue also make a positive knock-on effect on the stability of α C-helix. For the G719S/L858R system, the hydrophobic leucine residue mutates to the charged arginine residue, and the latter makes interactions with other charged residues (R836 and D837) located at the catalytic loop (Figure S5A). For the G719S/L861Q system, the hydrophobic leucine residue

mutates to the charged residue glutamine, similarly, the latter forms a hydrogen bond with R831 located at the catalytic loop and Y764 positioned at the α C-helix (Figure S5B,C). All these interactions are conducive to stabilizing the α C-helix indirectly. In words, it is reasonable to correlate the hydrogen bond interactions with the stability of α C-helix and, as observed in the G719S-IRE system, less stability of the α C-helix and broken KE salt bridge.

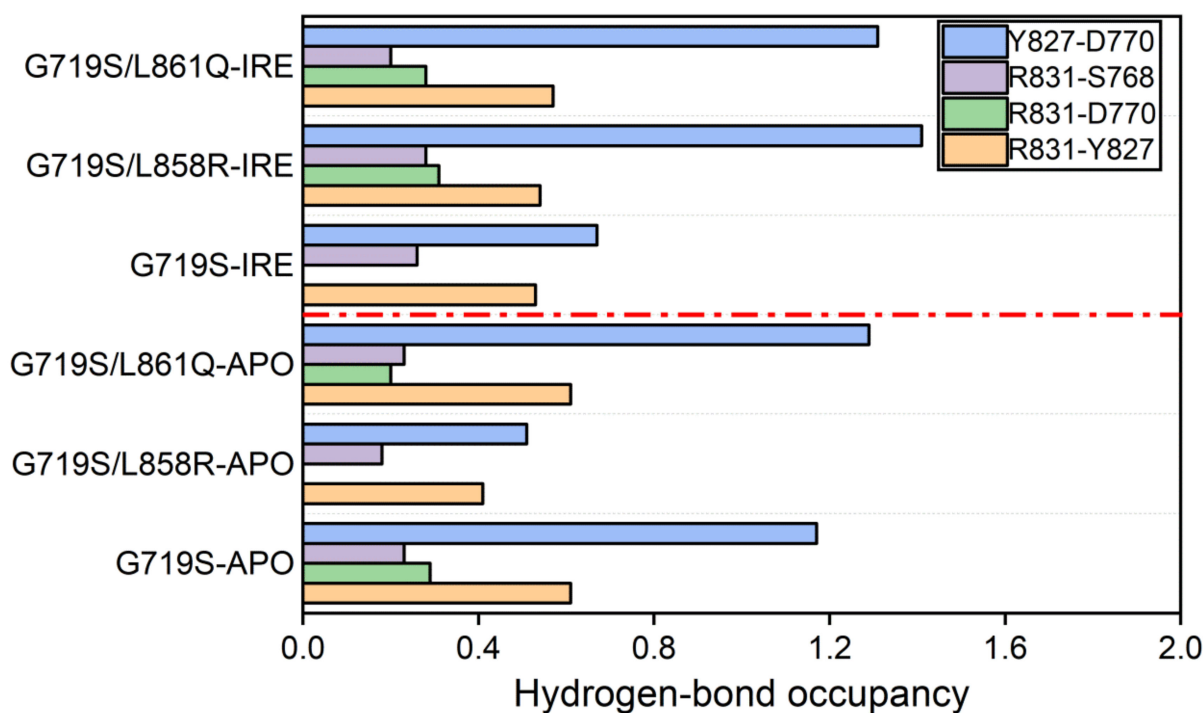


Figure 6. Hydrogen bond occupancy differs between the APO systems and the complex systems. Some residues have a hydrogen bond donor or acceptor of more than one, thus the hydrogen bond occupancy is above 1. A low panel of the red dotted line indicates the occupancy of hydrogen bonds (Y827-D770, R831-S768, R831-D770, and R831-Y827) in the APO systems, and the top panel indicates the occupancy of these hydrogen bonds in the complex systems.

2.5. Effects of IRE Binding on the Binding Pocket

To probe how IRE binding affects the binding pocket, the radius of gyration (R_g) was calculated for residues consisting binding pocket during the simulations. R_g depends on the relationship between the mass of certain atoms and the center of gravity of the molecule, which could be used to characterize the compactness of the binding pocket here. As presented in Figure 7A–C, though the average R_g value is increased, it shows a smaller distribution area in the complex systems compared with the APO systems. It is suggested that the binding pocket is more dynamic over the course of the simulation in absence of IRE, and a relatively stable pocket is induced by IRE binding though it stretches out in the presence of IRE.

We then estimated the contact information between IRE and EGFR. As presented in Figure S6, the residues contacted with IRE are similar among the three complex systems, and frequent contact residues are mostly hydrophobic residues, such as L718, F723, V726, A743, I744, M766, L777, L788, I789, M790, L792, M793, P794, L799, and L844. Therefore, the hydrophobic interactions play an important role in the stabilization of IRE at the binding pocket. To further consider the compactness of the hydrophobic pocket, the R_g values were also calculated for the mentioned hydrophobic residues consisting of the binding site. As shown in Figure 7D–F, the R_g values distribute broadly in the APO systems, whereas IRE binding shifts the R_g distribution to a relatively smaller area with increased R_g values. These observations indicate that the hydrophobic pocket is relatively more compact but less stable in the absence of IRE, and upon IRE binding, the pocket enlarges slightly to better

accommodate IRE and becomes more stable. Comparing the three complex systems, the R_g value of these hydrophobic residues in the G719S system is the biggest with an average of $8.37 \pm 0.12 \text{ \AA}$, whereas for the other two mutants, the average R_g values are similar with a smaller distribution area. In short, the results suggest that the compactness and stability of the hydrophobic cavity in the G719S-IRE system are weaker than the other two IRE-bound systems, which might affect the stability of IRE at the binding site.

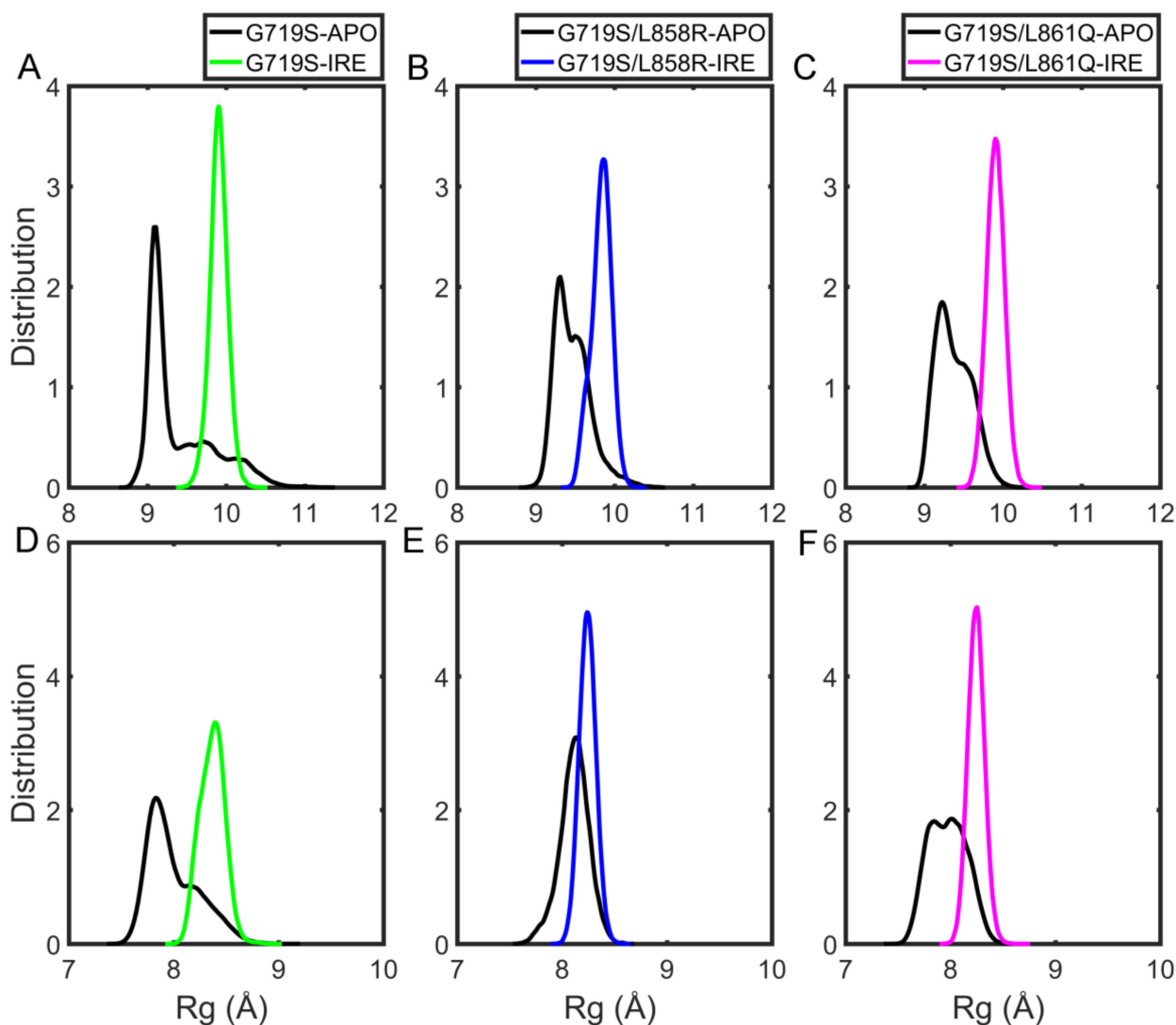


Figure 7. The probability distribution profiles of the R_g values of the binding site in the six systems considering three replicates together: (A) G719S mutants; (B) G719S/L858R mutants; (C) G719S/L861Q mutants. The probability distribution profiles of the R_g values of hydrophobic residues in the six systems considering three replicates together: (D) G719S mutants; (E) G719S/L858R mutants; (F) G719S/L861Q mutants.

IRE majorly occupies the ATP-binding regions with the benzene ring together with halogen atom substituents embedded in the non-ATP-contact region (hydrophobic pocket I) (Figure 1B). The vicinity residues around the substituted fluorine atom contain K745 and E762. The fluorine atom has a strong negative charge, which could attract the positively charged K745, thus the stability of the benzene ring part at the pocket could be enhanced by the positive electrostatic interaction. The distance between the fluorine atom and polar nitrogen atom of K745 was calculated. From Figure 8A, the distance distribution area is relatively larger in the G719S-IRE system than that in the other two systems. In addition, the electronegativity of the fluorine atom also makes an effect on the negatively charged

E762. Similarly, the distance between the fluorine atom and polar oxygen atoms of E762 was calculated (Figure 8B). For the G719S-IRE system, E762 is significantly distant from the fluorine atom, suggesting that the repulsion exerted by the fluorine atom on E762 is larger than that in other systems, which results in E762 moving to the outside of the binding pocket, thereby disrupting the KE salt bridge. Such a change might expand the solvent-exposed hydrophobic surface, and then weaken the stability of hydrophobic pocket I. Therefore, we quantified the area of the benzene ring part located exposed to the solvent through the solvent-accessible surface area (SASA) calculation. It is observed that obvious overlaps for SASA value distribution exist among the three complex systems, but the value is recognized to fluctuate to a larger area in the G719S-IRE system than in other systems. The observation suggested that the hydrophobic environment for the benzene ring binding is not stable as the one in other complex systems (Figure 8C), which might contribute less affinity of IRE toward EGFR^{G719S} mutant than other EGFR mutants with rare–rare and rare–classical combination.

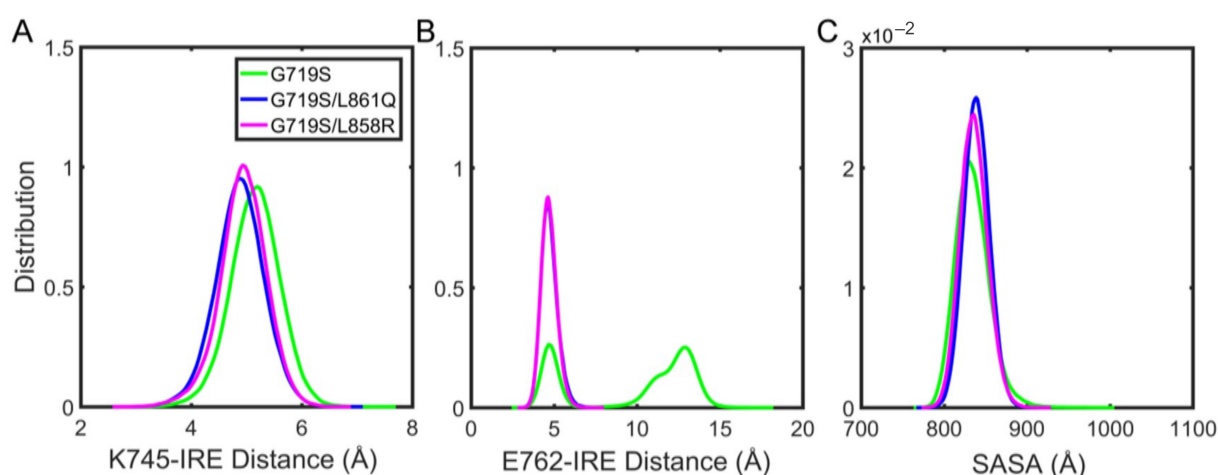


Figure 8. The probability distribution profiles of (A) the distance between atoms K745/NZ and IRE/F, (B) the distance between atoms E762/(OE1, OE2) and IRE/F, and (C) the solvent-accessible surface area (SASA) for the binding region of IRE benzene.

2.6. Stronger Binding Affinity of IRE with EGFR^{G719S/L858R} Mutant

Steered molecular dynamics (SMD) simulation is an approximate and quicker approach for simulating ligand dissociation by applying an external force. To explore the binding affinity of IRE toward EGFR with various mutations, SMD simulations were performed to pull IRE out of the binding site, which could provide physical quantity, such as free energy profile change along the chosen reaction coordinate. For the competitive inhibitor IRE, its dissociation from distinct EGFR mutants is closely related to the inhibitory activity. As seen in Figure 9, pulling IRE out of EGFR^{G719S/L858R} mutant needs the most work (roughly 83 kcal/mol) to be done, while they work in the G719S/L861Q system is a bit higher than the one in the G719S system. In short, the free energy change difference reveals that IRE has a stronger binding affinity with EGFR harboring combined mutations than a single rare mutation alone. Additionally, a rare mutation combined with a classical L858R mutation may be a strong indicator of EGFR sensitivity to EGFR inhibitors.

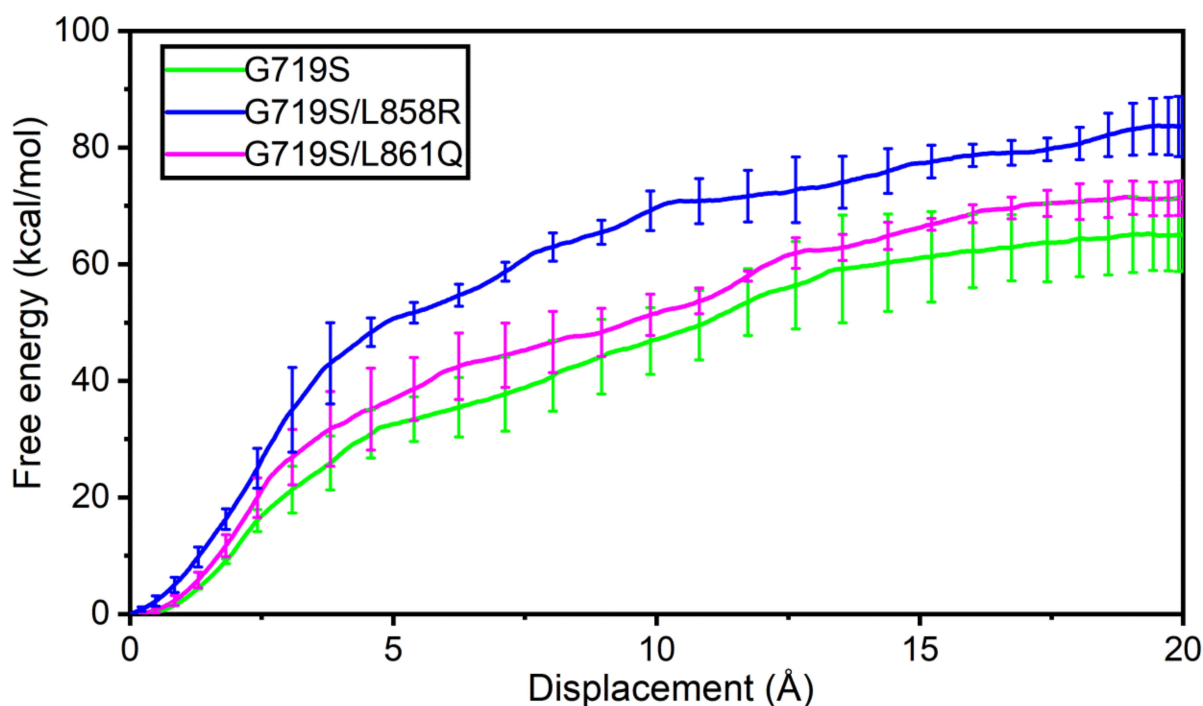


Figure 9. The average work profile for IRE dissociation with error bars of each complex system during the SMD simulations. The bars represent the standard deviation of the external works.

3. Conclusions

This study provides insights into the molecular basis by which EGFR with complex mutations shows better sensitivity for IRE than with a single rare mutation. In particular, the EGFR^{G719S-L858R} mutant exhibits the strongest sensitivity for IRE among the simulated systems, which suggests that a rare EGFR mutation combined with a classical L858R mutation may be a strong indicator of sensitivity to EGFR inhibitor. In the absence of IRE, the computational structural analyses uncovered that minor mutations favor the α C-in conformation, and the hydrogen bonds formed within the tetrad of residues S768 (in the α C-helix), D770 (in the α C- β 4 loop), Y827 (in the α E-helix), and R831 (in the catalytic loop) play an essential role for the stability of α C-helix and the maintenance of KE salt bridge. In the presence of IRE, the α C-helix is relatively more flexible for the EGFR^{G719S} mutant and the KE salt bridge is broken. The important interactions contributing to the stability of α C-helix are weakened in the IRE-bound EGFR^{G719S} system, whereas are enhanced in the IRE-bound EGFR^{G719S-L858R} system. The hydrogen bonds introduced by the second mutated residue also contribute to the stability of α C-helix. Moreover, the free energy change profiles reveal that the most work is needed to pull IRE out of the binding pocket in the EGFR^{G719S/L858R} system, and a bit above work is required to dissociate IRE in the EGFR^{G719S/L61Q} system than in the EGFR^{G719S} system, indicating that better affinity of IRE toward EGFR with complex mutations than single rare mutation. On balance, combined mutations show stronger sensitivity to inhibitor than the single minor mutation. An explicit understanding of rare mutations could provide useful insights for molecular drug design targeting EGFR uncommon mutations.

4. Methods

4.1. System Setup

The structures of mutated-EGFR in APO and IRE-bound forms were prepared based on the crystal structure of EGFR^{G719S/T790M} mutant in the complex of IRE (PDB ID: 3UG2 [34]). The completion of missing residues on the P-loop and α -loop was based on the template structure (PDB ID: 2ITY [35]). Moreover, the construction of the remaining missing amino acid residues in the crystal structure was achieved using the MODELLER [36] plugin in

Chimera [37]. Mutated residues substitution and the removal of small molecules both were carried out by PYMOL [38]. The full names and abbreviations of amino acids and other nouns in this work are listed in Tables S1 and S2.

4.2. Conventional Molecular Dynamics (CMD) Simulation

All prepared structures were employed as the initial structures for the conventional MD simulations, and all simulations were performed in AMBER 18 [39]. The resp force field parameters for inhibitor IRE were generated by using Gaussian 09 combined with the Antechamber module in the Amber 18 package. The AMBER force fields ff14SB [40] and GAFF [41] were used for describing the protein molecule and the ligand respectively. All simulated systems were neutralized with appropriate numbers of Na⁺ ions. Each neutralized system was solvated in an orthogonal box with TIP3P water molecules [42], leaving at least 10 Å between the solute atoms and the borders of the box. The particle mesh Ewald (PME) summation method [43] was applied to treat Coulombic interactions and the Shake algorithm [44] was used to constrain bond vibrations involving hydrogen atoms.

Three independent parallel molecular dynamics simulations were performed for each system. Four steps including the energy minimization, heating, equilibration, and production runs were carried out for the whole molecular simulation. Before MD simulations, each system was minimized for four 15,000 steps by means of the steepest descent and the conjugate gradient methods. The first three minimizations were performed with restraint force constants of 5.0, 2.0, and 1.0 kcal/mol/Å² applied to all heavy atoms, the protein backbone atoms, and the C α atoms, respectively. However, no restraint was imposed for the last minimization. Following the minimization steps, all simulated systems were heated from 0 to 298.15 K with a force constant of 5.0 kcal/mol/Å² applied to all heavy atoms. Next, each system was equilibrated through two 1000-ps and one 2000-ps steps with force constants of 1.0, 0.5, and 0.1 kcal/mol/Å² applied to all heavy atoms, the protein backbone atoms, and the C α atoms, respectively. Finally, an 800-ns production run without any restraints was performed for all equilibrated systems at the isothermal-isobaric (NPT) ensemble using Berendsen barostat [45] with a target pressure of 1 bar. Coordinates of the production runs were recorded every 2.0 ps.

4.3. Steered Molecular Dynamics (SMD) Simulation

SMD is a technique to manipulate the system from an initial configuration to a target configuration by applying an external force [46–48]. In our work, the SMD simulations were performed by applying forces along the line connecting the center of mass (COM) of IRE and residues within 6 Å of IRE. The applied force along the identified direction accelerated the release of IRE out of the binding pocket of EGFR. The three independent 2-ns SMD simulations in all systems were started with the spring constant of 10 kcal/mol/Å² controlling the strength of the pulling force. To avoid protein drift during the pulling process, the restriction was imposed on protein regions beyond 6 Å of IRE.

4.4. Principal Component Analysis

Principal component analysis (PCA) on EGFR in APO or inhibitor-bound forms based on the last 200-ns MD trajectories was performed using the CPPTRAJ module of AMBER18 software. The covariance matrix was firstly calculated using the coordinates of C α atoms of EGFR by RMSD fitting and then diagonalized to obtain the principal component eigenvectors. Each selected conformation was then projected in the collective coordinate space defined by the two largest principal component eigenvectors.

4.5. Probability Density Profile

The density distribution profiles were calculated in Matlab employing the Ksdensity function.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules27123844/s1>. Figure S1: The RMSDs of unbound EGFRs and IRE-bound EGFRs as a function of simulation time. Figure S2: The structure of EGFR with important regions highlighted. Figure S3: The flexibility of EGFRs during the simulations. Figure S4: The hydrogen-bond interactions contribute to the stability of the α C-helix. Figure S5: The hydrogen-bond interactions introduced by the co-occurring mutated residue. Figure S6: Contact map between IRE and EGFR in the three IRE-bound systems. Table S1: The full names and abbreviations of amino acids in this work. Table S2: The full names and abbreviations of nouns in this work. Movie S1: The motion of EGFR captured in PC1. Movie S2: The motion of EGFR captured in PC2.

Author Contributions: J.G. and M.L. (Miaomiao Li) conceived and designed the study. M.L. (Miaomiao Li) and M.L. (Mengrong Li) simulated the systems and analyzed the data. M.L. (Miaomiao Li), J.G., Y.X. and M.L. (Mengrong Li) wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the starting fund of Nanjing Agricultural University (680804014).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Molecular Dynamics trajectories and topology files of simulated systems in this work could be downloaded in <https://zenodo.org/record/6616150#YqWW6MjxTK>.

Acknowledgments: We thank the reviewers for their valuable comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yarden, Y.; Sliwkowski, M.X. Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 127–137. [[CrossRef](#)] [[PubMed](#)]
2. Day, K.C.; Lorenzatti Hiles, G.; Kozminsky, M.; Dawsey, S.J.; Paul, A.; Brose, L.J.; Shah, R.; Kunja, L.P.; Hall, C.; Palanisamy, N.; et al. HER2 and EGFR Overexpression Support Metastatic Progression of Prostate Cancer to Bone. *Cancer Res.* **2017**, *77*, 74–85. [[CrossRef](#)] [[PubMed](#)]
3. Gazdar, A.F. Activating and resistance mutations of EGFR in non-small-cell lung cancer: Role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene* **2009**, *28* (Suppl. 1), S24–S31. [[CrossRef](#)] [[PubMed](#)]
4. Arcila, M.E.; Nafa, K.; Chaft, J.E.; Rekhtman, N.; Lau, C.; Reva, B.A.; Zakowski, M.F.; Kris, M.G.; Ladanyi, M. EGFR exon 20 insertion mutations in lung adenocarcinomas: Prevalence, molecular heterogeneity, and clinicopathologic characteristics. *Mol. Cancer Ther.* **2013**, *12*, 220–229. [[CrossRef](#)]
5. Jackman, D.M.; Yeap, B.Y.; Sequist, L.V.; Lindeman, N.; Holmes, A.J.; Joshi, V.A.; Bell, D.W.; Huberman, M.S.; Halmos, B.; Rabin, M.S.; et al. Exon 19 deletion mutations of epidermal growth factor receptor are associated with prolonged survival in non-small cell lung cancer patients treated with gefitinib or erlotinib. *Clin. Cancer Res.* **2006**, *12*, 3908–3914. [[CrossRef](#)]
6. Sharma, S.V.; Bell, D.W.; Settleman, J.; Haber, D.A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **2007**, *7*, 169–181. [[CrossRef](#)]
7. Paez, J.G.; Jänne, P.A.; Lee, J.C.; Tracy, S.; Greulich, H.; Gabriel, S.; Herman, P.; Kaye, F.J.; Lindeman, N.; Boggon, T.J.; et al. EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science* **2004**, *304*, 1497–1500. [[CrossRef](#)]
8. Shigematsu, H.; Lin, L.; Takahashi, T.; Nomura, M.; Suzuki, M.; Wistuba, I.I.; Fong, K.M.; Lee, H.; Toyooka, S.; Shimizu, N.; et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J. Natl. Cancer Inst.* **2005**, *97*, 339–346. [[CrossRef](#)]
9. Skříčková, J.; Kadlec, B.; Venclíček, O. [Non-small cell lung cancer]. *Vnitř. Lek.* **2018**, *63*, 861–874. [[CrossRef](#)]
10. Kobayashi, Y.; Mitsudomi, T. Not all epidermal growth factor receptor mutations in lung cancer are created equal: Perspectives for individualized treatment strategy. *Cancer Sci.* **2016**, *107*, 1179–1186. [[CrossRef](#)]
11. Mok, T.S.; Wu, Y.L.; Thongprasert, S.; Yang, C.H.; Chu, D.T.; Saijo, N.; Sunpaweravong, P.; Han, B.; Margono, B.; Ichinose, Y.; et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N. Engl. J. Med.* **2009**, *361*, 947–957. [[CrossRef](#)] [[PubMed](#)]
12. Zhou, C.; Wu, Y.-L.; Chen, G.; Feng, J.; Liu, X.-Q.; Wang, C.; Zhang, S.; Wang, J.; Zhou, S.; Ren, S.; et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): A multicentre, open-label, randomised, phase 3 study. *Lancet Oncol.* **2011**, *12*, 735–742. [[CrossRef](#)]
13. Sequist, L.V.; Yang, J.C.; Yamamoto, N.; O’Byrne, K.; Hirsh, V.; Mok, T.; Geater, S.L.; Orlov, S.; Tsai, C.M.; Boyer, M.; et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J. Clin. Oncol.* **2013**, *31*, 3327–3334. [[CrossRef](#)]

14. Lynch, T.J.; Bell, D.W.; Sordella, R.; Gurubhagavatula, S.; Okimoto, R.A.; Brannigan, B.W.; Harris, P.L.; Haserlat, S.M.; Supko, J.G.; Haluska, F.G.; et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **2004**, *350*, 2129–2139. [[CrossRef](#)]
15. Kris, M.G.; Natale, R.B.; Herbst, R.S.; Lynch, T.J., Jr.; Prager, D.; Belani, C.P.; Schiller, J.H.; Kelly, K.; Spiridonidis, H.; Sandler, A.; et al. Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: A randomized trial. *JAMA* **2003**, *290*, 2149–2158. [[CrossRef](#)] [[PubMed](#)]
16. Lee, D.H.; Han, J.Y.; Lee, H.G.; Lee, J.J.; Lee, E.K.; Kim, H.Y.; Kim, H.K.; Hong, E.K.; Lee, J.S. Gefitinib as a first-line therapy of advanced or metastatic adenocarcinoma of the lung in never-smokers. *Clin. Cancer Res.* **2005**, *11*, 3032–3037. [[CrossRef](#)] [[PubMed](#)]
17. Beau-Faller, M.; Prim, N.; Ruppert, A.M.; Nanni-Metellus, I.; Lacave, R.; Lacroix, L.; Escande, F.; Lizard, S.; Pretet, J.L.; Rouquette, I.; et al. Rare EGFR exon 18 and exon 20 mutations in non-small-cell lung cancer on 10 117 patients: A multicentre observational study by the French ERMETIC-IFCT network. *Ann. Oncol.* **2014**, *25*, 126–131. [[CrossRef](#)]
18. Kobayashi, S.; Canepa, H.M.; Bailey, A.S.; Nakayama, S.; Yamaguchi, N.; Goldstein, M.A.; Huberman, M.S.; Costa, D.B. Compound EGFR mutations and response to EGFR tyrosine kinase inhibitors. *J. Thorac. Oncol.* **2013**, *8*, 45–51. [[CrossRef](#)]
19. Wu, J.Y.; Yu, C.J.; Chang, Y.C.; Yang, C.H.; Shih, J.Y.; Yang, P.C. Effectiveness of tyrosine kinase inhibitors on “uncommon” epidermal growth factor receptor mutations of unknown clinical significance in non-small cell lung cancer. *Clin. Cancer Res.* **2011**, *17*, 3812–3821. [[CrossRef](#)]
20. Liu, Y.; Wu, B.Q.; Zhong, H.H.; Hui, P.; Fang, W.G. Screening for EGFR and KRAS mutations in non-small cell lung carcinomas using DNA extraction by hydrothermal pressure coupled with PCR-based direct sequencing. *Int. J. Clin. Exp. Pathol.* **2013**, *6*, 1880–1889.
21. Peng, L.; Song, Z.G.; Jiao, S.C. Efficacy analysis of tyrosine kinase inhibitors on rare non-small cell lung cancer patients harboring complex EGFR mutations. *Sci. Rep.* **2014**, *4*, 6104. [[CrossRef](#)] [[PubMed](#)]
22. Chiu, C.H.; Yang, C.T.; Shih, J.Y.; Huang, M.S.; Su, W.C.; Lai, R.S.; Wang, C.C.; Hsiao, S.H.; Lin, Y.C.; Ho, C.L.; et al. Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor Treatment Response in Advanced Lung Adenocarcinomas with G719X/L861Q/S768I Mutations. *J. Thorac. Oncol.* **2015**, *10*, 793–799. [[CrossRef](#)] [[PubMed](#)]
23. Baek, J.H.; Sun, J.M.; Min, Y.J.; Cho, E.K.; Cho, B.C.; Kim, J.H.; Ahn, M.J.; Park, K. Efficacy of EGFR tyrosine kinase inhibitors in patients with EGFR-mutated non-small cell lung cancer except both exon 19 deletion and exon 21 L858R: A retrospective analysis in Korea. *Lung Cancer* **2015**, *87*, 148–154. [[CrossRef](#)] [[PubMed](#)]
24. Xu, J.; Jin, B.; Chu, T.; Dong, X.; Yang, H.; Zhang, Y.; Wu, D.; Lou, Y.; Zhang, X.; Wang, H.; et al. EGFR tyrosine kinase inhibitor (TKI) in patients with advanced non-small cell lung cancer (NSCLC) harboring uncommon EGFR mutations: A real-world study in China. *Lung Cancer* **2016**, *96*, 87–92. [[CrossRef](#)] [[PubMed](#)]
25. Shan, Y.; Eastwood, M.P.; Zhang, X.; Kim, E.T.; Arkhipov, A.; Dror, R.O.; Jumper, J.; Kuriyan, J.; Shaw, D.E. Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell* **2012**, *149*, 860–870. [[CrossRef](#)]
26. Yun, C.H.; Mengwasser, K.E.; Toms, A.V.; Woo, M.S.; Greulich, H.; Wong, K.K.; Meyerson, M.; Eck, M.J. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2070–2075. [[CrossRef](#)]
27. Harrison, P.T.; Vyse, S.; Huang, P.H. Rare epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer. *Semin. Cancer Biol.* **2020**, *61*, 167–179. [[CrossRef](#)]
28. Tsai, C.J.; Nussinov, R. The molecular basis of targeting protein kinases in cancer therapeutics. *Semin. Cancer Biol.* **2013**, *23*, 235–242. [[CrossRef](#)]
29. Liao, J.J. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J. Med. Chem.* **2007**, *50*, 409–424. [[CrossRef](#)]
30. Knapp, B.; Ospina, L.; Deane, C.M. Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J. Chem. Theory Comput.* **2018**, *14*, 6127–6138. [[CrossRef](#)]
31. Shan, Y.; Arkhipov, A.; Kim, E.T.; Pan, A.C.; Shaw, D.E. Transitions to catalytically inactive conformations in EGFR kinase. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 7270–7275. [[CrossRef](#)] [[PubMed](#)]
32. Zheng, J.; Knighton, D.R.; Xuong, N.H.; Taylor, S.S.; Sowadski, J.M.; Ten Eyck, L.F. Crystal structures of the myristylated catalytic subunit of cAMP-dependent protein kinase reveal open and closed conformations. *Protein Sci.* **1993**, *2*, 1559–1573. [[CrossRef](#)] [[PubMed](#)]
33. Taylor, S.S.; Knighton, D.R.; Zheng, J.; Ten Eyck, L.F.; Sowadski, J.M. cAMP-dependent protein kinase and the protein kinase family. *Faraday Discuss.* **1992**, 143–152. [[CrossRef](#)] [[PubMed](#)]
34. Yoshikawa, S.; Kukimoto-Niino, M.; Parker, L.; Handa, N.; Terada, T.; Fujimoto, T.; Terazawa, Y.; Wakiyama, M.; Sato, M.; Sano, S.; et al. Structural basis for the altered drug sensitivities of non-small cell lung cancer-associated mutants of human epidermal growth factor receptor. *Oncogene* **2013**, *32*, 27–38. [[CrossRef](#)] [[PubMed](#)]
35. Yun, C.H.; Boggon, T.J.; Li, Y.; Woo, M.S.; Greulich, H.; Meyerson, M.; Eck, M.J. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: Mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **2007**, *11*, 217–227. [[CrossRef](#)]
36. Fiser, A.; Sali, A. Modeller: Generation and refinement of homology-based protein structure models. *Methods Enzymol.* **2003**, *374*, 461–491.
37. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [[CrossRef](#)]

38. Pan, L.; Aller, S.G. Tools and procedures for visualization of proteins and other biomolecules. *Curr. Opin. Struct. Biol.* **2015**, *110*, 19.12.11–19.12.47. [[CrossRef](#)]
39. Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.V.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688. [[CrossRef](#)]
40. Maier, J.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [[CrossRef](#)]
41. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [[CrossRef](#)] [[PubMed](#)]
42. Price, D.; Brooks, C.L., 3rd. A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* **2004**, *121*, 10096–10103. [[CrossRef](#)] [[PubMed](#)]
43. Harvey, M.J.; De Fabritiis, G. An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware. *J. Chem. Theory Comput.* **2009**, *5*, 2371–2377. [[CrossRef](#)] [[PubMed](#)]
44. Ryckaert, J.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [[CrossRef](#)]
45. Eslami, H.; Mojahedi, F.; Moghadasi, J. Molecular dynamics simulation with weak coupling to heat and material baths. *J. Chem. Phys.* **2010**, *133*, 084105. [[CrossRef](#)]
46. Xu, L.; Hasin, N.; Shen, M.; He, J.; Xue, Y.; Zhou, X.; Perrett, S.; Song, Y.; Jones, G.W. Using steered molecular dynamics to predict and assess Hsp70 substrate-binding domain mutants that alter prion propagation. *PLoS Comput. Biol.* **2013**, *9*, e1002896. [[CrossRef](#)]
47. Shen, L.; Shen, J.; Luo, X.; Cheng, F.; Xu, Y.; Chen, K.; Arnold, E.; Ding, J.; Jiang, H. Steered molecular dynamics simulation on the binding of NNRTI to HIV-1 RT. *Biophys. J.* **2003**, *84*, 3547–3563. [[CrossRef](#)]
48. Zhang, Z.; Santos, A.P.; Zhou, Q.; Liang, L.; Wang, Q.; Wu, T.; Franzen, S. Steered molecular dynamics study of inhibitor binding in the internal binding site in dehaloperoxidase-hemoglobin. *Biophys. Chem.* **2016**, *211*, 28–38. [[CrossRef](#)]