

RESEARCH ARTICLE

Structural characterization of EGFR exon 19 deletion mutation using molecular dynamics simulation

Mahlet Z. Tamirat¹, Marika Koivu^{2,3,4}, Klaus Elenius^{2,3,5}, Mark S. Johnson^{1*}

1 Structural Bioinformatics Laboratory, Biochemistry, Faculty of Science and Engineering, Åbo Akademi University, Turku, Finland, **2** Medicity Research Laboratories and Institute of Biomedicine, University of Turku, Turku, Finland, **3** Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland, **4** Turku Doctoral Programme of Molecular Medicine, University of Turku, Turku, Finland, **5** Department of Oncology and Radiotherapy, University of Turku and Turku University Hospital, Turku, Finland

* johnson4@abo.fi

OPEN ACCESS

Citation: Tamirat MZ, Koivu M, Elenius K, Johnson MS (2019) Structural characterization of EGFR exon 19 deletion mutation using molecular dynamics simulation. PLoS ONE 14(9): e0222814. <https://doi.org/10.1371/journal.pone.0222814>

Editor: Freddie Salisbury, Jr, Wake Forest University, UNITED STATES

Received: May 20, 2019

Accepted: September 6, 2019

Published: September 19, 2019

Copyright: © 2019 Tamirat et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This research is part of the Drug Discovery and Diagnostics strategic area of Åbo Akademi University and University of Turku, and was supported by grants to M.S.J. from the Academy of Finland (308317, 320005), Sigrid Jusélius Foundation and Tor, Joe and Pentti Borg memorial fund. M.Z.T. is funded by the Åbo Akademi Doctoral Network of Informational and Structural Biology. K.E. is supported by Academy

Abstract

Epidermal growth factor receptor (EGFR) is a tyrosine kinase receptor important in diverse biological processes including cell proliferation and survival. Upregulation of EGFR activity due to over-expression or mutation is widely implicated in cancer. Activating somatic mutations of the EGFR kinase are postulated to affect the conformation and/or stability of the protein, shifting the EGFR inactive-active state equilibrium towards the activated state. Here, we examined a common EGFR deletion mutation, Δ^{746} ELREA⁷⁵⁰, which is frequently observed in non-small cell lung cancer patients. By using molecular dynamics simulation, we investigated the structural effects of the mutation that lead to the experimentally reported increases in kinase activity. Simulations of the active form wild-type and Δ ELREA EGFRs revealed the deletion stabilizes the α C helix of the kinase domain, which is located adjacent to the deletion site, by rigidifying the flexible β 3- α C loop that accommodates the ELREA sequence. Consequently, the α C helix is stabilized in the “ α C-in” active conformation that would prolong the time of the activated state. Moreover, in the mutant kinase, a salt bridge between E762 and K745, which is key for EGFR activity, was also stabilized during the simulation. Additionally, the interaction between EGFR and ATP was favored by Δ ELREA EGFR over wild-type EGFR, as reflected by the number of hydrogen bonds formed and the free energy of binding. Simulation of inactive EGFR suggested the deletion would promote a shift from the inactive conformation towards active EGFR, which is supported by the inward movement of the α C helix. The MDS results also align with the effects of tyrosine kinase inhibitors on Δ ELREA and wild-type EGFR lung cancer cell lines, where more pronounced inhibition was observed against Δ ELREA than for wild-type EGFR by inhibitors recognizing the active kinase conformation.

of Finland Grants 274728 and 316796, the Cancer Foundation of Finland, and Turku University Central Hospital. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Epidermal growth factor receptor (EGFR) kinase is a tyrosine kinase involved in multiple cellular processes, such as cell proliferation, differentiation, migration and survival [1]. EGFR, also known as ErbB1/Her1, is a member of the ErbB family of receptor kinases, which also includes ErbB2, ErbB3 and ErbB4 (ErbBs). The ErbBs are fundamental to the development and growth of organisms, however, anomalies with their regulation and/or signaling activity often associates them with various cancers, making them key therapeutic targets [2, 3]. Already in the 1980's it was observed that avian erythroblastosis retrovirus encoded chordate-species EGFR kinase domains that, untethered to a growth factor sensing ectodomain, was associated with development of cancers in the same chordate due to retrovirus infection [4, 5]. More recently, the extent of somatic mutations occurring in ErbBs from cancer patients has come to the forefront of research and patient treatment. For example, mutations in EGFR and amplification of ErbB2 predict sensitivity to EGFR and ErbB2 targeting cancer drugs, respectively [6, 7, 8]. Moreover, mutations observed in ErbB4 in non-small cell lung cancer patients experimentally lead to changes in ErbB2-ErbB4 heterodimer signaling promoting cell proliferation but not differentiation [9], and to increased phosphorylation likely because of stabilization of the ErbB4 active dimer state [10].

ErbBs are composed of an N-terminal extracellular growth factor binding domain, a helical transmembrane domain, a juxtamembrane sequence, an intracellular kinase domain that has an ATP binding / catalytic site, and a C-terminal tail that becomes phosphorylated during receptor activation (Fig 1A). ErbB activation is driven by growth factor binding to the ectodomain leading to dimer formation; ErbB2 itself does not bind growth factors but contributes to cell signaling by forming heterodimers with other EGFR family members that do bind the growth factor. The transition from monomer to either homodimer or heterodimer incurs a very large conformational change of the ectodomain. This in turn results in activation of the intracellular kinase domains through asymmetric dimer interactions, followed by autophosphorylation of tyrosine residues of the C-terminal tail, and triggering of multiple signal transduction pathways [1, 11–13].

The tyrosine kinase domain is a highly conserved bilobal structure consisting of the N-lobe and C-lobe, which are separated by the catalytic site where ATP binds (Fig 1B). The N-lobe is predominantly composed of β -sheets and constitutes key structural elements—such as the α C helix and P-loop—important for kinase activation and catalysis. The C-lobe, in contrast, is largely alpha helical and includes the functionally important activation and catalytic loops [14–16].

The EGFR kinase domain, like other kinases, is present in inactive and active conformations [17], the structures of which are well documented in the Protein Data Bank (PDB [18]). Monomeric ErbBs are inactive, but in the growth-factor induced dimer state the asymmetric interaction of the activator kinase domain and juxtamembrane segment B of the receiver kinase domain leads to the conformational changes required for phosphorylation [19]. The active and inactive conformations of kinase domain monomer structures mainly differ in the orientation of the α C helix and the activation loop (A-loop) (Fig 2A). In the active state, the α C helix is oriented towards the ATP binding pocket (“ α C-in” conformation), forming an ion-pair interaction between a conserved glutamate and a lysine residue. Additionally, the A-loop that contains the aspartate-phenylalanine-glycine “DFG” motif attains an open and extended conformation, with the catalytic aspartate of the motif pointing towards the ATP binding site (“DFG-in” conformation). In contrast, two conformations of the inactive EGFR kinase are observed; the Src-like inactive state and the “DFG-out” state. The majority of the inactive EGFR structures in the PDB exist in the Src-like inactive conformation, where the α C

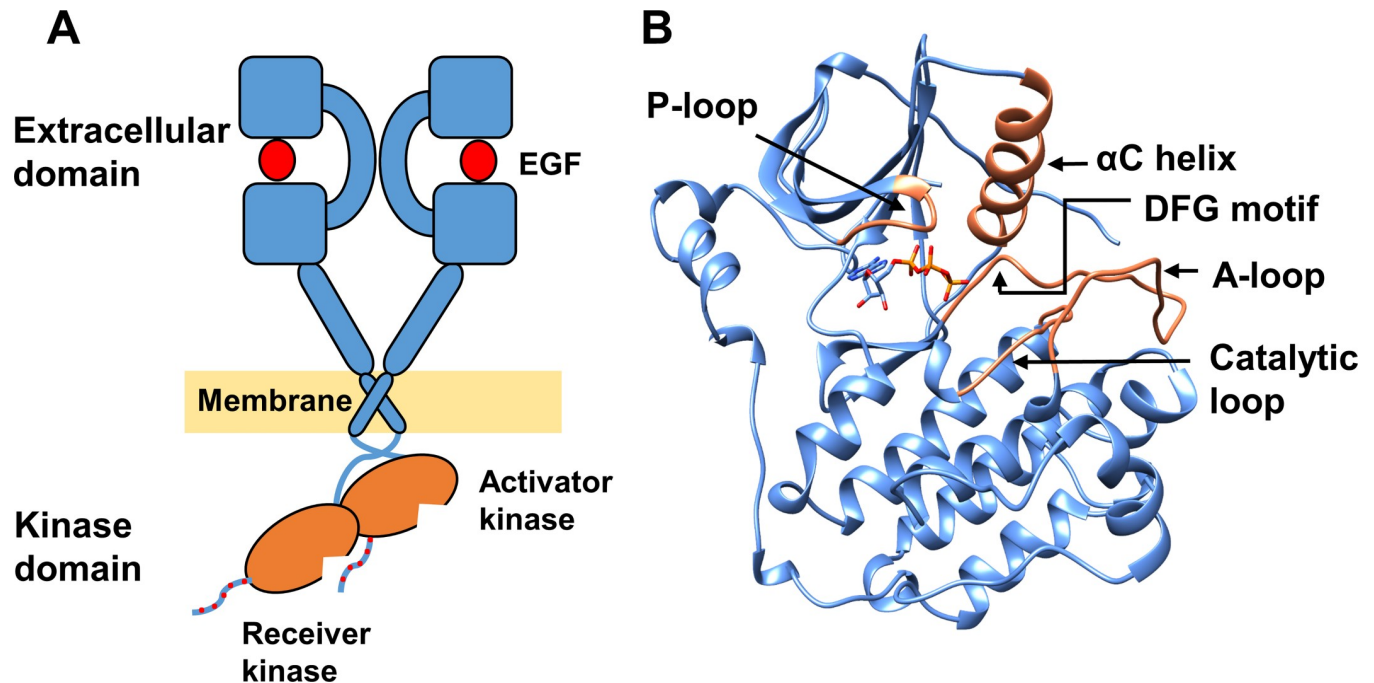


Fig 1. The EGFR structure. (A) EGFR dimer–growth factor (red circle) -bound extracellular domains, transmembrane domains and asymmetrically dimerized intracellular kinase domains (orange) with phosphotyrosines at the C-terminal tail marked in red dots. (B) The kinase domain of EGFR (PDB ID 2ITX). Missing loops have been built using other EGFR structures. Bound ANP in the crystal structure was replaced by ATP (sticks). Key structural features are shown in orange.

<https://doi.org/10.1371/journal.pone.0222814.g001>

helix is positioned away from the ATP binding site and assumes the “ α C-out” position, breaking the conserved glutamate-lysine salt bridge. Moreover, this conformation exhibits a small hydrophobic helix at the N-terminus of the A-loop that packs against the α C helix, with the DFG motif in the DFG-in conformation. On the other hand, the DFG-out inactive conformation displays an α C helix that partly overlaps with the Src-like conformation and has an extended A-loop with a flipped DFG motif that positions the aspartate away from the binding pocket in the DFG-out conformation. These distinct conformational changes play a vital role in the regulation of kinase activity in the ErbBs [19–22].

Increased EGFR tyrosine kinase activity, due to over-expression and/or somatic mutation (s), has been linked to a variety of human cancers [23, 24]. In particular, activating mutations of the EGFR kinase domain, including single amino acid substitutions, insertions and deletions, are commonly observed in non-small cell lung cancer (NSCLC) patients [6, 7, 21, 25]. The most prevalent EGFR mutations in NSCLC are L858R and exon 19 deletions (Fig 2B), which respectively account for 41% and 44% of all EGFR mutations [17, 26]. Although NSCLC patients with these activating mutations respond to first generation tyrosine kinase inhibitors (TKIs), they often develop resistance due to a secondary T790M “gatekeeper” mutation [27, 28]. Unlike exon 19 deletions, 3D structures of EGFR with the L858R and T790M mutations have been determined [22, 29–31].

Exon 19 deletions have been reported to increase EGFR autophosphorylation, and promote cell survival by selectively activating AKT and STAT pathways [7, 32, 33]. Exon 19 deletions include a number of variants differing in the length of the deleted amino acid sequence, the most common subtype being Δ^{746} ELREA⁷⁵⁰ [26, 34] (numbering from Uniprot ID P00533 and PDB ID 2ITX); and a majority of the subtypes share the deletion of the amino acids leucine-arginine-glutamic acid. These deletion mutations are located between the β 3 strand and

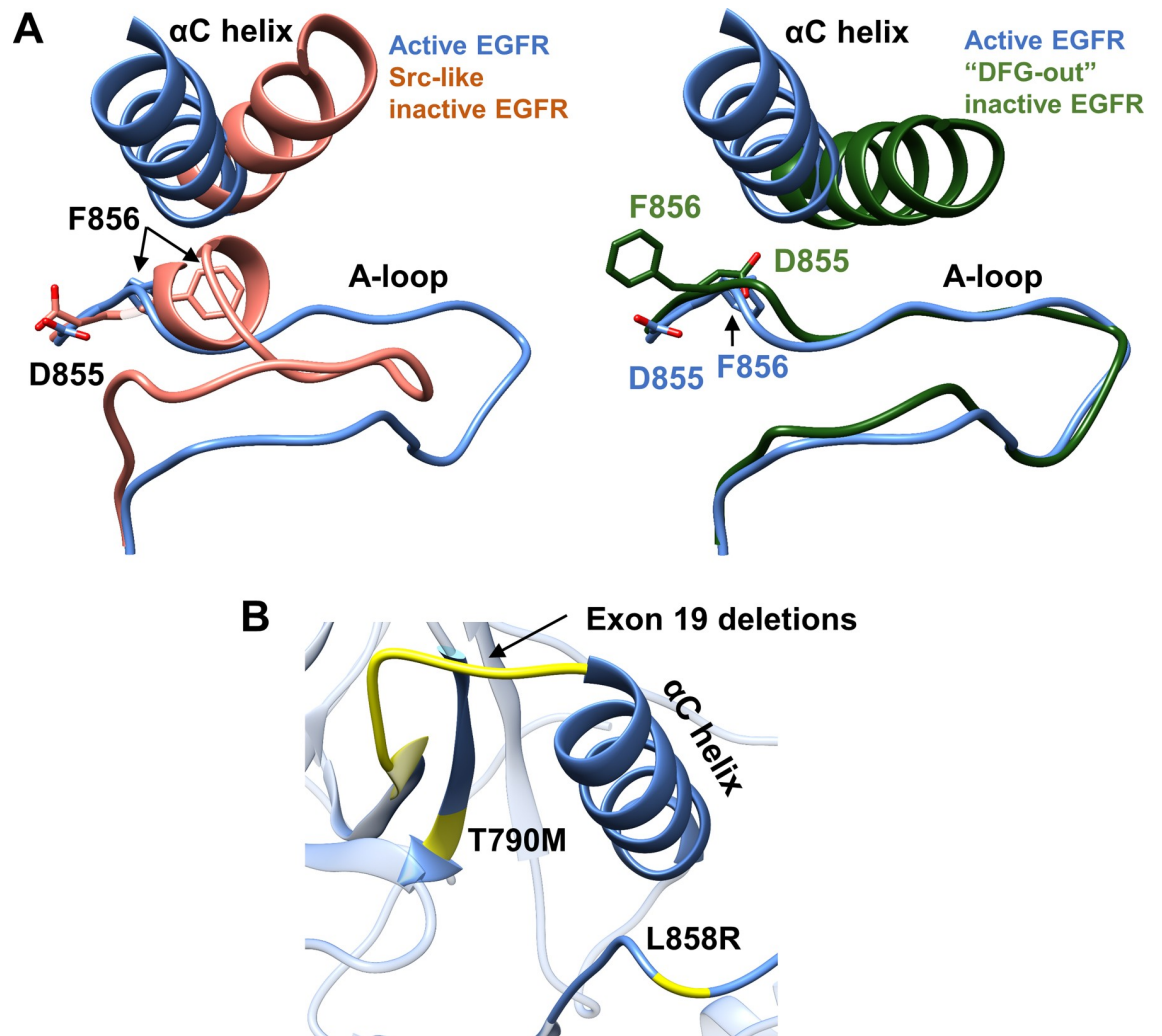


Fig 2. Active versus inactive structures of the EGFR kinase domain and common EGFR somatic mutations. (A) Superimposed structure of active EGFR (PDB ID 2GS2; blue) with (left) Src-like inactive EGFR (PDB ID 2GS7; orange); and (right) with DFG-out inactive EGFR (PDB ID 4I21; green). Aspartate and phenylalanine from the DFG motif are shown in sticks. (B) Main-chain location of the three common EGFR kinase somatic mutations (yellow) relative to the α C helix.

<https://doi.org/10.1371/journal.pone.0222814.g002>

the α C helix within the β 3- α C loop of the N-lobe of the kinase domain. The α C helix, positioned in the vicinity of the ATP binding / catalytic site and forming part of the asymmetric dimer interface, is fundamental for EGFR kinase activity. Hence, deletions within the β 3- α C loop can be expected to exert their effects by altering local structural features on or about the α C helix, as well as by modulating tyrosine kinase activity and subsequent events.

Here, we have sought to understand the structural changes that take place as a result of the most prevalent exon 19 deletion in EGFR, Δ ELREA, and the likely functional consequences due to changes in EGFR tyrosine kinase activity. To do so, the structures of the wild-type and mutant-active and inactive kinase domains—were assessed using molecular dynamics simulations (MDS) in order to probe the effects on domain conformation, local structure and consequences for biological activity.

Materials and methods

Structure preparation

X-ray structures of wild-type apo active (PDB code: 2GS2, 2.8 Å resolution; [19]), ATP-bound active (2ITX, 2.98 Å; [29]) and apo inactive (2GS7, 2.6 Å; [19]) served as a basis for the composite and model structures used in this study. The Src-like inactive structure is studied here as opposed to the DFG-out inactive conformation, on the basis that the majority of the experimentally resolved EGFR inactive structures are Src-like and include bound key TKIs unlike those of the DFG-out conformation. The bound ligand was removed from the apo inactive EGFR structure; for the ATP-bound active EGFR, the ligand ANP in 2ITX was replaced by ATP and a Mg^{2+} ion was added. Missing loops in all three structures were built using the loops from other EGFR structures (see S1 Table for details). Mutant Δ ELREA EGFRs were then modeled for each of the above wild-type structures using the Modeller program [35] available in the Chimera visualization tool [36]. Altogether, this resulted in three wild-type and three Δ ELREA EGFR structures, which were subsequently prepared for MDS using the protein preparation wizard in Maestro [37]: hydrogen atoms were added, optimal protonation states of ionizable side chains at pH 7.0 were determined using PROPKA [38], and the structures were energy minimized.

Molecular dynamics simulation

Classical MDS was used to probe the dynamics of the six wild-type and Δ ELREA EGFR structural models using the AMBER package (version 18) [39] and ff14SB force field [40]. Parameters for ATP [41] were used for the ATP-complexed systems. The proteins were solvated with explicit TIP3P water molecules [42] in an octahedral box, leaving 10 Å between protein surface atoms and edge of the box. The systems were neutralized by adding sodium counter ions. Additional Na^+/Cl^- ions were added to the simulation box to achieve a 150 mM salt concentration. Periodic boundary conditions were employed and the particle-mesh Ewald algorithm [43] was used to treat electrostatic interactions with a distance cutoff of 9 Å. Prior to conducting the production simulation, 5000 cycles of steepest descent and conjugate gradient energy minimization were carried out. The minimization was initiated by introducing a $25 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ restraint on solute atoms that was systematically reduced to $0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ over the total minimization step. The systems were then heated from 100 K to 300 K during 100 ps with a $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ restraint on solute atoms. Subsequently, a 900 ps equilibration at constant pressure was employed while reducing the restraint gradually to $0.1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. The equilibration protocol was concluded with an unrestrained 5 ns simulation. Finally, the production simulation was carried out for 100 ns at constant temperature (300 K) and pressure (1 bar) that was maintained using the Berendsen algorithm [44] with 5 ps coupling constant. In order to sample more conformational space, the simulations were performed in triplicate using different initial velocities, assigned by a pseudo-random number generator. Coordinates were saved every 10 ps and the resulting trajectories were analyzed further using the programs CPPTRAJ [45] and VMD [46]. CPPTRAJ was used to compute $C\alpha$ -atom root-mean-squared fluctuations (RMSF) and to examine hydrogen bonds. A hydrogen bond was defined as a donor-acceptor distance of less than or equal to 3.5 Å and a bond angle of greater than or equal to 135° . The root-mean-squared deviation (RMSD) was calculated over $C\alpha$ atoms using Chimera. VMD was critical for visualizing trajectories and monitoring distances between residues.

Principal component analysis

Principal component analysis (PCA) is a multivariate statistical method that was used in this study to reveal any dominant patterns of motion recorded during MDS. Initially, the recorded MD trajectory frames were superimposed on the average structure to remove global translational and rotational motions. Subsequently, a coordinate covariance matrix was generated for backbone atoms using the 3D positional coordinates from the trajectory frames. Diagonalizing this matrix generates eigenvectors and corresponding eigenvalues that respectively describe the direction and magnitude of motion. In this study, PCA was carried out for backbone atoms of the α C-helix and β 3- α C loop in both wild-type and Δ ELREA active EGFRs using the program CPPTRAJ. The normal mode data generated for the top three principal components (PC) were analyzed using the Normal Mode Wizard (NMWiz) [47] included in VMD.

Free energy calculations

To assess the relative binding free energy, ΔG_{bind} , of ATP with the wild-type and Δ ELREA EGFRs, the molecular mechanics generalized Born surface area (MM-GBSA) module [48] available in the AMBER package was applied to the resultant trajectories from the last 50 ns of the simulation. The MM-GBSA method can be summarized as follows:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}})$$

Where G_{complex} is the free energy of the protein-ligand complex, G_{protein} is the free energy of the protein and G_{ligand} is the free energy of the ligand.

Drug response data for the EGFR Δ ELREA mutant

Drug response data from three publicly available databases, The Cancer Cell Line Encyclopedia (CCLE) [49], The Genomics of Drug Sensitivity in Cancer (GDSC) [50], and The Cancer Therapeutics Response Portal (CTRP; second version) [51] were downloaded from the websites <https://portals.broadinstitute.org/ccle>, <http://www.cancerxgene.org/>, and <https://ocg.cancer.gov/programs/ctd2/data-portal>. RStudio version 3.5.1 [52] was used to sort through the data and to collect drug response data for lung cancer cell lines that contain the EGFR Δ ELREA mutation and for lung cancer cell lines that are wild-type for all four ErbB receptors, EGFR, ErbB2, ErbB3, and ErbB4 (“EGFR wild-type”). Mutations in other ErbBs, in addition to EGFR, may affect the sensitivity to EGFR/ErbB inhibitors. Thus, only ErbB wild-type cell lines were included in the control group to reduce the possibility of the control response curve shifting due to known and unknown ErbB mutations other than the EGFR Δ ELREA mutation. To further control the effect of including or excluding the ErbB2, ErbB3 and ErbB4 mutant cell lines from the control group, the mean area-under-the-curve (AUC) values were calculated for both the cell lines wild-type for EGFR irrespective of the mutation status of ErbB2, ErbB3 and ErbB4, and for cell lines wild-type for all four ErbBs. The values were very similar to each other (data not shown). The R code executed to process the data, including the libraries used, is provided as a supplementary document (S7 File). One hundred percent cell survival was assigned to a concentration of 0 μ M and drug response values were transformed into percentages. The analyses were carried out for the ErbB tyrosine kinase inhibitors erlotinib, gefitinib, afatinib, and lapatinib.

Results

Kinases are dynamic proteins that exist in an equilibrium between active and inactive states, and access to the active catalytic state is highly regulated and of restricted duration in order to

prevent signaling effects that lead to abnormal biological effects. Whereas kinases exhibit similar active state conformations, individual kinases can adopt different inactive conformations [15, 16].

Like other kinases of the receptor tyrosine kinase family, EGFR forms a dimer on binding the growth factor ligand to the ectodomain, serving to bring the cytoplasmic kinase domains together to form an asymmetric complex and leading to kinase activation. The interface of the kinase asymmetric dimer involves interactions of the α C helix with the juxtamembrane B peptide from the receiver kinase domain and helices α H and α I of the activator kinase domain (Fig 3). The β 3- α C loop from the receiver kinase domain is not itself in contact with the activator kinase domain, but this loop directly precedes the α C helix that is in direct contact.

Here, we investigated a commonly observed activating EGFR kinase deletion, Δ ELREA, in order to evaluate any structural changes resulting from the deletion reported to increase kinase activity [53, 54]. All-atom MDS was employed on apo forms of wild-type and Δ ELREA EGFRs, both in the active and inactive conformations. Additionally, simulation of wild-type and Δ ELREA EGFR-ATP complexes was carried out in order to determine the dynamic effect of the mutation on ATP binding. The simulations were performed in triplicate with different initial velocities and consistent observations were recorded, although simulation 2 of both ATP-bound wild-type and mutant EGFR showed a wider range of motions and larger estimated free energy of binding. Here we have described the results from simulation 1.

Active EGFR kinase: Dynamics of wild-type and Δ ELREA EGFR

In order to assess the overall dynamics of wild-type and Δ ELREA EGFRs during the simulations, the RMSF ($C\alpha$ atoms) was computed for each saved trajectory: it is evident that the loop regions fluctuate more than the secondary-structured regions in both EGFRs (Fig 4A). These loops include the N-terminal juxtamembrane segment, P-loop, A-loop and the C-terminal end of the proteins (note that due to the 5-residue deletion in Δ ELREA, the plots for wild-type and mutant EGFR (Fig 4A) are correspondingly offset from each other beginning from residue 746). A significant difference in fluctuation between wild-type and mutant EGFRs occurs at the β 3- α C loop—where the deletion mutation is located—and along the adjacent α C helix. These regions fluctuate more in wild-type EGFR in comparison to Δ ELREA. With the DFG motif there was no measurable change occurring as a result of the mutation, likely because only the side chain of F856 is in contact with the C-terminal end of the α C helix, which is itself positionally stabilized.

To better visualize the dynamic movements taking place in wild-type and Δ ELREA EGFRs, conformations sampled from the simulations over a given simulation-time interval were retrieved and superimposed on the median structure. The backbone traces are color coded according to the RMSD values over $C\alpha$ atoms for the conformational ensembles versus the median structure: blue indicates positionally stable regions, whereas white and red represent increased mobility. Both the wild-type (Fig 4B) and Δ ELREA (Fig 4C) EGFRs share similar stability profiles over a majority of each structure, with the exception of the β 3- α C loop and the α C helix. These regions appear to be less mobile in the mutant EGFR (average RMSD 1.91 Å) in comparison to wild-type EGFR (average RMSD 2.93 Å), a clear indication of the impact of the Δ ELREA deletion.

Active EGFR kinase: Δ ELREA constrains the α C helix

In the initial structural model of the active form Δ ELREA EGFR (Fig 5A, left), as a result of the deletion, the N-terminal portion of the α C helix (residues 753–755) deforms and relocates to the position formerly occupied by the β 3- α C loop. Similar features have been observed in X-

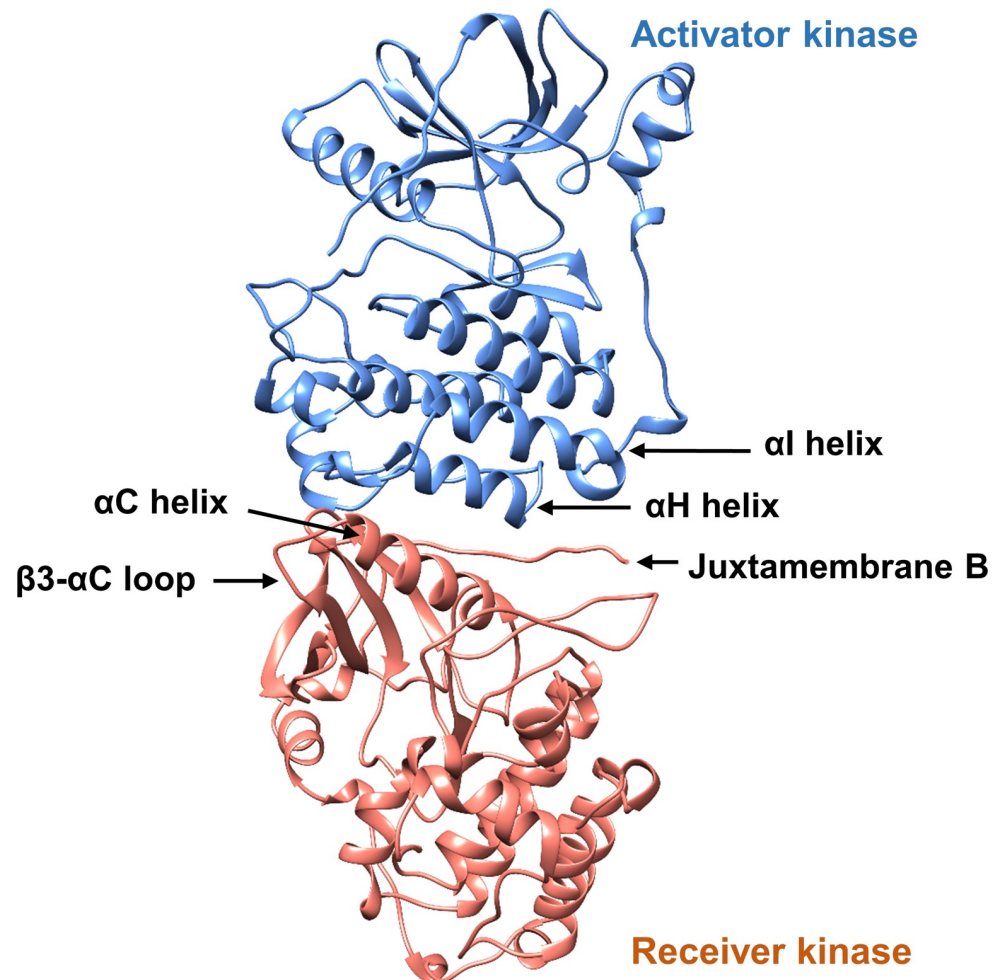


Fig 3. The EGFR asymmetric dimer. Interaction between the activator (blue) and receiver (orange) kinase domains in the EGFR asymmetric dimer; key structural elements at the dimer interface are highlighted, as well as the β 3- α C loop.

<https://doi.org/10.1371/journal.pone.0222814.g003>

ray structures of human B-RAF serine/threonine kinase with a β 3- α C loop deletion mutation, where this loop was highlighted as a critical feature—“a rheostat”—modulating kinase activity [55].

During the simulation of the active form Δ ELREA EGFR, the N-terminal part of the α C helix, nearest the site of the deletion, is uncoiled, as depicted by the average structure from the simulation of Δ ELREA EGFR (Fig 5A, right). The structural changes arising from Δ ELREA also places the α C helix in close proximity to the phosphate binding P-loop. This loop is involved in coordinating the nucleotide substrate, covering the ATP binding site where the α and β phosphates reside. During the simulation of wild-type active EGFR, Phe723 of the P-loop is on average 11.6 Å (95% confidence interval (CI) \pm 0.04) from Ile759 of the α C helix (measured between C α atoms). This distance is 1.5 Å shorter in the mutant, averaging 10.1 Å (95% CI \pm 0.04). The P-loop also packs closely with the uncoiled N-terminus of the α C helix, which in turn reduces the flexibility of the uncoiled helix.

The C α atoms of the α C helix (residues 753–768) fluctuate less in the mutant (average RMSF of 1.1 \pm 0.4 Å) than in wild-type EGFR (average RMSF of 1.5 \pm 0.57 Å) in comparison to their respective average structures (Fig 5B). This result is consistent with the shortened

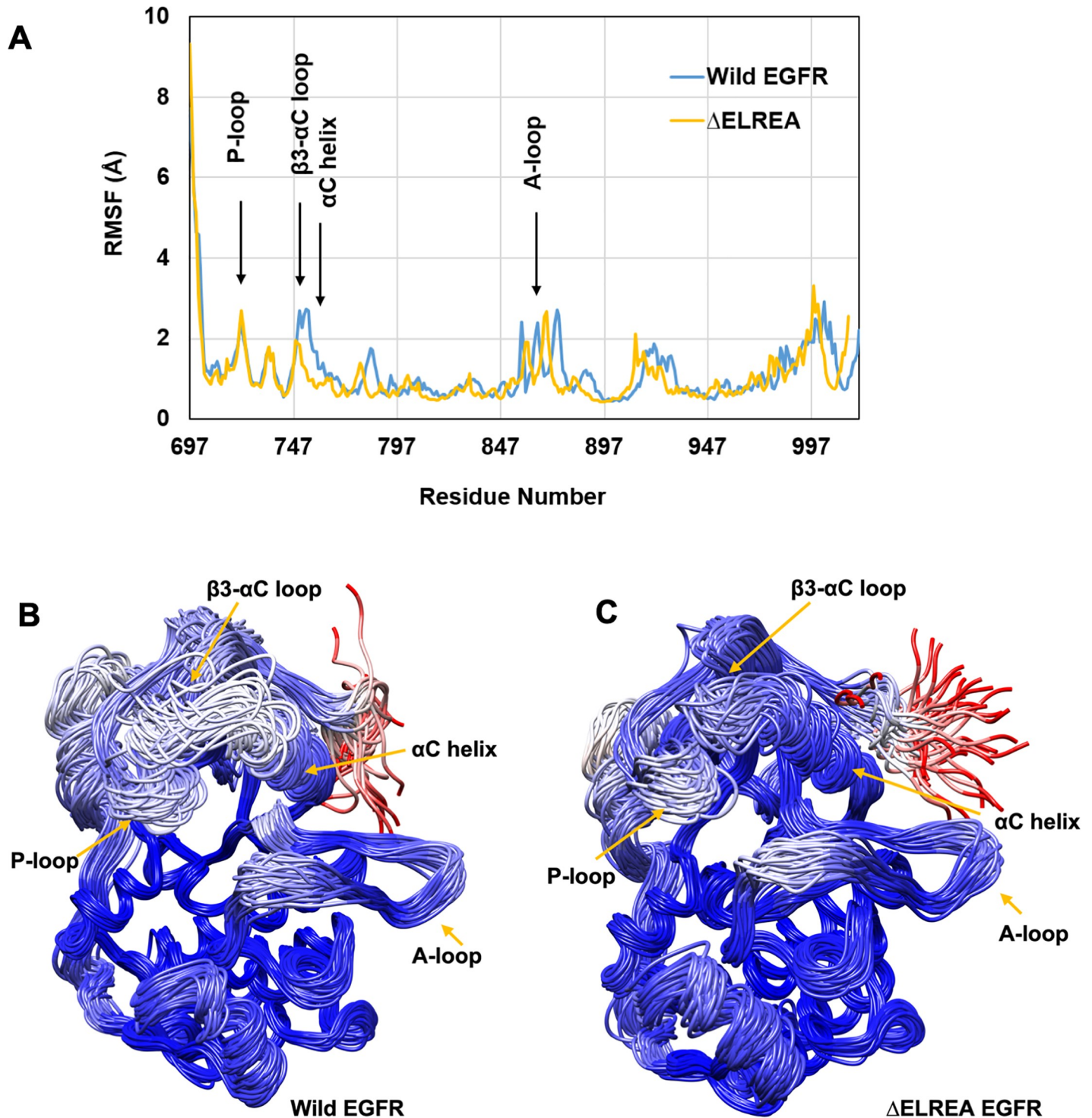


Fig 4. α -atom fluctuations and conformational ensemble of active wild-type and Δ ELREA EGFRs. (A) RMSF calculated over $C\alpha$ atoms for conformations sampled from MDS, showing large fluctuations for unstructured regions in both EGFRs; wild-type (Wild EGFR) and mutant (Δ ELREA) EGFR are offset by 5 residues starting from residue 746. Superimposed conformations of wild-type (B) and Δ ELREA (C). Chain traces are colored based on RMSDs for individual $C\alpha$ atoms within the ensemble according to the program Chimera: The gradient of coloring varies from blue with RMSD = 0.45 Å (minimum observed value), to white at 5.6 Å and to red = 11 Å (maximum observed value).

<https://doi.org/10.1371/journal.pone.0222814.g004>

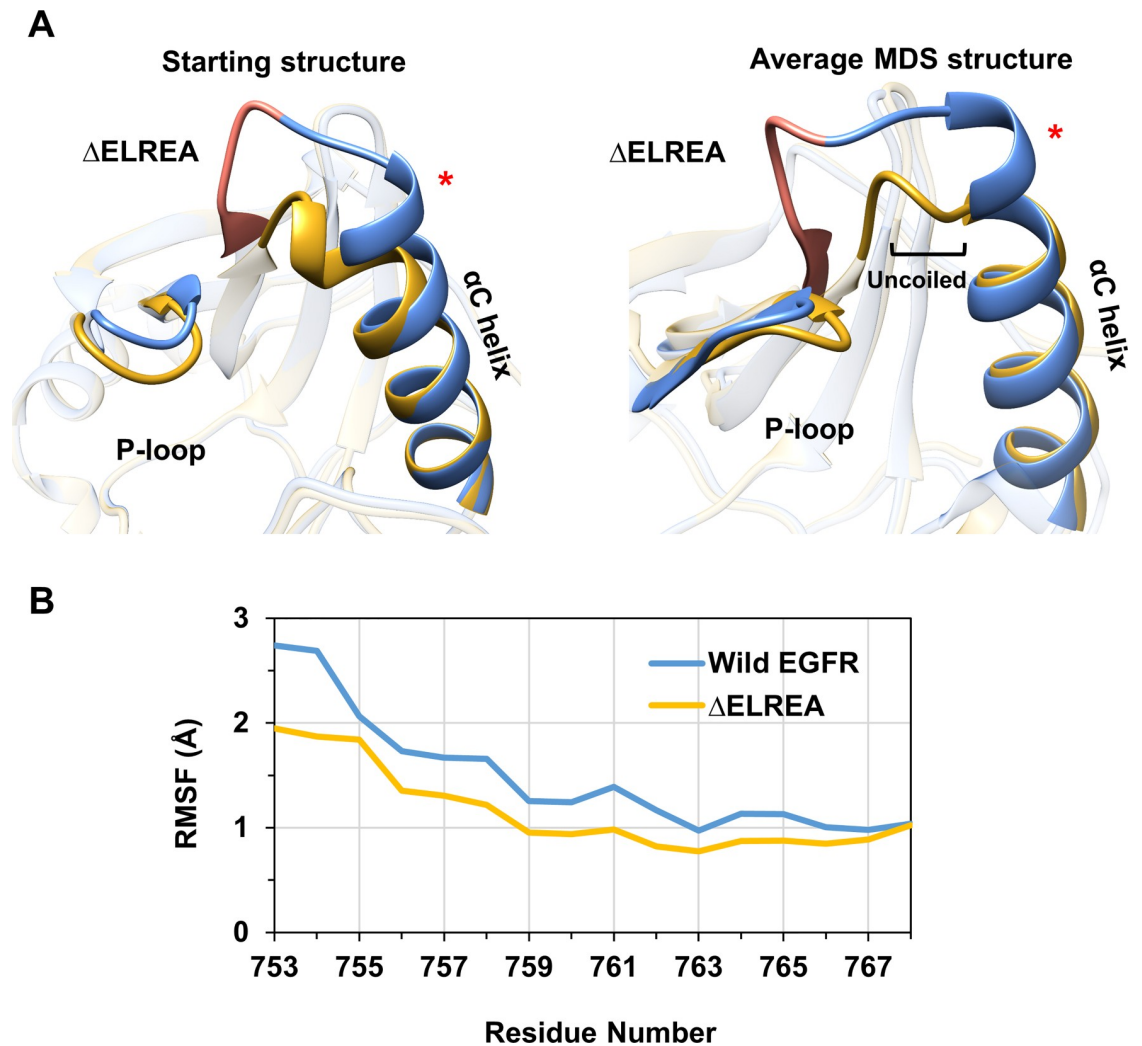


Fig 5. Wild-type vs Δ ELREA active EGFR structures and RMSF of the α C helix. (A) Superimposed active EGFR starting structure (left) and average MDS structure (right): wild-type (blue) and Δ ELREA (gold), highlighting the α C helix, P-loop and the deleted Δ ELREA section (orange). The N-terminal end of the α C helix is marked with a red asterisk in all relevant figures as a point of reference. (B) α C-atom RMSF over residues of the α C helix for the wild-type and mutant EGFR structures with respect to the average structure.

<https://doi.org/10.1371/journal.pone.0222814.g005>

linker loop of Δ ELREA EGFR between strand β 3 and the α C helix that constrains and limits the movement of the helix with respect to the strand. It is noteworthy that the β 3- α C loop is often missing in X-ray structures of the EGFR kinase domain, a testament to the flexibility of the wild-type loop. Deletion of a section of this loop would then help rigidify the loop and stabilize the succeeding α C helix in the “ α C-in” conformation, which would in turn shift the equilibrium in favor of the enzymatically active state EGFR and likely prolonging the lifetime of that activated state, too.

The principal motions revealed in the PCA calculations for the backbone atoms of the β 3- α C loop and α C helix (Fig 6) also indicate higher mobility over these structural units in wild-type EGFR relative to the Δ ELREA mutant. The top three PCs for β 3- α C loop and α C helix of wild-type and Δ ELREA EGFRs are represented as “porcupine” plots, where the arrows represent the direction and magnitude of motions over the backbone atoms. In Δ ELREA EGFR,

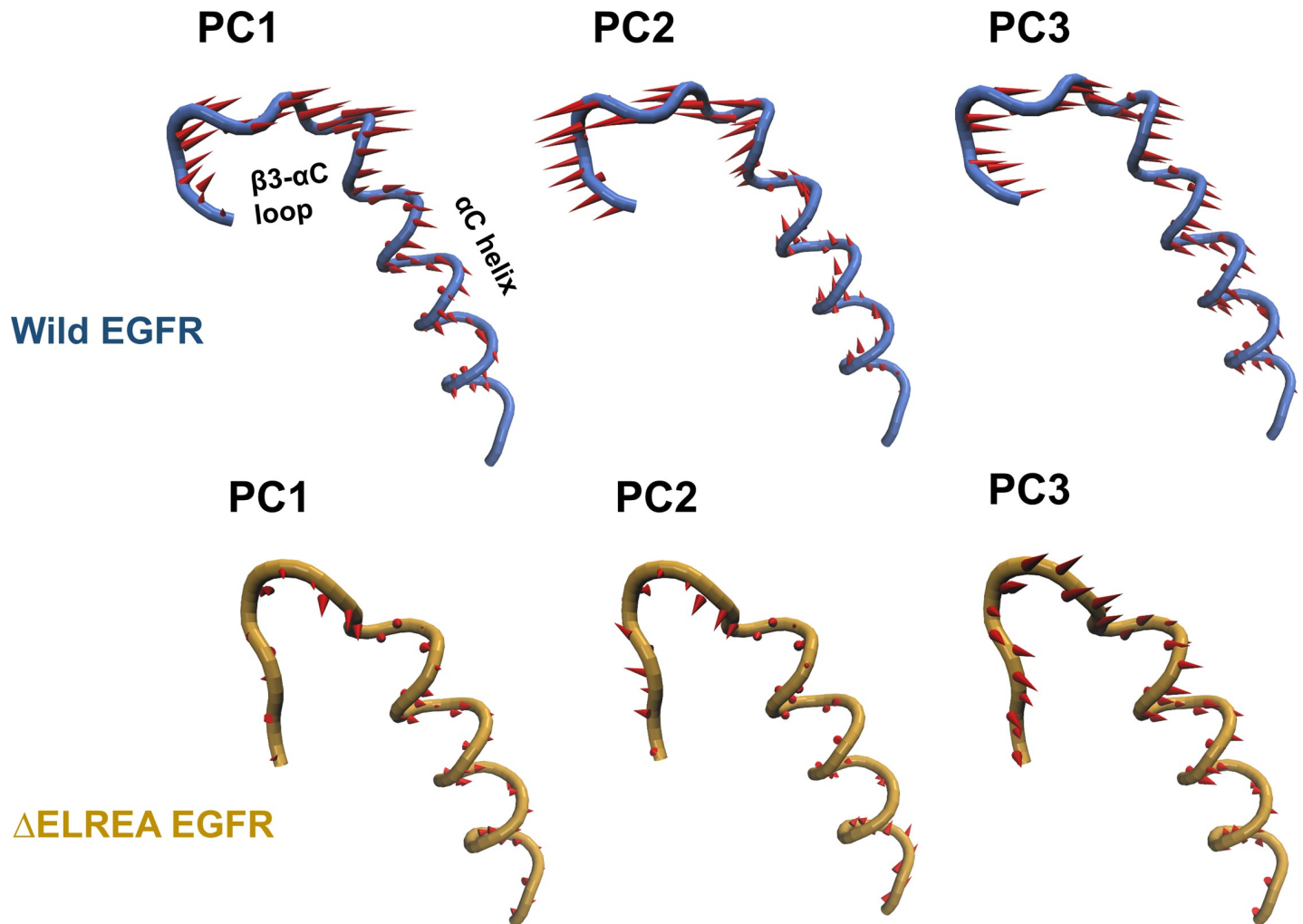


Fig 6. Principal motions of wild-type and Δ ELREA EGFR. The top three PCs for the backbone atoms of the α C helix and β 3- α C loop of wild-type (blue) and Δ ELREA (gold) EGFR are shown with a “porcupine” plot; red arrows represent the magnitude and direction of motion.

<https://doi.org/10.1371/journal.pone.0222814.g006>

both the α C helix and the remaining portion of the β 3- α C loop appear to be stable over all three PCs as indicated by the size of the arrows, in contrast to wild-type EGFR, exhibiting relatively large motions for both the α C helix and β 3- α C loop. Furthermore, the α C helix motions of wild-type EGFR are directed upwards and outwards, which may drive the conformational transition from the “ α C-in” to the “ α C-out” state.

Although the Δ ELREA deletion is not in direct contact with the asymmetric dimer interface (Fig 3), the deletion could also have an indirect effect on the dimer interaction through the stabilization exerted on the α C helix. The α C helix of the receiver kinase along with the juxta-membrane B segment is an integral part of the dimer interface, interacting with the α H and α I helices of the activator kinase. Therefore, it can be hypothesized that the deletion within the β 3- α C loop may also stabilize the interactions of the EGFR asymmetric dimer interface, and hence lead to a prolonged activation.

Active EGFR kinase: Δ ELREA and the E762··K745 salt bridge

A salt bridge conserved among kinases, forming between a glutamate from the α C helix and a lysine from the β 3 strand, is fundamental for tyrosine kinase activity. The ionic interaction helps to optimally orient and stabilize the lysine, which in turn interacts with the α - and β -phosphates of ATP, placing them properly for catalysis [14, 15] (Fig 7A). The E762··K745 salt bridge is present when the kinase is in the α C-in active conformation; conversely, the ionic interaction is broken and phosphate transfer is disrupted when the helix is oriented outwards as in the inactive conformation [16, 20]. To assess the dynamics of this interaction in the EGFR active state, we monitored the distance between the side-chain atoms C δ of Glu762 and N ζ of Lys745 in the wild-type and Δ ELREA trajectories (see S1A Fig).

With active Δ ELREA EGFR, the 4.2 ± 0.2 Å (95% CI ± 0.01) average distance between the side-chain atoms of the salt bridge is clearly more consistent and less variable than the 4.7 ± 1.0 Å (95% CI ± 0.04) observed in the simulation of wild-type active EGFR (Fig 7B). This reflects the mutual stabilization of both the α C helix and β 3 strand relative to each other due to the reduction of the intervening loop and consequential support for the formation and maintenance of the salt bridge that links both secondary structures to each other. The simulation of wild-type active EGFR results in longer E762 and K745 distances in multiple frames of the trajectory, implying frequent disruption of this interaction. The percentage occupancy for hydrogen bonds formed by the side-chain polar atoms of E762 and K745 also support this view (Fig 7C): the Glu762–Lys745 hydrogen bond is observed in more frames of the mutant trajectory than for wild-type EGFR. These findings suggest that the stability of the α C helix imparted by Δ ELREA mutation also enhances the stability of the E762··K745 salt bridge, which is key to EGFR kinase activity.

Active EGFR kinase: Δ ELREA and ATP binding

Protein kinases catalyze the phosphorylation of proteins fueled by Mg^{2+} -ATP. The cation coordinates with phosphate groups of ATP, helping to neutralize the negative charge, and optimally orienting the nucleotide for γ -phosphate transfer [56]. The adenosine ring of ATP binds in a hydrophobic pocket near the hinge region of EGFR, and the phosphate groups extend towards the N-terminal part of the A-loop (Fig 8A). During the simulation of the Mg^{2+} -ATP complexes of active Δ ELREA and wild-type EGFRs, a substantial difference is observed in the conformation of the α C helix, which moves away from the binding pocket in the case of wild-type EGFR (Fig 8A). In contrast, the α C helix of the Δ ELREA deletion mutant maintains its initial position near the active site, owing to the physical restraint introduced by the deletion. The orientation of the triphosphate moiety of ATP and the location of Mg^{2+} varies within all of the simulations of both mutant and wild-type EGFR (Fig 8A). Thus, these differences do not correlate with the Δ ELREA mutation, but instead follow the two observed orientations of the D855 side-chain from the DFG motif that coordinates Mg^{2+} .

The most stable interactions in both EGFR-ATP complexes occur between backbone atoms of Gln791 and Met793 and the adenosine ring of ATP (Fig 8A). These interactions exist in more than 90% of the snapshots obtained from the simulation. A key interaction which is differentially observed in the two EGFR-ATP forms takes place between Lys745 of the conserved salt bridge and phosphate groups of ATP. This interaction is more preserved in Δ ELREA EGFR, likely a result of a more stable α C helix (average C α atom RMSF of 0.5 Å), which in turn stabilizes the salt bridge between Glu762 and Lys745, therefore optimally positioning Lys745 to make the interaction with ATP. In contrast, with wild-type EGFR the full-length β 3- α C loop adds local flexibility and imposes fewer restraints, resulting in a less conformationally stable α C helix (average C α atom RMSF of 1.4 Å) and salt bridge between Glu762 and Lys745.

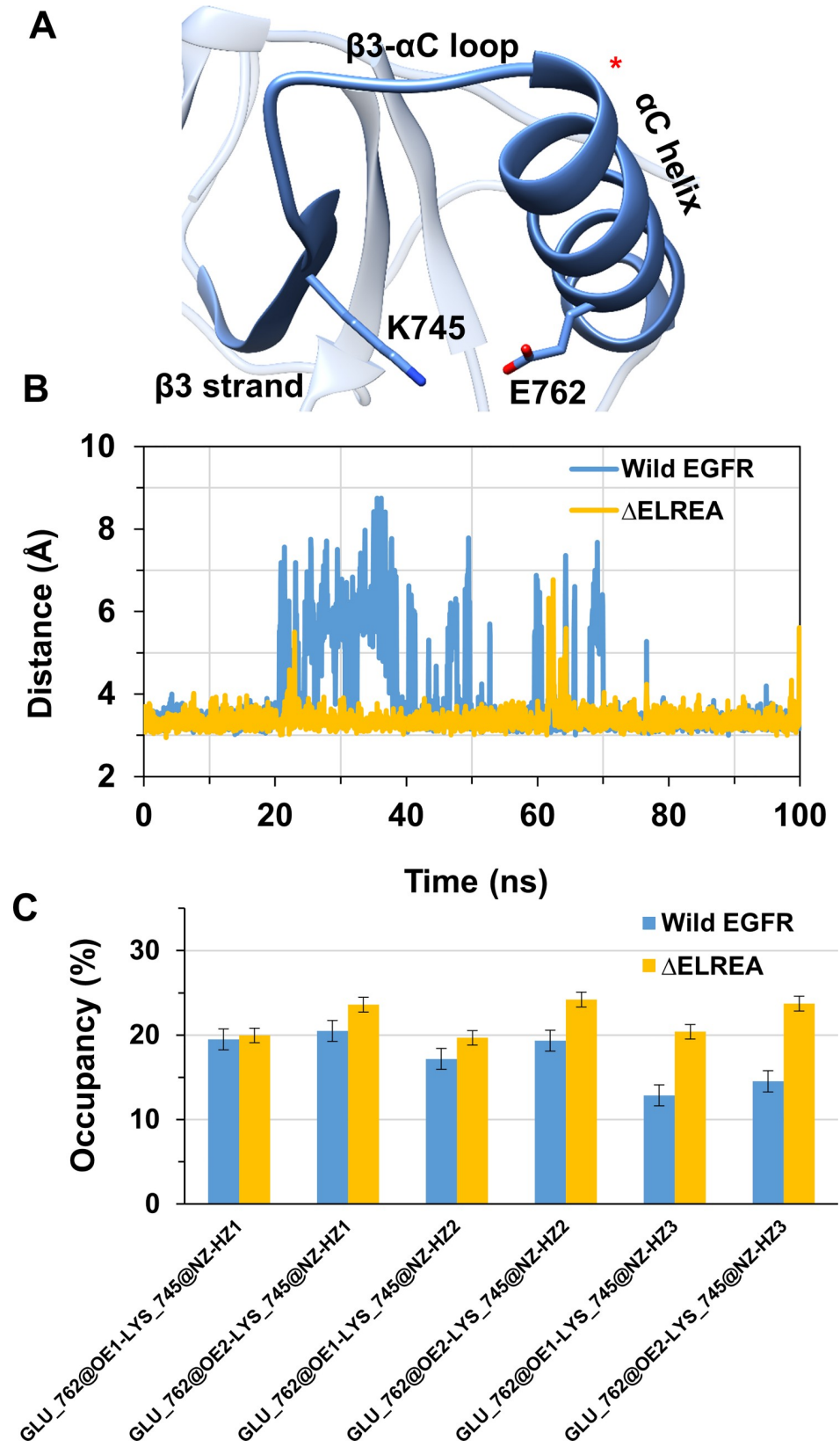


Fig 7. EGFR E762 · K745 salt bridge behaviour during the MDS. (A) Key salt bridge formed between K745 and E762 in the active EGFR kinase (shown in sticks). (B) Distance between K745 and E762 during the 100 ns simulation. Wild-type active EGFR (blue) displays longer distances and more fluctuations in the distances between C δ of Glu762 and N ζ of Lys745 of the ion pair as compared to Δ ELREA EGFR (gold). The C δ · · N ζ distances of the salt bridge are provided in S1A Fig. (C) Percentage occupancy of hydrogen bonds between terminal hydrogen atoms HZ1-3 of K745 and carbonyl oxygen atoms Oe1 and Oe2 of E762 in the wild-type and Δ ELREA EGFRs.

<https://doi.org/10.1371/journal.pone.0222814.g007>

As a result, Lys745 is not as strictly available as in the mutant EGFR to interact with the phosphate groups of the nucleotide. Δ ELREA EGFR also benefits from a well-maintained hydrogen bond between ATP and Arg841, which is less frequently observed in wild-type EGFR (see S2 Table for details). The number of hydrogen bonds within a distance of 3.5 Å formed between ATP and the two EGFRs during the simulation can provide a general view of the magnitude of interactions taking place: an average of 4.0 (95% CI \pm 0.03) hydrogen bonds were observed for the mutant and 3.2 (95% CI \pm 0.04) for wild-type EGFR (Fig 8B).

To estimate the relative binding affinity of ATP towards the wild-type and Δ ELREA EGFR kinases in the active conformation, the free energy of binding of the nucleotide was computed using the MM-GBSA method. The result (Fig 8C) shows that Δ ELREA EGFR exhibits a lower binding free energy, average ΔG_{bind} of -57 kcal/mol (95% CI \pm 0.43), as compared to the wild-type, which has an average ΔG_{bind} of -48 kcal/mol (95% CI \pm 0.33). This difference in free energy of binding is consistent with the larger number of favorable hydrogen bonds observed in the Δ ELREA EGFR-ATP complex in comparison to the wild-type. In summary, these findings support the notion that active Δ ELREA EGFR would have a higher binding affinity for ATP than the active wild-type EGFR, which in comparison to the active wild-type EGFR may be a contributing factor to the experimentally observed increased activity of the deletion mutant [53, 54].

Effect of Δ ELREA on the inactive EGFR kinase

The dynamic switching between inactive and active conformations of kinases is a fundamental process contributing to the regulation of their activity [15]. In ErbBs, it is not only the activation triggered by dimerization on binding growth factor, but also changes from the active to inactive state that can dampen down ErbB phosphorylation and signaling. Furthermore, the relative stabilization of features of the inactive versus active conformations can shift the equilibrium towards the more stable state, increasing its lifetime. Hence, it is of great importance to assess the possible structural effects of Δ ELREA not only on the active, but also on the inactive EGFR kinase.

In contrast to active EGFR structures, the inactive structures present in the PDB exhibit an α C helix with an uncoiled N-terminus. Hence, both the wild-type and mutant inactive EGFR structures used for the MDS are uncoiled at residues 753–755, which are now part of the β 3- α C loop. The only difference between the wild-type and Δ ELREA inactive EGFR initial models is therefore the length of the β 3- α C loop, which is shorter for the latter as a result of the deletion. To assess the effect of the mutation on the inactive structures during the simulation, the initial unminimized inactive structures and sampled conformations from both the wild-type EGFR and Δ ELREA inactive EGFR were superimposed on their respective median structures (solid colors, Fig 9A and 9B). With Δ ELREA a striking effect of the mutation is observed: the α C helix moves “inwards” towards the ATP cleft, a movement required to attain the active state conformation. In contrast, wild-type EGFR largely maintains the initial position of the α C helix. Moreover, the conformational ensemble (transparent chain-traces) shows a small “outward” motion for the wild-type α C helix (also for the A-loop helix) as compared to the initial unminimized structure. The α C helix of the mutant also becomes bent during the simulation, which is not observed for wild-type inactive EGFR where the central axis of the helix remains

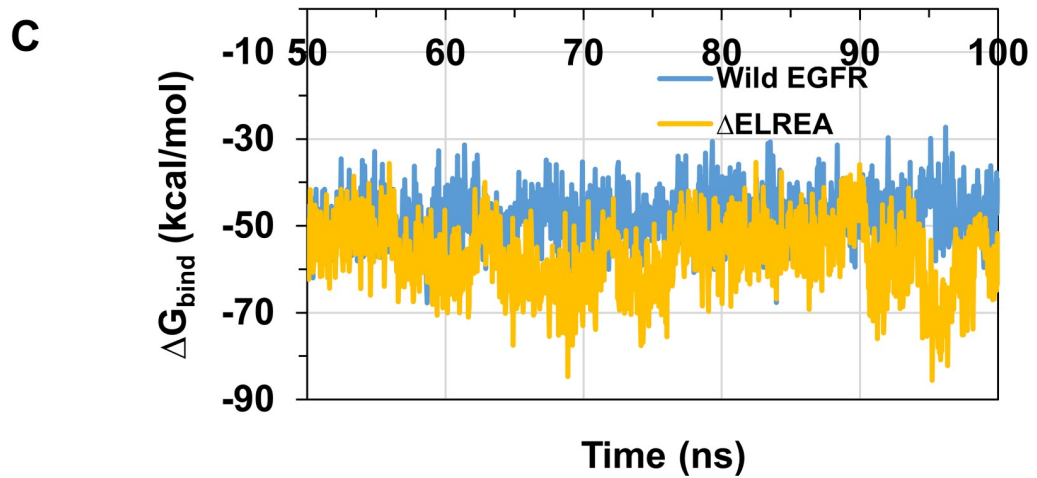
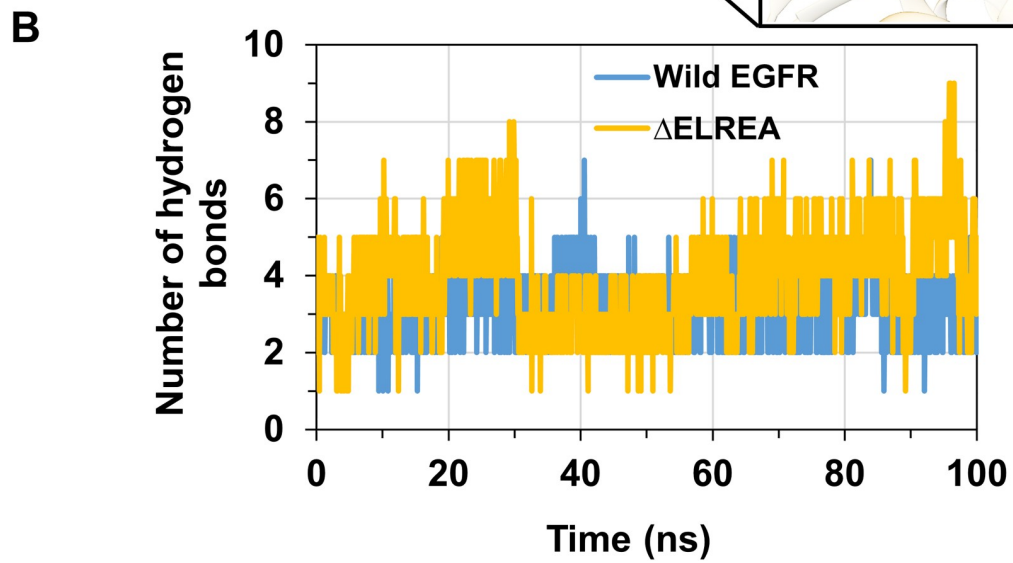
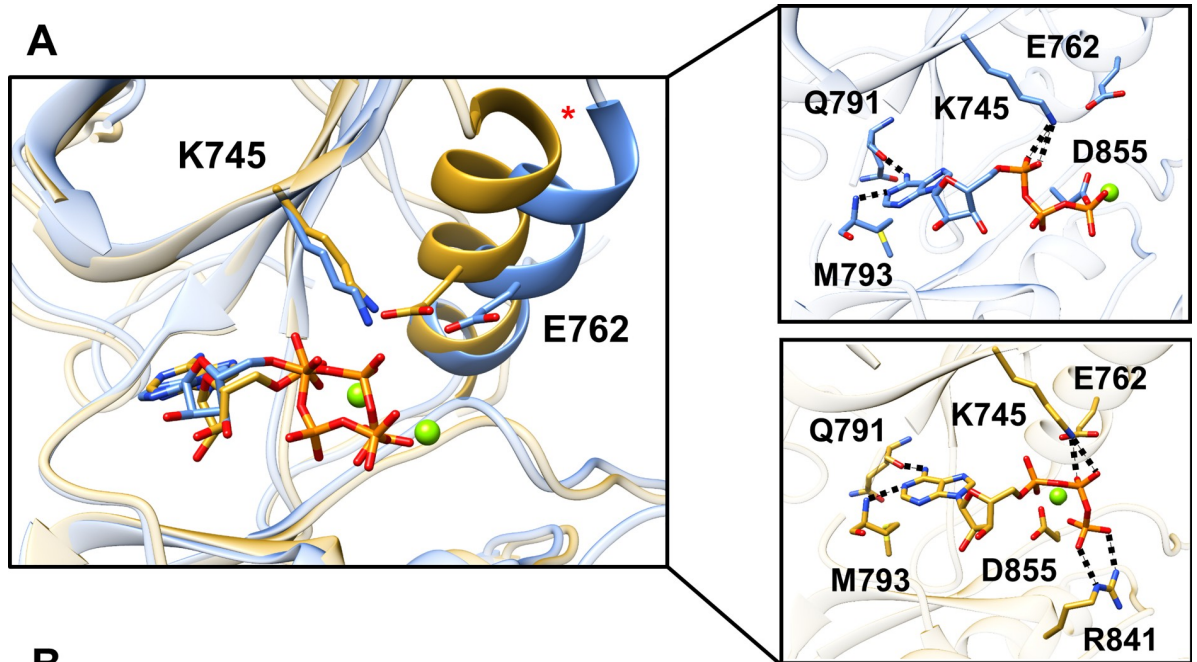


Fig 8. ATP binding to wild-type and Δ ELREA EGFR. Average structures of the (A) wild-type (blue) and Δ ELREA (gold) EGFR-ATP complexes from the MDS. The individual α C helices fluctuate more in wild-type EGFR than Δ ELREA EGFR (S2 Fig). The orientation of the triphosphate group of ATP and location of Mg^{2+} alternate relative to the side-chain location of D855 in the simulations. ATP is shown in the binding pocket with a magnesium ion (green) along with key residues interacting with Mg^{2+} -ATP or involved in catalysis; and hydrogen bond interactions observed frequently (dotted line). (B) The number of hydrogen bonds formed between ATP and the two EGFRs during the 100 ns simulation. (C) Binding free energy of ATP to wild-type and Δ ELREA EGFR during the 100 ns simulation. The number of hydrogen bonds observed between ATP and both wild-type and Δ ELREA EGFR, as well as the recorded binding free energy, is shown in S1B and S1C Fig.

<https://doi.org/10.1371/journal.pone.0222814.g008>

linear. The bent conformation in Δ ELREA results from the mutation-induced inward movement of the α C helix, which is then blocked by the small hydrophobic A-loop helix that packs against the α C helix and hinders a full inward movement (Fig 9B). Additionally, due to the partial inward movement of the Δ ELREA α C helix, the A-loop helix is “pressed down” from its initial position, making the distance between the two helices wider as compared to the wild-type inactive EGFR that maintains the initial positions of the two helices (Fig 9A). This notion is reflected in the distance between the C α atoms of Ile759 of the α C helix and Leu862 of the A-loop helix during MDS, averaging 5.8 Å (95% CI \pm 0.01) for wild-type and 7.4 Å (95% CI \pm 0.02) for mutant EGFR. Furthermore, compared to the wild-type inactive EGFR, the mutant displays a “loose” hydrophobic network between the two helices owing to the deletion of Leu747 of the β 3- α C loop (Δ ELREA), which normally contributes to the hydrophobic interaction in the wild-type EGFR (Fig 9C). Consequently, disruption of the obstructing hydrophobic cluster would be feasible in the mutant inactive kinase, which would predispose it to transition towards the active conformation.

It is noteworthy that observing large conformational changes in macromolecules requires longer simulations, in the range of micro and milliseconds that are computationally demanding. The same length of simulation time is needed to observe the transition among the conformational states of the EGFR tyrosine kinase [58]. Given our simulations were in nanosecond timescale, we do not expect to observe Δ ELREA driven conformational changes from the inactive to the active EGFR state. Nonetheless, from the 100 ns simulation it can be stipulated that Δ ELREA mutation favors an inward movement of the inactive EGFR α C helix and the mutation interferes with the obstructing hydrophobic cluster, features that support a shift from the inactive conformation towards that seen in active EGFR.

Discussion

EGFR is a large, complex receptor protein involved in various signaling pathways that regulate cell proliferation, migration and apoptosis [1]. The activity of EGFR and its family members are normally highly regulated and complex, since it is apparent that many regions of the structure can influence EGFR signaling and its active lifetime. Indeed, mutations in the extracellular, transmembrane and intracellular domains have been shown to enhance tyrosine kinase phosphorylation as well as cause changes in signaling by altering the active-inactive dynamics of the protein [2, 17, 26, 59]; over-expression, unregulated retroviral tyrosine kinase domains, and somatic mutations are all associated with cancers [23, 24].

In this study, we examined the Δ ELREA deletion mutation in human EGFR that is commonly observed in NSCLC patients. Indeed, Δ ELREA, located at the β 3- α C loop of the kinase domain, is reported to result in increased kinase activity [53, 54]; and of variable-length deletions introduced into the β 3- α C loop of EGFR, a five-residue deletion led to maximal activity [55]. Using MDS, we investigated the implications of the mutation on the active and inactive EGFR kinase structures, and the effect of Δ ELREA on ATP binding. In the active kinase structure, the mutation was shown to stabilize the α C helix in the “ α C-in” conformation by

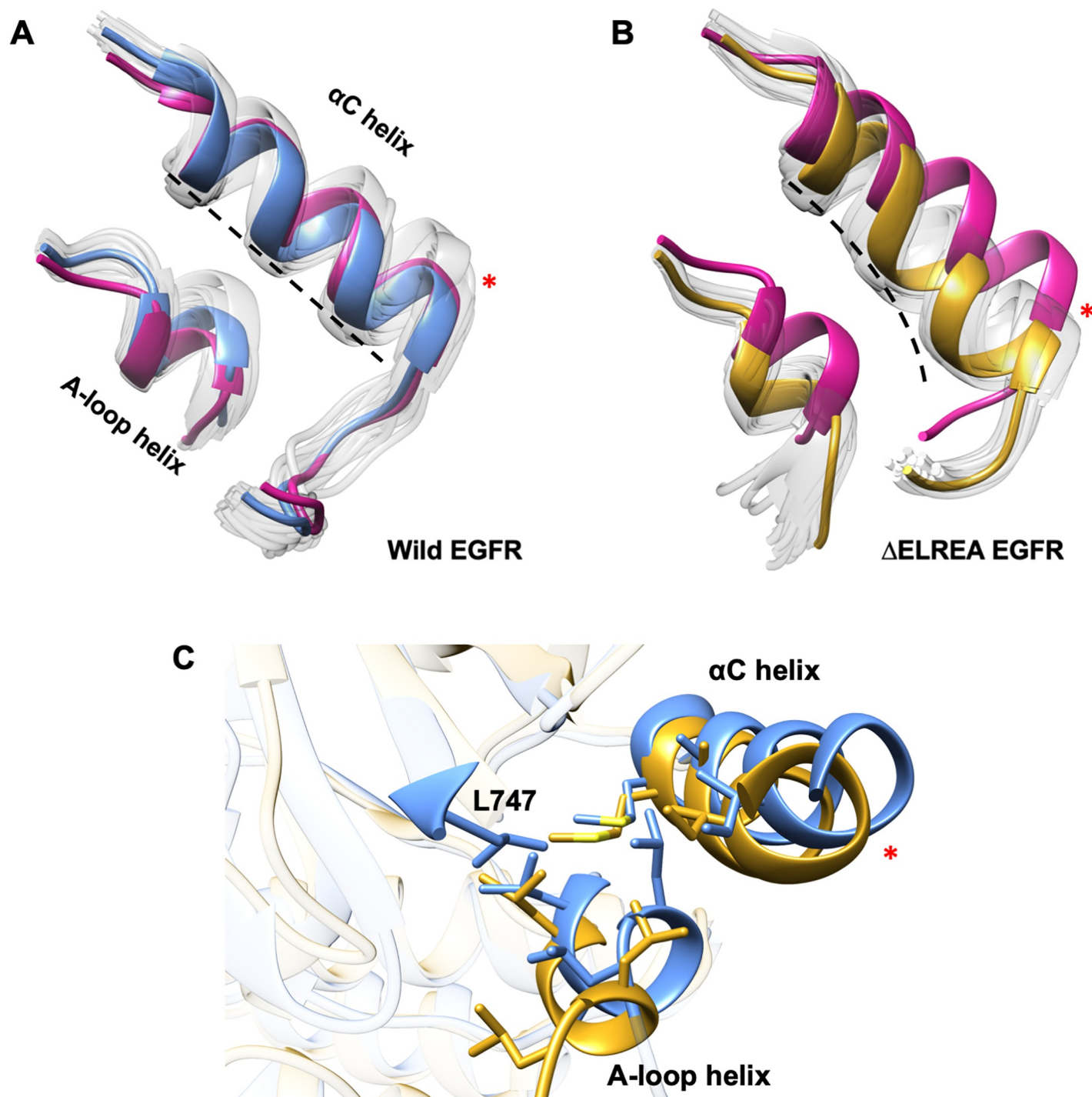


Fig 9. Inactive form of wild-type versus Δ ELREA EGFR during MDS. Superimposed conformations of the A-loop and α C helices of wild-type (A) and Δ ELREA (B) inactive EGFRs sampled during MDS. The median structure in the wild-type and mutant EGFRs is highlighted in blue and gold, respectively, initial unminimized structures in pink, conformational variations among sampled structures in transparent white (view inspired by [57]). Wild-type EGFR maintains the initial positions of the A-loop and α C helices, which are displaced in the mutant EGFR. The α C helix of the wild-type EGFR remains linear whereas the mutant is curved (dotted line). (C) Amino acids contributing to the hydrophobic network between the α C helix and the A-loop helix are shown as sticks. L747 from the β 3- α C loop, which is deleted in the Δ ELREA mutant, is labeled.

<https://doi.org/10.1371/journal.pone.0222814.g009>

EGFR wild-type
EGFR Δ ELREA

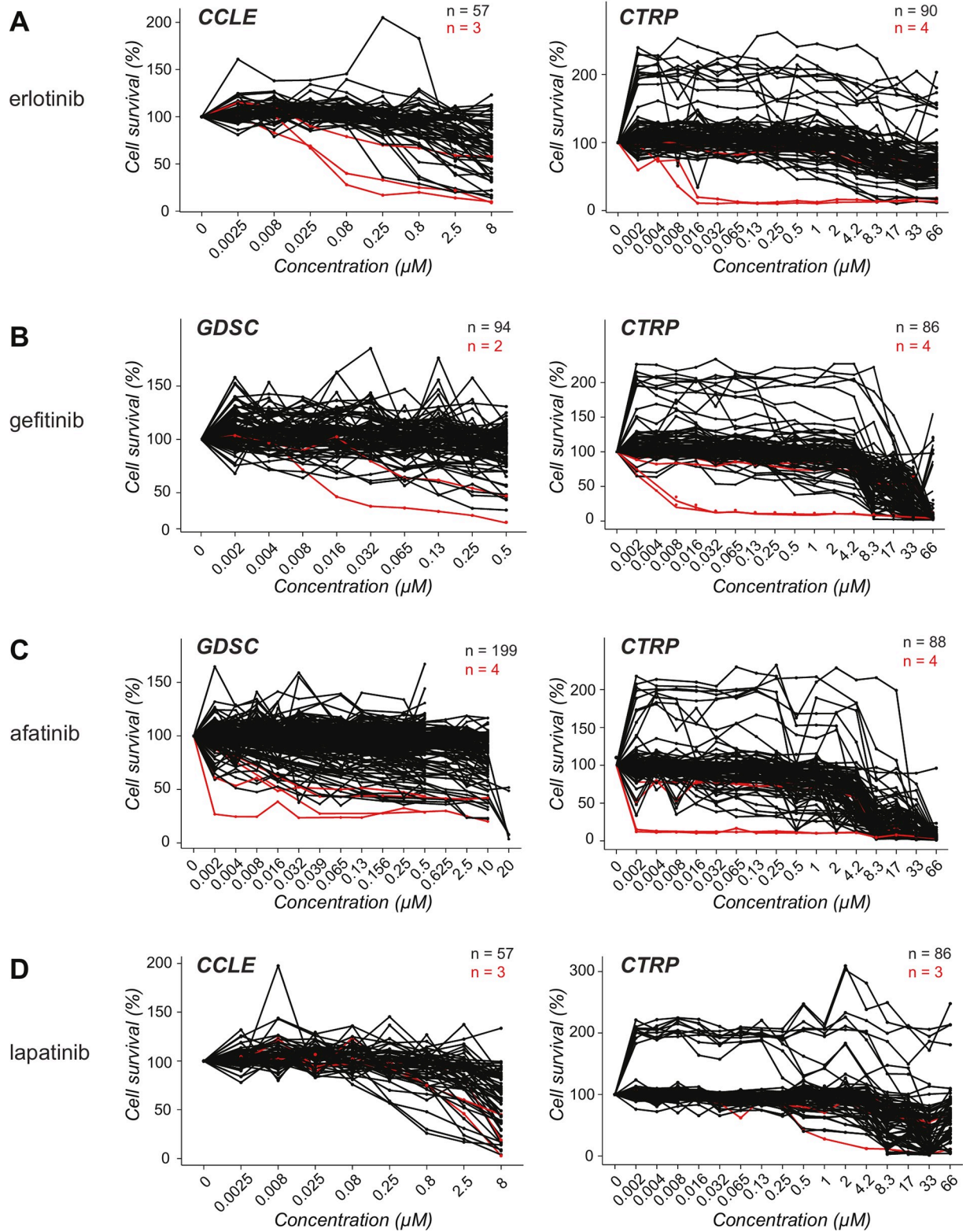


Fig 10. Response of EGFR Δ ELREA mutant lung cancer cells to tyrosine kinase inhibitors. Sensitivity (cell survival versus inhibitor concentration) of lung cancer cell lines harboring EGFR Δ ELREA mutant (red curves) versus wild-type EGFR (black curves) to (A) erlotinib, (B) gefitinib, (C) afatinib, and (D) lapatinib. The data were collected from the publicly available databases CCLE, CTRP and GDSC, as indicated.

<https://doi.org/10.1371/journal.pone.0222814.g010>

constraining its movement due to deletion of a section of the flexible β 3- α C loop. This would help maintain the salt bridge between Glu762 and Lys745 that is crucial for kinase activity. The deletion also results in a higher affinity interaction between ATP and EGFR, which is supported by a lower binding free energy value for the mutant-ATP complex in comparison to wild-type EGFR. Moreover, in the inactive kinase structure, Δ ELREA results in an inward movement of the α C helix that also disrupts a hydrophobic barrier between the active site and the α C helix, which would advocate a shift from an inactive conformation to the active conformation of EGFR. Each of these observations from MDS is consistent with a kinase domain with increased stability and lifetime in the active catalytic state.

The MDS results for the Δ ELREA mutation are also consistent with the effects of compounds (Fig 10A–10C) that inhibit EGFR mutant cell lines [49–51]: inhibitors recognizing the active conformation are more effective against the Δ ELREA mutant in lung cancer cell lines as compared to cell lines expressing wild-type EGFR. For example, the mean AUC values from CTRP data for wild-type and Δ ELREA mutant were 13.45 and 8.20 with erlotinib ($P < 0.0001$ –derived using unpaired two-sample t-test), 12.43 and 6.98 with gefitinib ($P < 0.0001$), and 10.39 and 5.29 with afatinib ($P < 0.0001$), respectively. In contrast to these three EGFR inhibitors, with lapatinib (Fig 8D), which selectively recognizes and binds to the inactive form of the receptor [60], the mean AUC for wild-type EGFR (mean AUC 12.86) is not significantly different ($P = 0.0844$) from that of the Δ ELREA mutant (mean AUC 11.28). Thus, the results of our study are consistent with the observed effects of conformation-specific EGFR ligands in cancer cell lines and further supported by simulations focused on inhibitor complexes with the active conformation of Δ ELREA EGFR, where the restricted α C helix movement resulted in a more compact binding site impacting the interactions and affinity for inhibitors [61].

In conclusion, this study suggests that the effects exerted by Δ ELREA are two-fold: the deletion mutation in EGFR increases kinase activity (1) by *stabilizing* that active catalytic conformation of the tyrosine kinase, and (2) by promoting a *conformational shift* towards the active catalytic state. Such “hot spot” somatic mutations, like Δ ELREA, offer possibilities to select ligands that may block conformational and stabilization effects and hence provide possible therapeutic tools aimed at controlling ErbB phosphorylation and signaling. Indeed, known EGFR inhibitors that preferentially recognize the active kinase conformation, as promoted by the Δ ELREA deletion, exert a more dramatic effect—reducing cell survival in cancer cell models—in comparison to conformationally malleable wild-type EGFR.

Supporting information

S1 Table. List of 3D structures used to build composite models for the MDS study.
(PDF)

S2 Table. Hydrogen bond interactions between wild-type/ Δ ELREA EGFR and ATP during MDS.
(PDF)

S1 Fig. Histogram plots of salt bridge distance, number of hydrogen bonds between EGFR and ATP, and free energy of binding of ATP from Figs 7B, 8B and 8C, respectively.
(TIF)

S2 Fig. Superposed coordinates of snapshots from each of the three ATP-bound replicate simulations of wild-type and mutant EGFR.

(TIFF)

S1 File. PDB coordinates of active form apo wild-type EGFR used for the simulation.

(PDB)

S2 File. PDB coordinates of active form apo Δ ELREA EGFR used for the simulation.

(PDB)

S3 File. PDB coordinates of active form ATP bound wild-type EGFR used for the simulation.

(PDB)

S4 File. PDB coordinates of active form ATP bound Δ ELREA EGFR used for the simulation.

(PDB)

S5 File. PDB coordinates of inactive form apo wild-type EGFR used for the simulation.

(PDB)

S6 File. PDB coordinates of inactive form apo Δ ELREA EGFR used for the simulation.

(PDB)

S7 File. R code used to process the drug response data.

(R)

Acknowledgments

We thank the CSC IT Center for Science for the computational resources and Dr. Outi Salon-Ahen for facilitating it. We also thank Dr. Jukka Lehtonen at Åbo Akademi University for scientific IT support within the Biocenter Finland bioinformatics network; and Biocenter Finland structural biology infrastructure network, and Dr. Parthiban Marimuthu for feedback on the simulations.

Author Contributions

Conceptualization: Mahlet Z. Tamirat, Mark S. Johnson.

Data curation: Mahlet Z. Tamirat, Marika Koivu.

Formal analysis: Mahlet Z. Tamirat, Marika Koivu.

Funding acquisition: Mark S. Johnson.

Investigation: Mahlet Z. Tamirat.

Methodology: Mahlet Z. Tamirat, Marika Koivu.

Resources: Mark S. Johnson.

Supervision: Mark S. Johnson.

Validation: Klaus Elenius, Mark S. Johnson.

Visualization: Mahlet Z. Tamirat.

Writing – original draft: Mahlet Z. Tamirat.

Writing – review & editing: Mahlet Z. Tamirat, Klaus Elenius, Mark S. Johnson.

References

1. Wieduwilt MJ, Moasser MM. The epidermal growth factor receptor family: biology driving targeted therapeutics. *Cell Mol Life Sci*. 2008; 65:1566–84. <https://doi.org/10.1007/s00018-008-7440-8> PMID: 18259690
2. Lemmon MA, Schlessinger J, Ferguson KM. The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harb Perspect Biol*. 2014; 6:a020768. <https://doi.org/10.1101/cshperspect.a020768> PMID: 24691965
3. Seshacharyulu P, Ponnusamy MP, Haridas D, Jain M, Ganti AK, Batra SK. Targeting the EGFR signaling pathway in cancer therapy. *Expert Opin Ther Targets*. 2012; 16(1):15–31. <https://doi.org/10.1517/14728222.2011.648617> PMID: 22239438
4. Downward J, Yarden Y, Mayes E, Scrace G, Totty N, Stockwell P, et al. Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. *Nature* 1984; 307(5951):521–7. <https://doi.org/10.1038/307521a0> PMID: 6320011
5. Vogt PK. Retroviral oncogenes: a historical primer. *Nat Rev Cancer*. 2012; 12(9):639–48. <https://doi.org/10.1038/nrc3320> PMID: 22898541
6. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, 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(21):2129–39. <https://doi.org/10.1056/NEJMoa040938> PMID: 15118073
7. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy. *Science*. 2004; 304(5676):1497–500. <https://doi.org/10.1126/science.1099314> PMID: 15118125
8. Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, et al. Efficacy and Safety of Trastuzumab as a Single Agent in First-Line Treatment of *HER2*-Overexpressing Metastatic Breast Cancer. *J Clin Oncol*. 2002; 20(3):719–26. <https://doi.org/10.1200/JCO.2002.20.3.719> PMID: 11821453
9. Tvorogov D, Sundvall M, Kurppa K, Hollmén M, Repo S, Johnson MS, et al. Somatic mutations of ErbB4: selective loss-of-function phenotype affecting signal transduction pathways in cancer. *J Biol Chem*. 2009; 284(9):5582–91. <https://doi.org/10.1074/jbc.M805438200> PMID: 19098003
10. Kurppa KJ, Denessiouk K, Johnson MS, Elenius K. Activating ERBB4 mutations in non-small cell lung cancer. *Oncogene*. 2016; 35(10):1283–91. <https://doi.org/10.1038/onc.2015.185> PMID: 26050618
11. Fuller SJ, Sivarajah K, Sugden PH. ErbB receptors, their ligands, and the consequences of their activation and inhibition in the myocardium. *J Mol Cell Cardiol*. 2008; 44(5):831–54.
12. Leahy DJ. Structure and Function of the Epidermal Growth Factor (EGF/ErbB) Family of Receptors. *Adv Protein Chem*. 2004; 68:1–27. [https://doi.org/10.1016/S0065-3233\(04\)68001-6](https://doi.org/10.1016/S0065-3233(04)68001-6) PMID: 15500857
13. Roskoski R. ErbB/HER protein-tyrosine kinases: Structures and small molecule inhibitors. *Pharmacol Res*. 2014; 87:42–59. <https://doi.org/10.1016/j.phrs.2014.06.001> PMID: 24928736
14. Hubbard SR, Till JH. Protein Tyrosine Kinase Structure and Function. *Annu Rev Biochem*. 2000; 69(1):373–98.
15. Huse M, Kuriyan J. The Conformational Plasticity of Protein Kinases. *Cell*. 2002; 109(3):275–82. [https://doi.org/10.1016/S0092-8674\(02\)00741-9](https://doi.org/10.1016/S0092-8674(02)00741-9) PMID: 12015977
16. Jura N, Zhang X, Endres NF, Seeliger MA, Schindler T, Kuriyan J. Catalytic Control in the EGF Receptor and Its Connection to General Kinase Regulatory Mechanisms. *Mol Cell*. 2011; 42:9–22. <https://doi.org/10.1016/j.molcel.2011.03.004> PMID: 21474065
17. Kumar A, Petri ET, Halmos B, Boggon TJ. Structure and clinical relevance of the epidermal growth factor receptor in human cancer. *J Clin Oncol*. 2008; 26(10):1742–51. <https://doi.org/10.1200/JCO.2007.12.1178> PMID: 18375904
18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28(1):235–42. <https://doi.org/10.1093/nar/28.1.235> PMID: 10592235
19. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J. An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell*. 2006; 125(6):1137–49. <https://doi.org/10.1016/j.cell.2006.05.013> PMID: 16777603
20. Hasenahuer MA, Barletta GP, Fernandez-Alberti S, Parisi G, Fornasari MS. Pockets as structural descriptors of EGFR kinase conformations. *PLoS One*. 2017; 12(12):e0189147. <https://doi.org/10.1371/journal.pone.0189147> PMID: 29228029
21. Zhang X, Chang A. Somatic mutations of the epidermal growth factor receptor and non-small-cell lung cancer. *J Med Genet*. 2007; 44(3):166–72. <https://doi.org/10.1136/jmg.2006.046102> PMID: 17158592

22. Gajiwala KS, Feng J, Ferre R, Ryan K, Brodsky O, Weinrich S, et al. Insights into the Aberrant Activity of Mutant EGFR Kinase Domain and Drug Recognition. *Structure*. 2013; 21(2):209–19. <https://doi.org/10.1016/j.str.2012.11.014> PMID: 23273428
23. Irmer D, Funk JO, Blaukat A. EGFR kinase domain mutations—functional impact and relevance for lung cancer therapy. *Oncogene*. 2007; 26(39):5693–701. <https://doi.org/10.1038/sj.onc.1210383> PMID: 17353898
24. Oliveira S, van Bergen en Henegouwen PM, Storm G, Schiffelers RM. Molecular biology of epidermal growth factor receptor inhibition for cancer therapy. *Expert Opin Biol Ther*. 2006; 6(6):605–17. <https://doi.org/10.1517/14712598.6.6.605> PMID: 16706607
25. Siegelin MD, Borczuk AC. Epidermal growth factor receptor mutations in lung adenocarcinoma. *Lab Invest*. 2014; 94(2):129–37. <https://doi.org/10.1038/labinvest.2013.147> PMID: 24378644
26. Gazdar AF. 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(Suppl 1):S24–31.
27. Sgambato A, Casaluce F, Maione P, Rossi A, Rossi E, Napolitano A, et al. The Role of EGFR Tyrosine Kinase Inhibitors in the First-Line Treatment of Advanced Non Small Cell Lung Cancer Patients Harboring EGFR Mutation. *Curr Med Chem*. 2012; 19(20):3337–52. <https://doi.org/10.2174/092986712801215973> PMID: 22664249
28. Sullivan I, Planchard D. Next-Generation EGFR Tyrosine Kinase Inhibitors for Treating EGFR-Mutant Lung Cancer beyond First Line. *Front Med*. 2016; 3:76.
29. Yun C-H, Boggon TJ, Li Y, Woo MS, Greulich H, Meyerson M, et al. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell*. 2007; 11(3):217–27. <https://doi.org/10.1016/j.ccr.2006.12.017> PMID: 17349580
30. Yun C-H, Mengwasser KE, Toms A V, Woo MS, Greulich H, Wong K-K, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci*. 2008; 105(6):2070–5. <https://doi.org/10.1073/pnas.0709662105> PMID: 18227510
31. Sogabe S, Kawakita Y, Igaki S, Iwata H, Miki H, Cary DR, et al. Structure-Based Approach for the Discovery of Pyrrolo[3,2-*d*]pyrimidine-Based EGFR T790M/L858R Mutant Inhibitors. *ACS Med Chem Lett*. 2013; 4(2):201–5. <https://doi.org/10.1021/ml300327z> PMID: 24900643
32. Sordella R, Bell DW, Haber DA, Settleman J. Gefitinib-Sensitizing EGFR Mutations in Lung Cancer Activate Anti-Apoptotic Pathways. *Science*. 2004; 305(5687):1163–7. <https://doi.org/10.1126/science.1101637> PMID: 15284455
33. de Gunst MM, Gallegos-Ruiz MI, Giaccone G, Rodriguez JA. Functional analysis of cancer-associated EGFR mutants using a cellular assay with YFP-tagged EGFR intracellular domain. *Mol Cancer*. 2007; 6:56. <https://doi.org/10.1186/1476-4598-6-56> PMID: 17877814
34. Su J, Zhong W, Zhang X, Huang Y, Yan H, Yang J, et al. Molecular characteristics and clinical outcomes of EGFR exon 19 indel subtypes to EGFR TKIs in NSCLC patients. *Oncotarget*. 2017; 8(67):111246–57. <https://doi.org/10.18632/oncotarget.22768> PMID: 29340050
35. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 1993; 234(3):779–815. <https://doi.org/10.1006/jmbi.1993.1626> PMID: 8254673
36. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25(13):1605–12. <https://doi.org/10.1002/jcc.20084> PMID: 15264254
37. Schrödinger Release 2018–3: Maestro, version Schrödinger LLC, New York, NY, 2018.
38. Olsson M. H. M.; Søndergard C. R.; Rostkowski M.; Jensen J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa predictions. *J. Chem Theor Comput*. 2011; 7(2): 525–537.
39. Case DA, Be-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE, Cruzeiro VWD, et al. AMBER 2018, University of California, San Francisco.
40. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. 2015; 11(8):3696–713. <https://doi.org/10.1021/acs.jctc.5b00255> PMID: 26574453
41. Meagher KL, Redman LT, Carlson HA. Development of polyphosphate parameters for use with the AMBER force field. *J Comput Chem*. 2003; 24(9):1016–25. <https://doi.org/10.1002/jcc.10262> PMID: 12759902
42. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983; 79(2):926–35.
43. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995; 103(19):8577–93.

44. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys.* 1984; 81(8):3684–90.
45. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput.* 2013; 9(7):3084–95. <https://doi.org/10.1021/ct400341p> PMID: 26583988
46. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph.* 1996; 14(1):33–8. PMID: 8744570
47. Bakan A, Meireles LM, Bahar I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics.* 2011; 27(11):1575–7. <https://doi.org/10.1093/bioinformatics/btr168> PMID: 21471012
48. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. *MMPBSA.py*: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput.* 2012; 8(9):3314–21. <https://doi.org/10.1021/ct300418h> PMID: 26605738
49. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* 2012; 483:603–607. <https://doi.org/10.1038/nature11003> PMID: 22460905
50. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 2013; 41:955–61.
51. Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price E V., Coletti ME, et al. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov.* 2015; 5:1210–23. <https://doi.org/10.1158/2159-8290.CD-15-0235> PMID: 26482930
52. RStudio Team (2018). RStudio; Integrated Development for R. RStudio, Inc., Boston, MA.
53. Guha U, Chaerkady R, Marimuthu A, Patterson AS, Kashyap MK, Harsha HC, et al. Comparisons of tyrosine phosphorylated proteins in cells expressing lung cancer-specific alleles of EGFR and KRAS. *Proc Natl Acad Sci.* 2008; 105(37):14112–7. <https://doi.org/10.1073/pnas.0806158105> PMID: 18776048
54. Furuyama K, Harada T, Iwama E, Shiraishi Y, Okamura K, Ijichi K, et al. Sensitivity and kinase activity of epidermal growth factor receptor (EGFR) exon 19 and others to EGFR-tyrosine kinase inhibitors. *Cancer Sci.* 2013; 104(5):584–9. <https://doi.org/10.1111/cas.12125> PMID: 23387505
55. Foster SA, Whalen DM, Özen A, Wongchenko MJ, Yin J, Yen I, et al. Activation Mechanism of Oncogenic Deletion Mutations in BRAF, EGFR, and HER2. *Cancer Cell.* 2016; 29(4):477–93. <https://doi.org/10.1016/j.ccell.2016.02.010> PMID: 26996308
56. Wang Z, Cole PA. Catalytic mechanisms and regulation of protein kinases. *Methods Enzymol* 2014; 548:1–21. <https://doi.org/10.1016/B978-0-12-397918-6.00001-X> PMID: 25399640
57. Melvin RL, Salsbury FR. Visualizing ensembles in structural biology. *J Mol Graph Model.* 2016; 67:44–53. <https://doi.org/10.1016/j.jmgm.2016.05.001> PMID: 27179343
58. Shan Y, Arkhipov A, Kim ET, Pan AC, Shaw DE. Transitions to catalytically inactive conformations in EGFR kinase. *Proc Natl Acad Sci.* 2013; 110(18):7270–5. <https://doi.org/10.1073/pnas.1220843110> PMID: 23576739
59. Ou S-HI, Schrock AB, Bocharov E V., Klempner SJ, Haddad CK, Steinecker G, et al. HER2 Transmembrane Domain (TMD) Mutations (V659/G660) That Stabilize Homo- and Heterodimerization Are Rare Oncogenic Drivers in Lung Adenocarcinoma That Respond to Afatinib. *J Thorac Oncol* 2017; 12(3):446–57. <https://doi.org/10.1016/j.jtho.2016.11.2224> PMID: 27903463
60. Wood ER, Truesdale AT, McDonald OB, Yuan D, Hassell A, Dickerson SH, et al. A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* 2004; 64(18):6652–9. <https://doi.org/10.1158/0008-5472.CAN-04-1168> PMID: 15374980
61. Kannan S, Pradhan MR, Tiwari G, Tan W-C, Chowbay B, Tan EH, et al. Hydration effects on the efficacy of the Epidermal growth factor receptor kinase inhibitor afatinib. *Sci Rep.* 2017; 7(1):1540. <https://doi.org/10.1038/s41598-017-01491-z> PMID: 28484248