Journal of Visualized Experiments

Deciphering the Structural Effects of Activating EGFR Somatic Mutations with Molecular Dynamics Simulation

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE61125R1
Full Title:	Deciphering the Structural Effects of Activating EGFR Somatic Mutations with Molecular Dynamics Simulation
Section/Category:	JoVE Biochemistry
Keywords:	Molecular dynamics simulations, Protein structure analyses, Epidermal growth factor receptor family, Somatic mutations, Cancer
Corresponding Author:	Mark Stuart Johnson, Ph.D. Åbo Akademi Faculty of Science and Engineering Turku, Varsinais-Suomi FINLAND
Corresponding Author's Institution:	Åbo Akademi Faculty of Science and Engineering
Corresponding Author E-Mail:	johnson4@abo.fi
Order of Authors:	Mark Stuart Johnson, Ph.D.
	Mahlet Z. Tamirat
	Kari J. Kurppa
	Klaus Elenius
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Open Access (US$4,200)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Turku, Varsinais-suomi, Finland

TITLE:

Deciphering the Structural Effects of Activating EGFR Somatic Mutations with Molecular Dynamics Simulation

AUTHORS:

Mahlet Z. Tamirat[1], Kari J. Kurppa[2], Klaus Elenius[2,3,4], 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]Department of Oncology and Radiotherapy, University of Turku and Turku University Hospital, Turku, Finland

Email addresses of co-authors:
Mahlet Z. Tamirat (mtamirat@abo.fi)
Kari J. Kurppa (kjkurp@utu.fi)
Klaus Elenius (klaele@utu.fi)

Corresponding author:
Mark S. Johnson (johnson4@abo.fi)

Keywords:

Molecular dynamics simulations, Protein structure analyses, Epidermal growth factor receptor family, Somatic mutations, Cancer

Summary

The objective of this protocol is to use molecular dynamics simulations to examine the dynamic structural changes that occur due to activating mutations of the EGFR kinase protein.

Abstract

Numerous somatic mutations occurring in the epidermal growth factor receptor (EGFR) family (ErbB) of receptor tyrosine kinases (RTK) have been reported from cancer patients, although relatively few have been tested and shown to cause functional changes in ErbBs. The ErbB receptors are dimerized and activated upon ligand binding, and dynamic conformational changes of the receptors are inherent for induction of downstream signaling. For two mutations shown experimentally to alter EGFR function, A702V and the ΔELREA deletion mutation, we illustrate in the following protocol how molecular dynamics (MD) simulations can probe the (1) conformational stability of the mutant tyrosine kinase structure in comparison with wild-type EGFR; (2) structural consequences and conformational transitions and their relationship to observed functional changes; (3) effects of mutations on the strength of binding ATP as well as for binding between the kinase domains in the activated asymmetric dimer; and (4) effects of the mutations on key interactions within the EGFR binding site associated with the activated enzyme. The protocol provides a detailed step-by-step procedure as well as guidance that can be more generally

useful for investigation of protein structures using MD simulations as a means to probe structural dynamics and the relationship to biological function.

Introduction:

The human epidermal growth factor receptor (EGFR) family (ErbB) of receptor tyrosine kinases (RTKs) includes four members – EGFR/ErbB1/HER1, ErbB2/HER2, ErbB3/HER3 and ErbB4/HER4. The ErbB receptors regulate fundamental cellular processes such as cell growth and proliferation, differentiation, migration and survival[1,2], and are thus potent proto-oncogenes. Aberrant activity of the ErbB receptors, especially EGFR and ErbB2, has been frequently associated with human cancers making ErbB receptors key targets for cancer therapeutics[2,3].

Several somatic alterations of *ERBB* genes have been reported from human malignancies[3,4,5]. The best characterized examples include the recurrent, activating point mutations and short in-frame deletions in the EGFR kinase domain in non-small cell lung cancer (NSCLC). These EGFR mutations represent key drivers of cancer growth, and predict sensitivity to EGFR targeting cancer drugs[6,7,8]. However, in most cancers, somatic mutations in EGFR occur outside of these recurrent "hotspots" and are distributed over the entire 1210-residue span of the receptor. Indeed, most of the residues along the EGFR primary sequence have been found to be mutated in human cancer[9]. Nevertheless, apart from the few hotspots, the functional significance of the vast majority of the cancer-associated EGFR mutations remains unknown.

The monomeric structure of ErbBs consists of a large amino terminal extracellular domain, followed by a single transmembrane helix leading to the intracellular tyrosine kinase domain and C-terminal tail region that contains docking sites for intracellular signaling proteins. Ligand binding triggers a dramatic conformational change in the extracellular domain, which facilitates the formation of receptor dimers by exposing the dimerization arms that symmetrically cross over each other and interact with their aromatic/hydrophobic surfaces. Upon receptor dimer formation the tyrosine kinase domains come into contact asymmetrically (**Figure 1**), resulting in the activation of the kinases that phosphorylate the C-terminal tails of the receptor monomers, and subsequently in activation of downstream signaling[10, 11].

[Place **Figure 1** here]

Because of the dynamic structural rearrangements that occur during the monomer ⇄ dimer transitions, along with kinase activation that is associated with the formation of an asymmetric dimer, mutations along the full length of the receptor structure can potentially have an effect on receptor function. Here we describe several examples from our previous studies in which modeling of the mutation and visualization were sufficient to explain the consequences for function.

Example 1: One reported mutation, D595V in ErbB4[13], led to increased ErbB4 dimerization and phosphorylation[14]. Visualization of the location of the mutation was a critical factor in understanding the observed functional effects: D595V occurred at the symmetrical crossover of the dimeric arms of the ectodomain (**Figure 2A**). The arms are largely aromatic and

hydrophobic, and the replacement of polar aspartic acid by valine would be expected to increase the "sticky" hydrophobic interactions, stabilizing the dimer and hence increase the length of time when phosphorylation takes place[14]. It was a surprise at first to find aspartate in each arm, but in retrospect one might think of it as a timing mechanism for activity, where the polar acid side chains reduce the affinity and lifetime of the intact dimer and hence limit kinase-mediated phosphorylation and signaling. Replacement by valine would then remove this safeguard by further stabilizing the ErbB4 dimer.

[Place **Figure 2** here]

Example 2: One might anticipate that somatic mutations that target the ATP-binding site of the kinase domain would alter or eliminate enzymatic activity leading to an impaired or kinase-dead receptor incapable of signaling. Of nine reported mutations from patients with breast, gastric, colorectal, or NSCLC[15], two of the nine mutations when tested had highly diminished phosphorylation activity[16]: G802dup (G → GG) and D861Y. Both inactivating somatic mutations were found within the ATP binding site of the tyrosine kinase domain structure (**Figure 2B**): flexible glycine, duplicated, would alter the adenine ring site and small aspartic acid replaced by the bulky tyrosine near the terminal phosphates would physically prevent Mg^{2+}–ATP from binding. However, since ErbB4 can form a heterodimer with ErbB2 – ErbB2 does not bind a growth factor and depends on association with an ErbB that does in order to heterodimerize – the ErbB2(active)-ErbB4(kinase-dead) heterodimer would stimulate cell proliferation via the Erk/Akt signaling pathway yet cells would not differentiate because of kinase-dead ErbB4 and lack of STAT5 pathway activation[16].

In more recent studies, it became apparent that the dynamic movements of ErbBs were relevant to understanding the effects of some mutants on ErbB function, especially mutations that occur within the tyrosine kinase domain. The tyrosine kinase domain consists of an N-lobe (mainly β-sheets) and C-lobe (largely alpha helical), which are separated by the catalytic site where ATP binds. The N-lobe includes the αC helix and P-loop, whereas the activation (A-loop) and catalytic loops are present in the C-lobe[17,18,19]. Crystal structures of the tyrosine kinase domain revealed two inactive conformations, the majority of structures have the Src-like inactive state. In the active conformation the catalytic aspartate of the A-loop points towards the ATP binding site and the αC helix is oriented towards the ATP binding pocket ("αC-in" conformation), forming a strong glutamate-lysine ion-pair interaction.

Because ErbBs and the component kinase domain are highly dynamic entities, and especially for instances where the effects of mutations on function and biological activity are likely to be tightly linked to the conformational states of the ErbBs, it is important to assess mutations with respect to the range of dynamic changes they would experience. X-ray crystal structures of ErbBs provide static snapshots of the 3D structure, which may or may not be relevant to understanding the dynamic consequences of a mutation. In order to probe the range of dynamic changes corresponding to the "energy landscape" available to a three-dimensional (3D) structure, molecular dynamics (MD) simulations are widely used[20]. In the case of mutations that would lead to local conformational changes within the tyrosine kinase domain or stabilization of a complex, simulations on the order of 100 ns may be sufficient. However, larger scale conformational changes (e.g., transitions between the active and inactive

conformations of the kinase domain) require a longer simulation time – on the order of microseconds[21].

With respect to the protocol described below we consider two activating mutations within the tyrosine kinase domain (**Figure 3**). Both mutations are located within the kinase domain at locations that experience *local* conformational changes that dictate whether the kinase is active or not, and thus MD simulations were applied in both instances. In the first case, we consider changes that directly affect the ATP binding site and catalytic machinery of the EGFR receiver kinase domain, specifically examining the consequences of an exon 19 deletion mutation that is widely implicated in NSCLC[4,7]. The Δ^{746}ELREA^{750} mutation, which reduces the length of the β3-αC loop preceding the αC helix – the helix that moves towards the binding/active site on kinase activation and participates in forming the critical electrostatic interaction between E762 of the helix and K745 by positioning the lysine for interaction with ATP – predisposes the domain for activation[12]. In the second case, we consider the A702V mutation of EGFR, shown to be a novel gain-of-function activating mutation revealed by the iScream platform[9] and identified in an NSCLC patient[22]. Alanine-702 on the receiver kinase domain is located on juxtamembrane segment B at the interface of the receiver and the activator kinase domains, in which this asymmetric kinase dimer complex and kinase conformational changes are required for activation[9].

[Place **Figure 3** here]

PROTOCOL

NOTE: The detailed steps taken to examine the effects of the ΔELREA and A702V mutation on the EGFR structure using MD simulations are discussed as follows:

1. Structure preparation

NOTE: In order to study the structural impacts of the ΔELREA mutation, wild-type and mutant forms of apo active, ATP-bound active and apo inactive EGFR monomer structures are prepared as follows.

1.1. Open the **Chimera**[23] visualization program (https://www.cgl.ucsf.edu/chimera/) to prepare the wild-type apo active EGFR kinase structure. In the **File** menu click the **Fetch by ID** option and select the Protein Data Bank (PDB[24]) database and specify PDB code **2GS2**[25] (2.8 Å resolution). The PDB is a repository for 3D structures solved by different experimental techniques, including X-ray crystallography, nuclear magnetic resonance spectroscopy, cryo-electron microscopy and neutron diffraction.

1.2. Build the missing structural elements of 2GS2 by taking these segments from the EGFR PDB structures 1M14[26] (2.6 Å) and 3W2S[27] (1.9 Å). To do so, open 1M14 and 3W2S and superimpose them on 2GS2 using the **MatchMaker** option in the **Tools → Structure comparison** menu.

1.2.1. Crop out the segments to be added from 1M14 and 3W2S. Select the terminal atoms of the residue before the gaps in 2GS2 and the atoms to be added from 1M14 and 3W2S (for

detailed information on the added segments, see **Table 1**). On the command line type **bond sel** and hit **enter**. This structure is the *template* structure.

1.3. Use the template structure from step 1.2 to construct the ΔELREA deletion mutant form of the EGFR kinase domain. Produce the FASTA format sequence of ΔELREA EGFR by saving the sequence of the template structure (**Favorite → Sequence → File → save as**) and then deleting the ELREA sequence at residue numbers 746-750.

1.3.1. Open the ΔELREA EGFR sequence in Chimera and align it with the sequence of the 2GS2 template structure using the **Sequence** menu. In the alignment window select the **Structure → Modeller (homology)**[28] option.

1.3.2. On the pop-up window specify the 2GS2 composite structure as the template and the mutant sequence as the query to be modeled. Then press **OK**. Select a mutant model among the resulting models, based on the zDOPE score (typically the lowest score) and visual inspection.

1.4. To prepare the ATP-bound wild-type EGFR kinase structure, use PDB structure 2ITX[29] (2.98 Å) as the principal structure. Build missing segments (see **Table 1**) using the structures 2GS6[25] (2.6 Å) and 3W2S following the procedure in step 1.2. Convert the ligand ANP in the resulting structure to ATP by opening the PDB file in a text editor and changing the N3B nitrogen atom of ANP to an oxygen atom.

1.5. Open the structure in step 1.4 in Chimera as stated in step 1.1. Add a magnesium ion to this structure from PDB structure 2ITN[29] (2.47 Å) in order to attain a similar positioning for the Mg^{2+} ion.

1.6. With the structure resulting from step 1.5, model the ΔELREA mutant form following the step 1.3.

1.7. To prepare the wild-type apo inactive EGFR kinase structure, open PDB structure 2GS7[25] (2.6 Å) as in step 1.1 and delete the bound ligands and crystallographic waters. Add the missing segments in 2GS7 (see **Table 1**) from structures 3W2S and 4HJO[30] (2.75 Å) using the procedure in step 1.2. Based on the final inactive EGFR structure, prepare the mutant model using the procedure in step 1.3.

NOTE: For the A702V mutation study, the EGFR asymmetric dimer structure is studied as the mutation is located at the juxtamembrane B segment of the kinase domain that forms a large part of the dimer interface. The wild-type and A702V mutant EGFR structures are prepared as follows:

1.8. The wild-type asymmetric dimer structure is constructed from PDB structure 2GS2, which initially is displayed in the monomeric form. To convert to the biological assembly that contains the activator and receiver kinases in the asymmetric arrangement, open 2GS2 in Chimera as in (1.1) and carry out symmetry calculations by clicking the **Tools → Higher-Order Structure → Unit Cell** menu. Select the 2GS2 structure and enter **Make copies**. Finally, select

and save a single asymmetric dimer from the multiple copies of the dimer resulting from the symmetry operations.

1.9. Using the wild-type asymmetric EGFR structure from step 1.8, build the A702V mutant by replacing alanine 702 with valine using the **Tools → Structure editing → Rotamers** option in Chimera.

NOTE: Collectively, six monomeric and two dimeric EGFR structures are prepared for the ΔELREA and A702V mutation studies, respectively. Each structure is subsequently processed for simulation using the protein preparation wizard in the Maestro program[31] and MD simulations are made with the Amber program[32].

1.10. Open the structure in **Maestro** by using the **File → import structure** option. Then click on the **protein preparation wizard** button and select the following: add hydrogen atoms, build missing side-chain atoms, determine protonation states of ionizable residues at pH 7.0 using PROPKA, optimize the orientation of asparagine, glutamine and histidine residues for hydrogen bonding, and finally minimize the structure.

2. System setup

2.1. Open the **leap** program included in the Amber software package. Import the ff14SB force field[33] (**source leaprc.protein.ff14SB**) and TIP3P water molecules[34] **(source leaprc.water.tip3p**). For the ATP-bound systems also import parameters for ATP[35] (**loadamberparams frcmod.phosphate**, **loadamberprep ATP.prep**). Then, load the structure (**mol = loadpdb structure.pdb**).

2.2. Solvate the structure in an octahedral box with explicit TIP3P water molecules that extends 10 Å in all directions from the surface atoms of the protein (**solvateoct mol TIP3PBOX 10.0**).

2.3. Check the built system (**Check mol**) and neutralize it by adding necessary ions (**addions mol Na+ 0**). To sufficiently model biomolecular systems, add additional Na^+/Cl^- atoms to the simulation box to bring the system salt concentration to 0.15 M (**addions mol Na+ X**, **addions mol Cl- X**), where X is replaced by the result of: desired salt concentration * number of water molecules * volume per water molecule * Avogadro's number.

2.4. Generate and save the topology and coordinate files of the system, which serve as inputs for the subsequent production simulation (**saveamberparm mol X.prmtop X.inpcrd**).

3. Molecular dynamics simulation

3.1. Using Amber, initially subject the simulation system to 5000 cycles of steepest descent and conjugate gradient energy minimization to circumvent unfavorable configurations. Carry out the minimization in multiple steps, gradually lowering the restraint applied on solute atoms from 25 mol^{-1} $Å^{-2}$ to 0 kcal mol^{-1} $Å^{-2}$.

3.1.1. In the minimization input file, **min.in,** adjust the **maxcyc** variable for the total minimization cycle (maxcyc = 5000) and **ncyc** to state the number of cycles for the steepest descent algorithm. Use the **restraint_wt** variable to apply the restraint force on solute atoms specified by the **restraintmask** parameter. Then run the minimization as follows:

$AMBERHOME/bin/sander -O -i min.in -o min.out -p X.prmtop -c X.inpcrd -r min.rst -ref X.inpcrd

NOTE: The strategy and actual parameters used may vary according to one's own preferences. Details and guidance can be found from the Amber manual and website (https://ambermd.org/index.php).

3.2. Heat the system for 100 ps from 0 K to 300 K setting a 10 kcal mol^{-1} $Å^{-2}$ restraint on solute atoms. To do so, set **tempi = 0.0**, **temp0 = 300.0**, **dt = 0.002 ps**, **nstlim = 50000** and **restraint_wt = 10** in the **heat.in** input file. Carry out the heating with the following command:

$AMBERHOME/bin/sander -O -i heat.in -o heat.out -p X.prmtop -c min.rst -r heat.rst -x heat.mdcrd -ref min.rst

3.3. Equilibrate the system for 900 ps under an NPT ensemble; constant number of atoms, temperature (**temp0 = 300.0**) and pressure (**ntp = 1**), controlling it with the Berendsen method (**ntt = 1**). Set a 9 Å distance cutoff (**cut = 9.0**) for long range electrostatic interactions. Gradually lower the solute atom restraint to 0.1 mol^{-1} $Å^{-2}$ (**restraint_wt = 0.1**). Run the equilibration input file **equil.in** that describes the above parameters as follows:

$AMBERHOME/bin/sander -O -i equil.in -o equil.out -p X.prmtop -c heat.rst -r equil.rst -x equil.mdcrd -ref heat.rst

3.4. Finalize the equilibration with an unrestrained 5 ns simulation (set **dt = 0.002 ps**, **ntslim = 2500000**).

$AMBERHOME/bin/sander -O -i equil_final.in -o equil_final.out -p X.prmtop -c equil.rst -r equil_final.rst -x equil_final.mdcrd -ref equil.rst

3.5. Check that the system has equilibrated by examining the temperature, pressure, density and energy values.

$AMBERHOME/bin/process_mdout.perl heat.out equil.out equil_final.out
xmgrace summary.TEMP/DENSITY/ETOT/EPTOT/EKTOT

3.6. Carry out the production simulation for 100 ns (set **dt = 0.002 ps**, **ntslim = 50000000** in **prod.in**) and save conformations every 10 ps (**ntwx = 5000**).

$AMBERHOME/bin/sander -O -i prod.in -o prod.out -p X.prmtop -c equil_final.rst -r prod.rst -x prod.mdrcd -ref equil_final.rst

4. Analysis

4.1. Visual inspection

4.1.1. Visualize the conformations sampled during the wild-type and mutant EGFR kinase simulations by opening the **X.prmtop** amber topology files and the corresponding **prod.mdcrd** trajectory files in **VMD**[36]. Using convenient secondary structure representations,

analyze the overall structural dynamics of the proteins from the recorded trajectory. View specific interactions between atoms/residues of interest, such as the catalytically essential K745 — E762 salt bridge.

4.1.2. Alternatively, save multiple conformations sampled during the simulation in PDB format and open them using the **Chimera** program. Superimpose the structures on the initial or median structure using the **MatchMaker** option. Display the initial/median structure in solid and the rest of the aligned structures in faded white. This approach allows one to visualize the recorded structural movements with more clarity.

NOTE: Suggestions for effective representations and processing of conformational ensembles from MDS can be found in Melvin et al.[37].

4.2. RMSD and RMSF analysis

4.2.1. Compute root-mean square deviation (RMSD) and root-mean square fluctuation (RMSF) calculations with the **Cpptraj** program[38] to analyze the global stability of the proteins and examine the flexibility of the different structural units. In the **rmsd.in** and **rmsf.in** input files, indicate the backbone atoms (for RMSD) and Cα atoms (for RMSF) of the initial structure as the reference for RMS fitting. In the **rmsd/rmsf.in** files import the amber topology files (**parm X.promtop**) and the corresponding trajectory files (**trajin prod.mdcrd**). Then run the command **Cpptraj -i rmsd/rmsf.in**. Plot the output data for analysis.

4.2.2. Alternatively, align conformational ensembles and color each residue based on the Cα atom RMSD. To do so, open the conformations in Chimera and align them with **Matchmaker** option.

4.2.2.1. Go to **Tools → Depiction → Render by attribute.** Select **Residues** of the conformational ensemble and **Cα RMSD** as the attributes and click **OK.** The chain trace of the conformations will then be colored from blue → white → red, respectively reflecting regions of high, medium and low structural stability.

4.3. Hydrogen bond analysis

4.3.1. Analyze the hydrogen bond interaction between ATP and wild-type/ΔELREA EGFRs. Prepare a Cpptraj script, **hbond.in**, to carry out this task. Define a hydrogen bond with a donor-acceptor distance of less than or equal to 3.5 Å and a bond angle of greater than or equal to 135˚. Specify analysis only for intermolecular hydrogen bonds with the **nointramol** variable i.e. hydrogen bonds between ATP and EGFR (**hbond All nointramol dist 3.5 out nhb.agr avgout avghb.dat**). Run the script as **Cpptraj -i hbond.in**.

4.3.2. Use this script to assess intramolecular interactions, for instance between residues K745 and E762, which are key residues for EGFR kinase activity. To do so, specify K745 as the hydrogen bond donor and E762 as the hydrogen bond acceptor in **hbond.in** and run the script accordingly.

4.4. Monitoring distance between atoms

4.4.1. Measure the distance between K745 and E762 by opening the trajectories of wild-type and ΔELREA apo EGFRs in **VMD**. Select the Cδ of Glu762 and Nz of Lys745 by clicking **Mouse → label → bond**. Monitor the distance during the simulation by plotting a graph with **graphics → labels → bond → graph**.

4.5. Free energy calculations

4.5.1. To compute the estimated binding free energies between ATP and wild-type/ΔELREA EGFRs, and between the activator and receiver kinases of wild-type/A702V EGFRs, use the molecular mechanics generalized Born surface area (MM-GBSA) module[39] available in the AMBER package. Set ATP as the ligand and EGFR as the receptor in the ΔELREA study. In the A702V study, specify the receiver kinase as the ligand and the activator kinase as the receptor.

4.5.2. First prepare the ligand, receptor and ligand-receptor complex PDB files separately in the **leap** program setting the **PBRadii** value to **mbondi2**. For the PDB files, save the amber topology (.prmtop) and coordinate (.inpcrd) files.

4.5.3. Then, in the mmgbsa input file, **mmgbsa.in**, set **igb = 2**, **saltcon = 0.1**. Execute binding energy calculations using the trajectories of the simulations, the prepared receptor/ligand amber files and the parameters in the **mmgbsa.in** with the **MMPBSA.py** script available in Amber as follows:

$AMBERHOME/bin/MMPBSA.py -O -i mmgbsa.in -o mmgbsa.dat -sp X.prmtop -cp complex.prmtop -rp receptor.prmtop -lp ligand.prmtop -y prod.mdcrd -eo output.csv

4.5.4. Analyze the output data, **output.csv**, by plotting graphs.

REPRESENTATIVE RESULTS

The described protocol was used to study the structural effects of the ΔELREA and A702V mutations on the EGFR kinase structure. One application of the protocol was to investigate the effect of the mutations on local structural/conformational stability by computing RMSD and RMSF values from the MD simulations. As the A702V mutation is located at the juxtamembrane B segment, the RMSD of this segment of the receiver kinase relative to the starting structure was calculated for both the wild-type and A702V EGFRs. The result (**Figure 4A**) revealed that the juxtramembrane B segment of the mutant has increased conformational stability during the 100 ns simulation (average RMSD 0.7 Å – 95% confidence interval (CI) ± 0.009) as compared to the wild-type EGFR kinase domain (average RMSD 1.1 Å – 95% CI ± 0.01). This is very likely the result of the tighter hydrophobic interactions at the dimer interface due to replacement of alanine 702 (methyl group side chain) to a bulkier hydrophobic residue, valine (isopropyl group side chain), leading to increased hydrophobic interactions of V702 on the receiver kinase domain with isoleucine 941 of the activator kinase domain.

The ΔELREA mutation is located at the β3-αC loop, adjacent to the functionally critical αC helix; the conformation of the αC helix is key to shifts between the active and inactive states of the EGFR kinase. The conformational stability of the αC helix in the active state was assessed by examining the RMSF over the Cα atoms of residues within the helix during MD

simulations (**Figure 4B**): overall, there are lower fluctuations in the mutant (average RMSF 1.1 Å – 95% CI ± 0.4) in comparison to the wild-type (average RMSF 1.5 Å – 95% CI ± 0.57); with the highest difference in fluctuations recorded for the N-terminal residues. The sampled conformations respectively superposed on the median structure of the wild-type kinase domain and of the ΔELREA kinase domain also support these results (**Figure 4C**): both the wild-type and ΔELREA kinase domains have overall similar stability for the superposed conformations except for the β3-αC loop and the αC helix, which are clearly more stable in ΔELREA EGFR. These findings indicate that deletion of the ELREA sequence restrains the movement of the active state αC helix, hence restraining and thus stabilizing the active conformation. Furthermore, since the αC helix forms part of the asymmetric dimer interface, the restraints on the αC helix in the mutant would very likely stabilize the asymmetric dimer, prolonging the duration of the activated state.

Another application of the protocol is to investigate the behavior of key intra- and intermolecular interactions taking place during the simulation. Thus, the interaction between K745 and E762, which is fundamental to EGFR enzymatic activity, was analyzed for both the active form wild-type and ΔELREA EGFR kinase by measuring the percentage occupancy of hydrogen bonds formed between the side-chain polar atoms of the two residues during the MD simulations (**Figure 5A**): this key electrostatic interaction was formed more often in the ΔELREA kinase domain in comparison to the wild-type kinase domain, owing to the more stable αC helix that accommodates E762. The interactions between Mg^{2+}–ATP and the wild-type and ΔELREA EGFR kinase domains (**Figure 5B**) during the simulation were also assessed (**Figure 5C**): the number of hydrogen bonds were greater for ΔELREA (average value 4.0 – 95% CI ± 0.03) than for the wild-type EGFR (average value 3.2 – 95% CI ± 0.04). Further analysis of the hydrogen bonds revealed that K745 interacts more frequently with the phosphate groups of ATP in ΔELREA EGFR, which is linked to the more stable K745-E762 interaction noted in the simulation of the ΔELREA mutant EGFR kinase domain.

MD simulations as described in the protocol are also useful in assessing the relative free energy of binding for protein-protein and protein-ligand interactions. The binding energies between the activator and receiver kinase domains of wild-type and A702V EGFRs, and between ATP and the wild-type and ΔELREA mutant EGFR kinase domains, were computed from molecular mechanics generalized Born surface area (MMGBSA) calculations (**Figure 6A**): the A702V mutant produced a lower average ΔG_{bind} value (average ΔG_{bind} = -76 kcal/mol – 95% CI ± 0.47), representing more favorable dimer interactions, in contrast to the wild-type EGFR domain (average ΔG_{bind} = -61 kcal/mol – 95% CI ± 0.61). This observation is consistent with the more stable juxtamembrane B segment and tighter dimer interface due to increased hydrophobic interactions observed for the A702V EGFR kinase domain. In the case of ATP binding to the wild-type and ΔELREA EGFR kinase domains (**Figure 6B**), MMGBSA calculations predict stronger ATP binding with the ΔELREA mutant (average ΔG_{bind} -57 kcal/mol – 95% CI ± 0.43) in comparison to wild-type EGFR (average ΔG_{bind} -48 kcal/mol – 95% CI ± 0.33). This outcome is in line with the greater number of hydrogen bonds recorded between ATP and ΔELREA EGFR (**Figure 5C**) in comparison to the wild-type domain.

The protocol can also be employed to investigate conformational changes observed during a simulation. In the current study, the effects of the ΔELREA mutation on the inactive EGFR conformation was studied by visual inspection and superpositioning of the sampled

conformations from the simulation. The analysis uncovered an inward movement of the αC helix in the ΔELREA EGFR kinase domain (**Figure 7A**), a structural change expected during the transition to the active state. In contrast, the αC helix of wild-type inactive EGFR maintained its initial conformation (**Figure 7B**). Thus, the MD simulations support the proposal that the deletion mutation, shown experimentally to increase kinase activity[40,41], promotes a conformational shift from the inactive kinase towards the active state.

Table 1: Structures used to build composite models of apo active, apo inactive and ATP-bound active structures. Missing regions (amino acid range in parentheses) in the principal structure were constructed from the listed structures.

Figure 1: Structure of the EGFR dimer. EGFR dimerizes when the extracellular domains bind growth factor (EGF, epidermal growth factor). The receiver kinase domain is then activated through asymmetric interaction with the activator kinase domain, and the C-terminal tails are autophosphorylated at tyrosine residues (Modified from Tamirat et al.[12]).

Figure 2: Location of an ErbB4 activating mutation and mutations producing kinase-dead ErbB4. (**A**) D595 (activating D595V mutation) is located on the aromatic/hydrophobic dimeric arms of the ErbB4 ectodomain model; the arms associate on growth factor binding; (nearby residues are shown as sticks). (**B**) In ErbB4, G802 (inactivating G802dup mutation) helps form the binding pocket around the adenine ring of ATP and catalytic D861 (inactivating D861Y mutation) binds both Mg^{2+} (not shown) and the γ-phosphate group of ATP.

Figure 3: The asymmetric kinase domain dimer of EGFR. The A702V mutation would be located at the critical interface of the activator and receiver kinase domains, adjacent to the αC helix and close to isoleucine 941 of the activator kinase. Conformational changes induced by formation of the asymmetric dimer lead to kinase activation. The β3-αC loop containing the ELREA sequence directly precedes the αC helix; during activation, the αC helix moves inward towards the ATP binding site.

Figure 4: Wild-type and mutant conformational stability of the active EGFR kinase domain during MD simulations. (**A**) RMSD (backbone atoms) over the juxtamembrane B segment of the wild-type (blue) and A702V (red) receiver kinase domain. (**B**) RMSF (Cα atoms) over residues of the αC helix: wild-type (blue) and ΔELREA (gold). (**C**) Superposed sampled conformations of wild-type (left) and ΔELREA (right) EGFR kinase domain; chain traces colored based on RMSD (Cα atoms) of each residue with respect to the median structure. Coloring ranges from blue to white to red, representing regions of high to low conformational stability. Note that the "free" N-terminal regions of the isolated kinase domain, colored red, would not exhibit this level of mobility in the intact EGFR structure. Figures adapted from Chakroborty et al.[9] (Figure 4A reproduced with permission of the *Journal of Biological Chemistry*) and Tamirat et al.[12].

Figure 5: Key features seen in the active receiver kinase during MD simulations: the K745-E762 salt bridge, the αC helix and interactions with. (**A**) Percentage occupancy of the K745-E762 interaction during the simulation of the wild-type (blue) and ΔELREA (gold) EGFR kinase domains. (**B**) Residues of the wild-type and ΔELREA mutant interacting with ATP (sticks). Mg^{2+} (green) coordinates with ATP and D855. (**C**) The number of hydrogen bonds formed by ATP

with both the wild-type and ΔELREA EGFR kinase domains during MD simulations. Figure from Tamirat et al.[12].

Figure 6: Lower relative free energies of binding are observed for the mutant kinase domains during the simulations. (**A**) Binding energies calculated for the interaction between the activator and receiver kinase domains of the wild-type (blue) and A702V (red) EGFRs. (**B**) ΔG_{bind} of ATP to wild-type (blue) and ΔELREA (gold) EGFR kinase domains. Figures adapted from Chakroborty et al.[9] (Figure 6A reproduced with permission of the *Journal of Biological Chemistry*) and Tamirat et al.[12].

Figure 7. Superposed conformations from the wild-type and ΔELREA inactive EGFR kinase domain. Conformation of the αC helix and A-loop helix of (**A**) wild-type (median structure in blue) and (**B**) ΔELREA EGFRs (gold). Other sampled conformations, faded white; initial structures prior to the MD simulations, pink. Figure from Tamirat et al.[12].

DISCUSSION

The protocol described in this study focuses on using molecular dynamics simulations to investigate local and global structural alterations that arise from activating somatic mutations of the EGFR kinase domain. Although X-ray crystal structures of wild-type and mutant EGFRs provide invaluable structural insight, they depict one or a few static representations. However, inherent to the biological function of ErbBs is the necessary transitions between the enzymatically inactive and active tyrosine kinase, invoking dynamic changes in both the structure and intramolecular interactions between kinase monomers. MD simulations were thus carried out to probe the dynamic nature of the EGFR tyrosine kinase domain, including the wild-type structure, the introduced ΔELREA deletion mutation, and the A702V mutation. These simulations were successful in elucidating the likely role of these mutations in the structures and how their effects on the conformation of the tyrosine kinase domain would lead to the experimentally observed increases in EGFR kinase activity.

A crucial step in this protocol is the use of a relevant structure to assess the impact of the mutation. One way to select a relevant simulation input structure is to visualize the location of the mutation in the static 3D structure and examine its possible impact with respect to the neighboring amino acids and structural units. In this study, for instance, since the A702V EGFR mutation is located at the juxtamembrane B segment that forms the asymmetric dimer interface, the use of the dimer structure for the simulation as opposed to the monomer is critical. The use of a monomeric structure would have exposed the juxtamembrane B segment of the receiver kinase to the solvent, depriving it from the stabilizing interactions, enhanced by the mutation to a larger hydrophobic residue and interactions with isoleucine 941 from the C-lobe residues of the activator kinase. Moreover, it is noteworthy that the 3D structure represented by the coordinates in a PDB file do not necessarily correspond to the biologically relevant structure that should be used for study. For example, with the structure of ErbB4, PDB code 3BCE, the PDB coordinates correspond to a trimer, but this is due to crystal contacts (few contacts between the monomers are seen when visualizing this structure). Matrices within the PDB file can be used (e.g., within Chimera) to reconstruct the crystallographically related structures, which can be visualized to identify chains that correspond to the biologically relevant 3D structure as reported in the original publication[42]. Another essential step of the protocol is to properly prepare the simulation input structure, such as building

missing amino acids in different loop regions, and especially where located in the vicinity of the mutation. Although numerous wild-type EGFR structures exist in the PDB, only a limited number of mutant EGFR structures are available. Consequently, the mutant structures also need to be modeled; for a single residue mutation like A702V, Chimera was used to mutate the residue; whereas, for the ΔELREA deletion mutation, Modeller was used.

The various parameters utilized in the simulation input files – for example, the number of minimization cycles, heating the system to the desired temperature in one go or instead heating slowly through several intermediate temperatures, the time period for the equilibration and for the production simulations – can be modified based on the molecule of study, the aim of the work and one's own preferences. While carrying out MD simulations, it is also common to come across errors that can arise from the input files, issues related to the simulation software in use or even a user error. Hence, it is very important to understand the source of the errors by carefully examining any error messages. Most simulation programs have a mailing list where users can pose questions to the software developers and to other users by which most problems can be resolved. Additionally, user manuals provide significant assistance to understanding the details of the simulation protocol, including assumptions and limitations. Although MD simulation is an important tool to explore the dynamic properties of molecules, remember that computational results need to be carefully evaluated in conjunction with other sources of information to assess their validity. Whenever possible, work together with researchers that are experts on the proteins under study, especially where relevant wet-lab experimental studies are made, which serve to provide results for structural interpretation as well as to suggest experiments that may be made based on structural observations to test hypotheses.

In this study, the protocol was effective in examining the dynamic structural impacts of the ΔELREA and A702V mutations on the EGFR kinase structures. The simulations revealed that ΔELREA restrains the functionally essential αC helix and promotes a conformational shift from the inactive kinase to a stabilized active kinase. The simulation results are independently supported by drug response data that demonstrated the effects of tyrosine kinase inhibitors on lung cancer cell lines having the ΔELREA deletion mutation and wild-type EGFR, where greater inhibition by drugs recognizing the active kinase conformation was reported for ΔELREA than for wild-type EGFR[12]. With the A702V mutation, the MD simulations indicate, in comparison to the wild type, increased stabilization of the activator-receiver kinase interface as well as higher affinity of the activator and receiver kinase for each other, together supporting maintenance of the activated conformation of the EGFR kinase. The A702V mutation, located on the juxtamembrane B segment of the receiver kinase, would increase hydrophobic interactions with the activator kinase, functioning to prolong the duration of the activated state. The A702V mutation supports cell survival in the *absence* of growth factor and was identified in an in vitro screening for EGFR mutations[9].

Disclosures

The authors have nothing to disclose.

Acknowledgments

This research is funded by grants to M.S.J from the Academy of Finland (308317, 320005), Sigrid Juselius Foundation and Tor, Joe and Pentti Borg memorial fund, and to K.E. from the

Academy of Finland (274728, 316796), the Cancer Foundation of Finland, and the Turku University Central Hospital. M.Z.T. is funded by the Åbo Akademi Doctoral Network of Informational and Structural Biology. We thank the CSC IT Center for Science for the computing resources and Dr. Jukka Lehtonen for the IT support under the Biocenter Finland bioinformatics network; and Biocenter Finland structural biology infrastructure network.

References

1. Yarden, Y., Sliwkowski, M. X. Untangling the ErbB signalling network. Vol. 2, *Nature Reviews Molecular Cell Biology*. **2**, 127–37 (2001).
2. Lemmon, M. A., Schlessinger, J., Ferguson, K. M. The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harbor Perspectives in Biology*. **6**, a020768 (2014).
3. Arteaga, C. L., Engelman. J. A. ERBB receptors: From oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell*. **2**, 282–303 (2014).
4. Mishra, R., Hanker, A. B., Garrett, J. T. Genomic alterations of ERBB receptors in cancer: Clinical implications. *Oncotarget*. **8**, 114371–92 (2017).
5. Cerami E. et al. [webpage]. https://www.cbioportal.org/ (2012).
6. Lynch, T. J. et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non–small-cell lung cancer to gefitinib. *New England Journal of Medicine*. **350** (21), 2129–2139 (2004).
7. Paez, J. G. et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*. **304** (5676), 1497–1500 (2004).
8. Pao, W. et al. EGF receptor gene mutations are common in lung cancers from ‘‘never smokers’’ and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proceedings of the National Academy of Sciences U.S.A*. **101** (36), 13306–13311 (2004).
9. Chakroborty, D. et al. Unbiased in vitro screen for activating EGFR mutations. *Journal of Biological Chemistry*. **294** (24), 9377–9389 (2019).
10. Leahy, D. J. Structure and Function of the Epidermal Growth Factor (EGF/ErbB) Family of Receptors. *Advances in Protein Chemistry*. **68**, 1–27 (2004).
11. Roskoski, R. ErbB/HER protein-tyrosine kinases: Structures and small molecule inhibitors. *Pharmacological Research*. **87**, 42–59 (2014).
12. Tamirat, M. Z., Koivu, M., Elenius, K., Johnson, M.S. Structural characterization of EGFR exon 19 deletion mutation using molecular dynamics simulation. *PLoS ONE*. **14** (9), e0222814 (2019).
13. Ding, L. et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*. **455**, 1069–1075 (2008).
14. Kurppa, K. J., Denessiouk, K., Johnson, M. S., Elenius, K. Activating somatic ERBB4 mutations in non small-cell lung cancer. *Oncogene*. **35** (10), 1283–1291 (2016).
15. Soung, Y. H. et al. Somatic mutations of the ERBB4 kinase domain in human cancers. *International Journal of Cancer*. **118**, 1426–1429 (2006).
16. Tvorogov, D. et al. Somatic mutations of ERBB4: selective loss-of-function phenotype affecting signal transduction pathways in cancer. *Journal of Biological Chemistry*. **284**, 5582–5591 (2009).
17. Hubbard, S. R., Till, J. H. Protein tyrosine kinase structure and function. *Annual Review of Biochemistry*. **69** (1), 373–398 (2000).
18. Huse, M., Kuriyan, J. The conformational plasticity of protein kinases. *Cell*. **109** (3), 275–282 (2002).

19. Jura, N. et al. Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. *Molecular Cell*. **42**, 9–22 (2011).
20. M. Karplus, M., Kuriyan, J. Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences U.S.A*. **102** (19), 6679–6685 (2005).
21. Shan, Y., Arkhipov, A., Kim, E. T., Pan, A. C., Shaw, D. E. Transitions to catalytically inactive conformations in EGFR kinase. *Proceedings of the National Academy of Sciences U.S.A*. **110** (18), 7270–5 (2013).
22. Reckamp K. L. et al. A phase I trial to determine the optimal biological dose of celecoxib when combined with erlotinib in advanced non-small cell lung cancer. *Clinical Cancer Research*. **12**(11 I), 3381–8 (2006).
23. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. **25** (13), 1605–12 (2004).
24. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Research*. 28 (1), 235–42 (2000).
25. Zhang, X., Gureasko, J., Shen, K., Cole, P. A., Kuriyan, J. An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell*. **125** (6), 1137–49 (2006).
26. Stamos, J., Sliwkowski, M. X., Eigenbrot, C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *Journal of Biological Chemistry*. **277** (48), 46265–72 (2002).
27. Sogabe, S. et al. Structure-Based Approach for the Discovery of Pyrrolo[3,2-d]pyrimidine-Based EGFR T790M/L858R Mutant Inhibitors. *ACS Medicinal Chemistry Letters*. **4** (2), 201–5 (2013).
28. Sali, A., Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*. **234** (3), 779–815 (1993).
29. Yun, C. H. et al. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell*. **11** (3), 217–27 (2007).
30. Park, J. H., Liu Y., Lemmon, M. A., Radhakrishnan, R. Erlotinib binds both inactive and active conformations of the EGFR tyrosine kinase domain. *Biochemical Journal*. **448** (3), 417–23 (2012).
31. Schrödinger Release 2018–3: Maestro, version Schrödinger LLC, New York, NY, 2018.
32. Case, D. A. et al. AMBER 2018, University of California, San Francisco.
33. Maier, J. A, Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*. **11** (8), 3696–713 (2015).
34. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*. **79** (2), 926–35 (1983).
35. Meagher, K. L., Redman, L. T., Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *Journal of Computational Chemistry*. **24** (9), 1016–25 (2003).
36. Humphrey, W., Dalke, A., Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. **14** (1), 33–8 (1996).
37. Melvin, R. L., Salsbury, F. R. Visualizing ensembles in structural biology. *Journal of Molecular Graphics and Modelling*. **67**, 44–53 (2016).

38. Roe, D. R., Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*. **9** (7), 3084–95 (2013).
39. Miller, B. R. et al. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation*. **8** (9), 3314–21 (2012).
40. Guha, U. et al. Comparisons of tyrosine phosphorylated proteins in cells expressing lung cancer-specific alleles of EGFR and KRAS. *Proceedings of the National Academy of Sciences U.S.A*. **105** (37), 14112–7 (2008).
41. Furuyama. K. et al. Sensitivity and kinase activity of epidermal growth factor receptor (EGFR) exon 19 and others to EGFR-tyrosine kinase inhibitors. *Cancer Science*. **104** (5), 584–9 (2013).
42. Qiu, C. et al. Mechanism of Activation and Inhibition of the HER4/ErbB4 Kinase. *Structure*. **6** (3), 460–7 (2008).

Figure 1

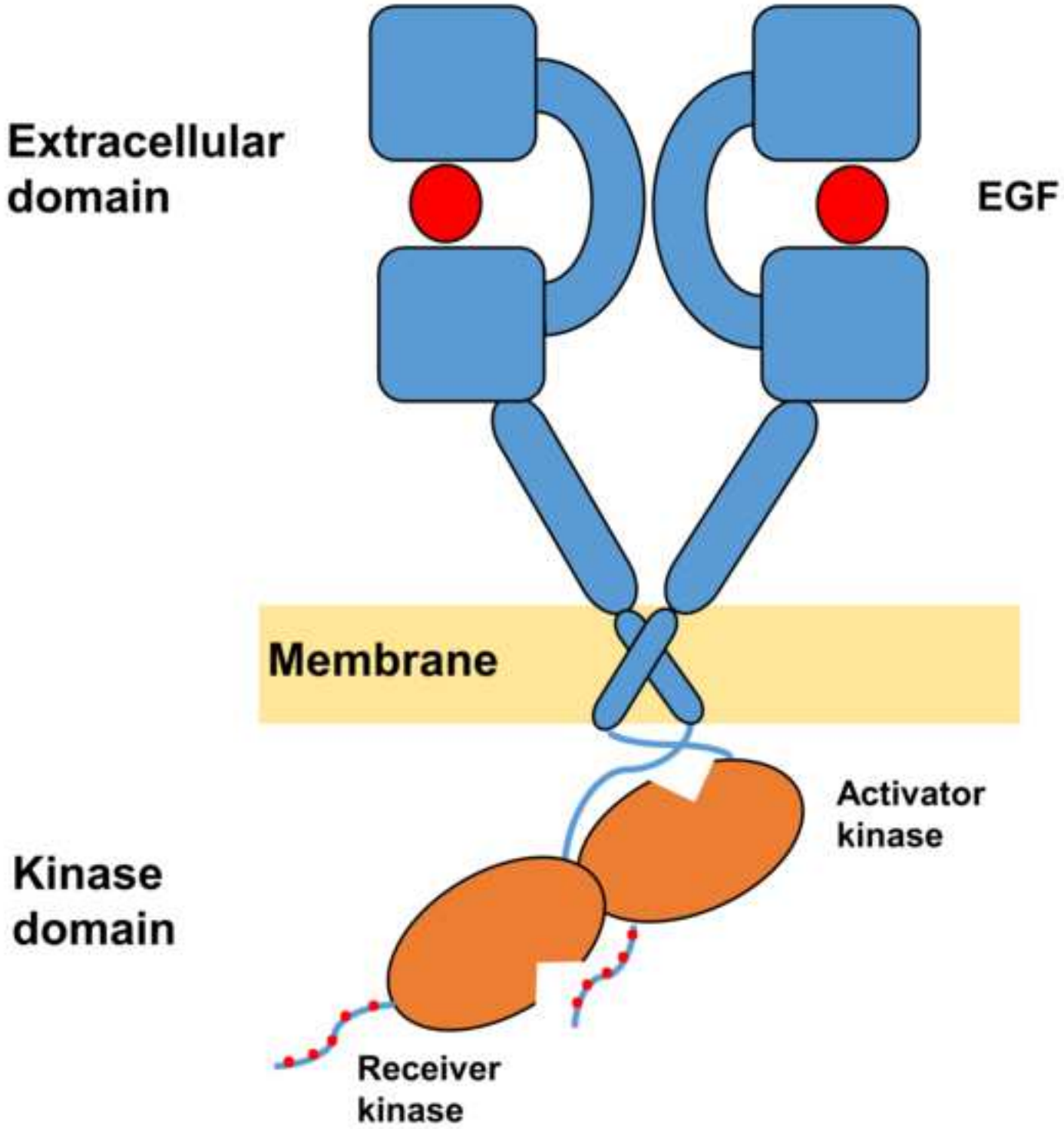


Figure 2

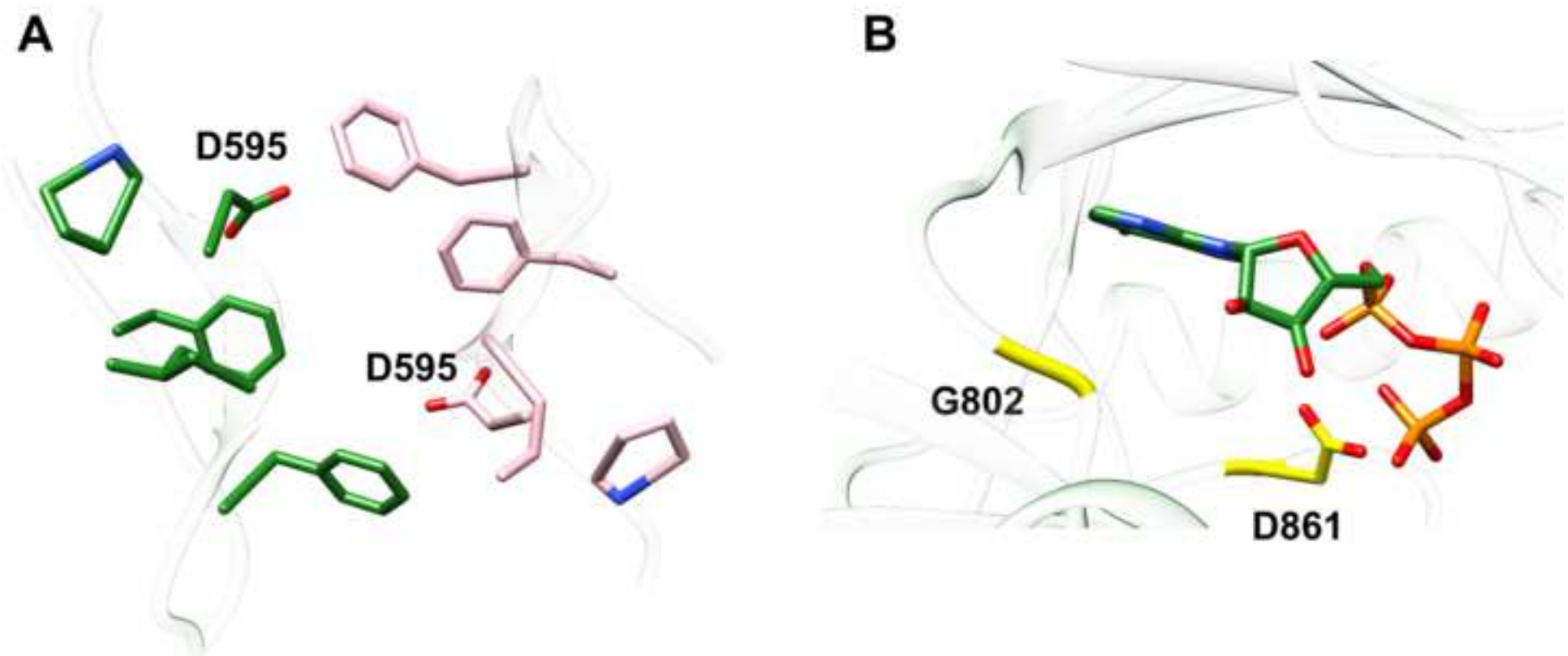


Figure 3

Click here to access/download;Figure;Figure 3.tif

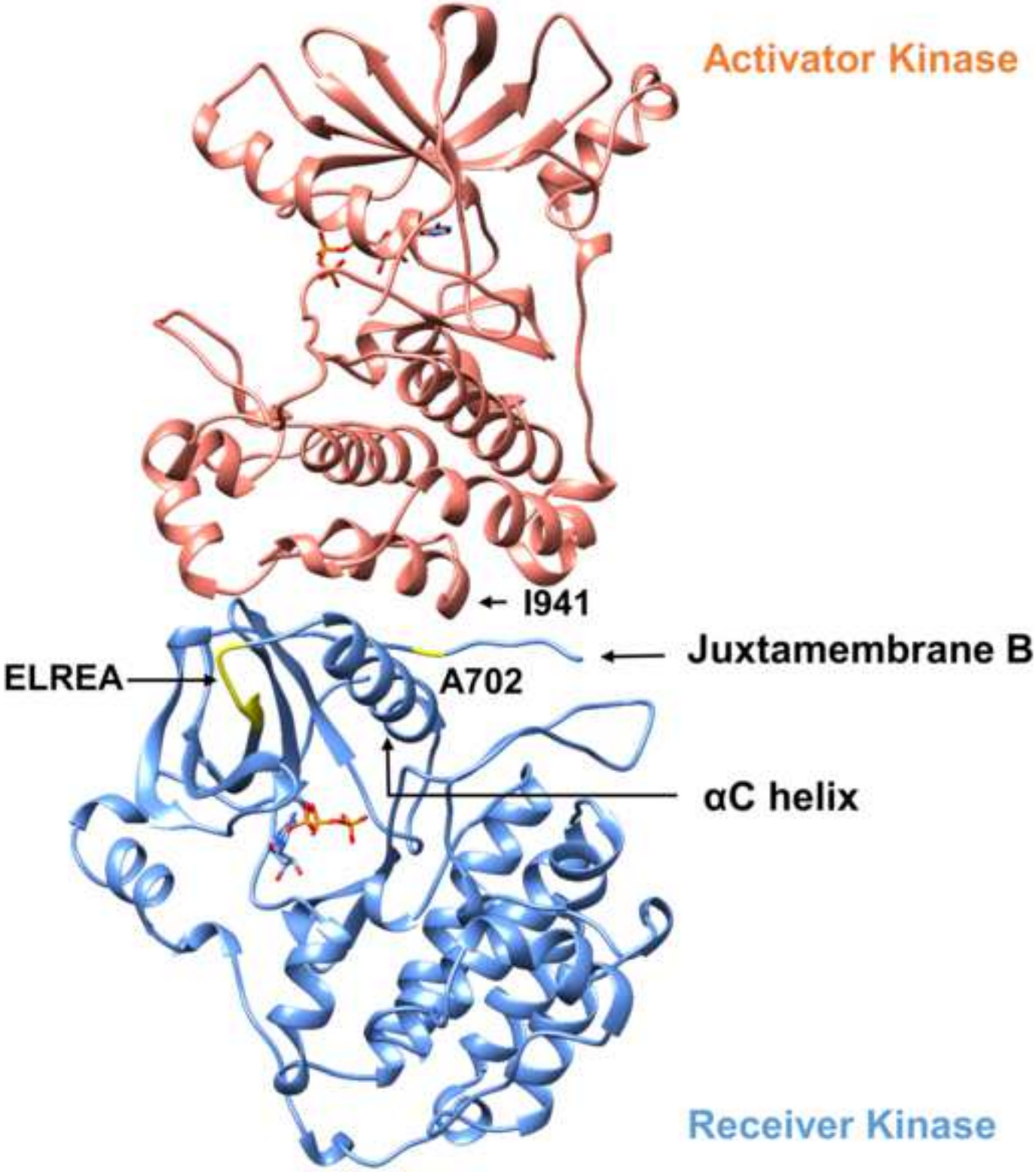


Figure 4

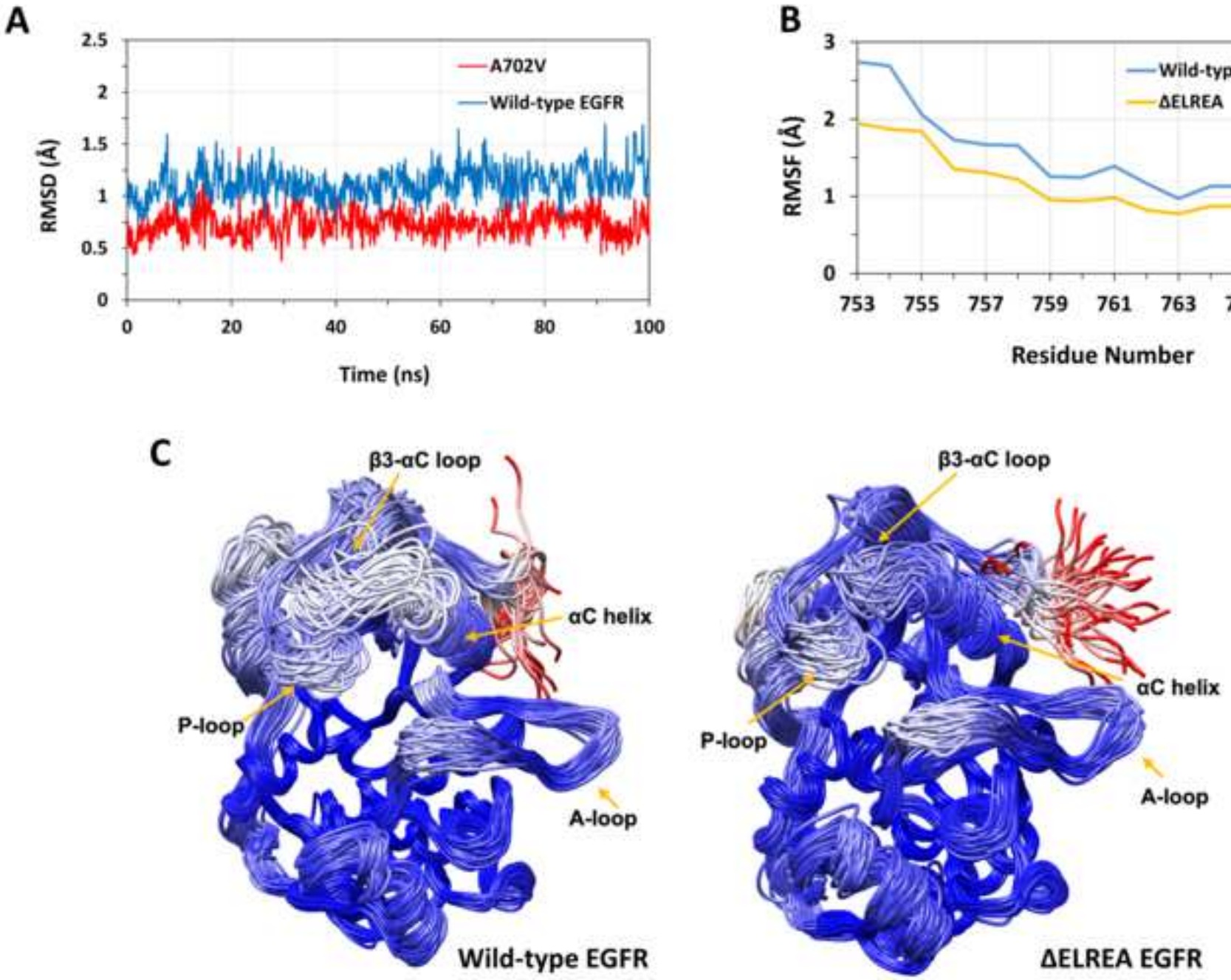


Figure 5

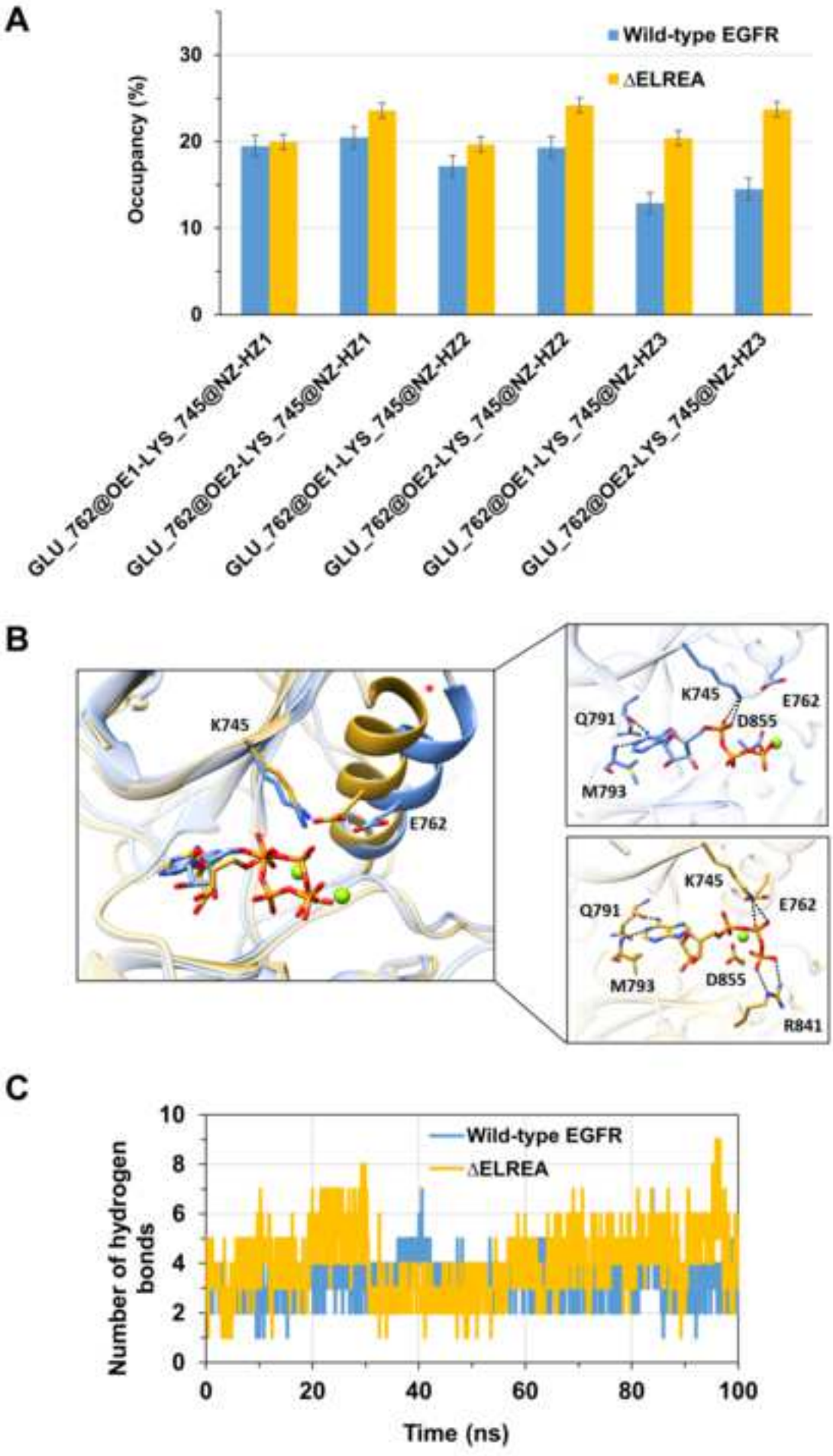


Figure 6

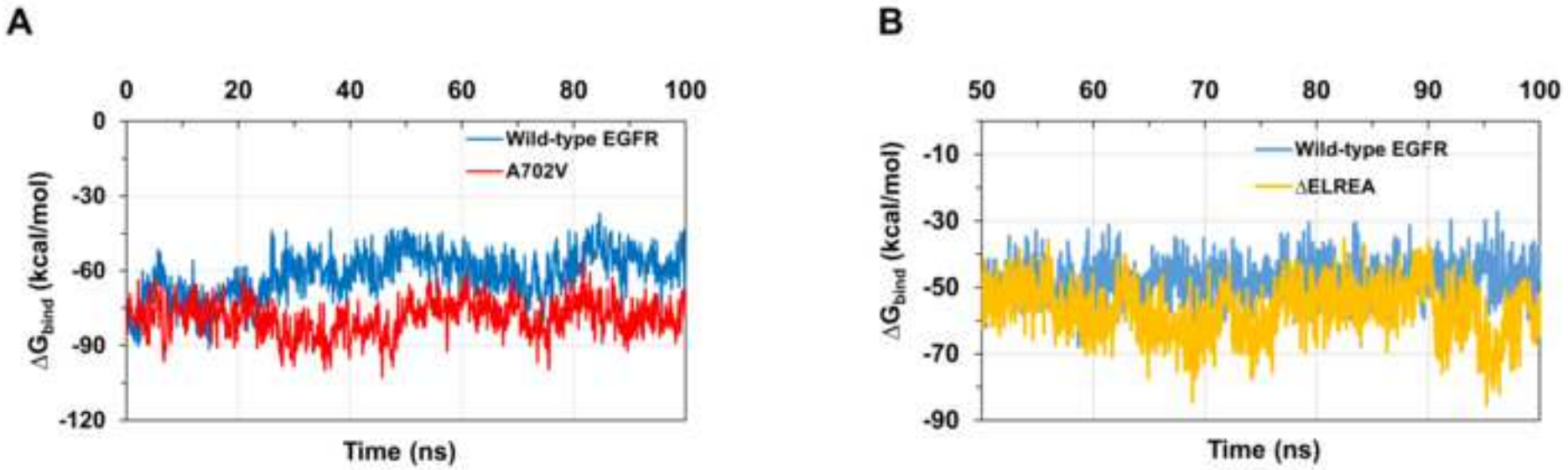
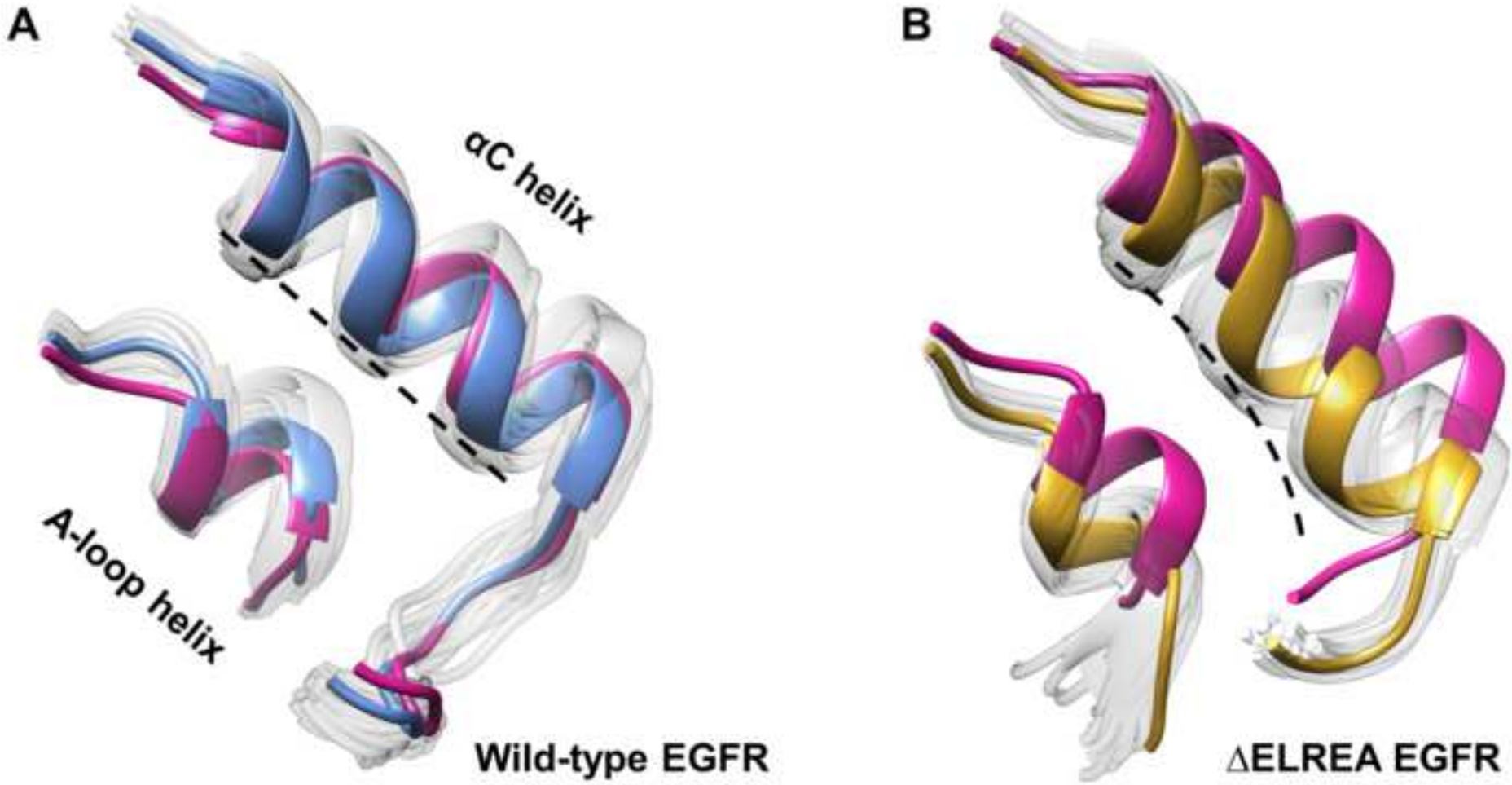


Figure 7



Table

	Apo active EGFR	Apo inactive EGFR
Principal structure	2GS2	2GS7
Structures used to build missing loops	1M14 (723-725)	3W2S (958-984)
	3W2S (967-981)	4HJO (848-850)

ATP-bound active EGFR
2ITX
2GS6 (862-865)
3W2S (990-1001)

Name of Material/ Equipment	Company	Catalog Number
Amber software	University of California, San Francisco	Version 2018
Chimera program	Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco	Version 1.13.1
EGFR struture files	The Protein Data Bank	
Maestro	Schrödinger LLC	Version 2018-3
Modeller program	The Andrej Šali Lab, Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, University of California San Francisco	
VMD software	Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign	Version 1.9.3

Comments/Description
Executable
Executable
3D coordinates of EGFR structures
Executable
Included in the Chimera program
Executable

First of all, we would like to thank the editor and the reviewers for their constructive comments. Replies to the raised questions/comments are written in bold.

Editorial comments:
Changes to be made by the Author(s):
1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues. The JoVE editor will not copy-edit your manuscript and any errors in the submitted revision may be present in the published version.

The manuscript has been thoroughly checked for spelling/grammar issues.

2. Please revise the table of the essential supplies, reagents, and equipment. The table should include the name, company, and catalog number of all relevant materials in separate columns in an xls/xlsx file. Please sort the Materials Table alphabetically by the name of the material.

The table has now been edited by sorting the materials name alphabetically. Other details on the materials (name, company) has already been included.

3. Please obtain explicit copyright permission to reuse any figures from a previous publication. Explicit permission can be expressed in the form of a letter from the editor or a link to the editorial policy that allows re-prints. Please upload this information as a .doc or .docx file to your Editorial Manager account. The Figure must be cited appropriately in the Figure Legend, i.e. “This figure has been modified from [citation].”

Copyright permission for reused figures (Figure 4A and Figure 6A) has been attached. The reused figures are cited in the text as recommended by JoVE. All other figures are originals or from a PLoS ONE article in which we own the copyright.

4. Please revise the text to avoid the use of any personal pronouns (e.g., "we", "you", "our" etc.).

Personal pronouns are not used in the Protocol. However, in the Introduction section we have used few pronouns as we are referring our previous works.

5. The Protocol should contain only action items that direct the reader to do something. Please move the discussion about the protocol to the Discussion.

The discussion section present at the beginning of the protocol has now been moved to the Discussion.

6. Please highlight 2.75 pages or less of the Protocol (including headings and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol. Remember that non-highlighted Protocol steps will remain in the manuscript, and therefore will still be available to the reader.

Parts of the protocol to be included in the video are highlighted in yellow.

7. Please ensure that the highlighted steps form a cohesive narrative with a logical flow from one highlighted step to the next. Please highlight complete sentences (not parts of sentences). Please ensure that the highlighted part of the step includes at least one action that is written in imperative tense.

We have highlighted the protocol considering a cohesive narrative. Parts of a sentence are however highlighted in some cases to leave out unnecessary or contextual details for the video recording.

Reviewers' comments:
Reviewer #1:

Manuscript Summary:
The manuscript "Deciphering the Structural Effects of Activating EGFR Somatic Mutations with Molecular Dynamics Simulation" is not a interesting manuscript for publication. The idea to study two new mutation is interesting however there is lack of the deep interpretreation the data from MD.

Major Concerns:
The abstract has no information about the finding from this study. The proposal is not new. It's a trival method which can be following from any tutorial from internet. Characterization of interaction comparing between wildtype and mutants can be more details than described in the manuscript.

JoVE articles are aimed to thoroughly discuss the method/protocol used in a study rather than the results, as stated on the journal's page. Therefore, in the abstract and elsewhere we have focused more on how the described protocol can be used to study the structural impacts of mutations that have been published in detail elsewhere. Furthermore, since only representative results are discussed in the article, a more detailed analysis of the MD data has not been presented.

Reviewer #2:

Manuscript Summary:
The authors have well drafted the manuscript entitled "Deciphering the Structural Effects of Activating EGFR Somatic Mutations with Molecular Dynamics Simulation". The authors have tried to reveal the structural impacts of the mutations in the EGFR protein. Though the script is interesting, it needs several modifications before accepting for publication.

Major Concerns:
The AMBER dynamics tool has its own protocol page at https://ambermd.org/tutorials/ How is your proposed protocol serve more significance than the existing?

The AMBER webpage indeed describes several protocols that can be used to assess systems with molecular dynamics simulations. Here, we have tailored the protocol to be used to probe mutation-induced structural changes of the EGFR protein. Moreover, our protocol describes not only the step-by-step procedures of performing simulations with AMBER, but also properly preparing input proteins structures and analyzing the output data, which are key parts of the process.

Minor Concerns:

1. Why authors have highlighted few paragraphs in yellow? Does it mean that, it should be reviewed and it is not the clean copy?

The yellow highlighted section of the protocol is the part that is going to be included in the video, as instructed by JoVE.

2. The references for the last Table are missing. Authors should mention the reference of each software.

The references for each software are already included in the main text. However, the journal does not require references to be included in the excel file containing the table.

3. In the RMSD plot, the deviation in the mutation condition is lesser than the native. How does author see this? Explain

As described on lines 368-377 of the manuscript, the lower deviation values recorded for the juxtamembrane segment of the mutant show that this segment has a higher stability in the mutant as compared to the wild-type EGFR. This results from the replacement of A702 in wild-type EGFR by valine in the mutant, which results in tighter hydrophobic interactions at the dimer interface that accommodates the juxtamembrane segment.

4. What is the clinical significance of the chosen mutations?

The ΔELREA mutation is one of the most common activating EGFR mutation that is widely implicated in non-small cell lung cancer. We have now included this information on lines 148-149. The A702V mutation was studied as it was identified in our previous study that screened somatic mutations of EGFR, and A702V was determined as a gain-of-function mutation by the iSCREAM platform and reported in one cancer patient. This information is mentioned on lines 152-154 of the manuscript.

5. Why authors have not attempted to compare the SNP and deletion mutation against native in a single graph plot?

The data/variable being compared for the SNP and the deletion mutation are different. For instance, in Figure 4A we have compared RMSD for the SNP against wild-type, whereas in Figure 4B we have compared RMSF of the deletion mutant against wild-type. In Figure 6A we have assessed the free binding energy between EGFR dimers for the SNP Vs wild-type. In contrast, the binding energy in Figure 6B is between ATP and ELREA/wild-type EGFR. The variables being studied are different hence they were plotted in separate graphs.

6. The hydrogen bonds in mutation condition is more than native condition. You mean mutation is more needed for the protein to maintain its secondary structure?

No, in Figure 5A the percentage occupancy of hydrogen bonds formed between K745 and E762 of the salt bridge are shown. In Figure 5C the number of hydrogen bonds between ATP and wild-type/ΔELREA EGFR formed during the simulations are plotted. The figure shows that the mutant EGFR makes more hydrogen bond interaction with ATP as compared to the wild-type. This was linked to the stability of the K745-E762 bond in the mutant EGFR, which led to a more conserved interaction between ATP and K745. This information is present on lines 412-418 of the manuscript.

7. Author might check few latest publications on dynamics simulations and mutations for linking genotype-phenotype
(i) Genotype-phenotype correlation in 18 Egyptian patients with glutaric acidemia type I

Genetic Epidemiology of Glucose-6-Dehydrogenase Deficiency in the Arab World
(ii) Evolution- and structure-based computational strategy reveals the impact of deleterious missense mutations on MODY 2 (maturity-onset diabetes of the young, type 2)
(iii) Extrapolating the effect of deleterious nsSNPs in the binding adaptability of flavopiridol with CDK7 protein: A molecular dynamics approach
(iv) Path to facilitate the prediction of functional amino acid substitutions in red blood cell disorders - A computational approach

We thank the reviewer for the kind suggestions and we will consider them in future studies.

Copyright Clearance Center Marketplace™

American Soc for Biochemistry & Molecular Biology - License Terms and Conditions

Order Date	02-Mar-2020
Order license ID	1020888-1
ISSN	1083-351X
Type of Use	Republish in a journal/magazine
Publisher	AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BI
Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Journal of biological chemistry	Country	United States of America
Author/Editor	AMERICAN SOCIETY FOR BIOCHEMISTRY & MOLECULAR BIOL	Rightsholder	American Soc for Biochemistry & Molecular Biology
Date	01/01/1905	Publication Type	e-Journal
Language	English	URL	http://www.jbc.org/

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	Distribution	Worldwide
Number of charts / graphs / tables / figures requested	2	Translation	Original language of publication
Format (select all that apply)	Electronic	Copies for the disabled?	No
Who will republish the content?	Author of requested content	Minor editing privileges?	No
Duration of Use	Life of current edition	Incidental promotional use?	No
Lifetime Unit Quantity	Up to 999	Currency	USD
Rights Requested	Main product		

NEW WORK DETAILS

Title	Deciphering the Structural Effects of Activating EGFR Somatic Mutations with Molecular Dynamics Simulation	Publisher imprint	N/A
		Expected publication date	2020-09-30
Author	Mahlet Z. Tamirat, Kari J. Kurppa, Klaus Elenius, Mark S. Johnson	Expected size (number of pages)	8
		Standard identifier	N/A
Publication	Journal of Visualized Experiments		
Publisher	MyJove Corp.		

ADDITIONAL DETAILS

Order reference number	N/A

The requesting person / organization to appear on the license	Mahlet Z. Tamirat

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure S8B and Figure S8C	**Title of the article/chapter the portion is from**	An unbiased in vitro screen for activating epidermal growth factor receptor mutations
Editor of portion(s)	Eric R. Fearon	**Author of portion(s)**	AMERICAN SOCIETY FOR BIOCHEMISTRY & MOLECULAR BIOL
Volume of serial or monograph	294	**Issue, if republishing an article from a serial**	N/A
Page or page range of portion	Page 8 of supplementary material	**Publication date of portion**	2019-04-05

PUBLISHER TERMS AND CONDITIONS

Journal of Biological Chemistry permissions policies: • If you are an author of the content for which you are seeking permission, or if you are not an author but are requesting permission to copy, distribute, transmit and adapt the work for noncommercial purposes (e.g. reproduction of a figure for educational purposes such as schoolwork, or appending a reprinted article to a PhD dissertation), you do not need to seek permission using the options listed below, as long as any reuse includes the credit line in the reuse policies listed above. • Parties who are not authors on the article who wish to reuse content for commercial purposes such as reproducing a figure in a book, journal, or coursepack published by a commercial publisher, do need permission and should request permission by completing the form below. For more information please see Journal of Biological Chemistry: http://www.jbc.org/site/misc/edpolicy.xhtml#copyright

CCC Republication Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

    3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

    3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically

set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER;

USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

    8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

    8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here: https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

    8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

    8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

    8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

v 1.1

Supplemental File (Screenshot 1)

Click here to access/download;Supplemental File (Figures, Permissions, etc.);JoVE61125_screenfile1.tif

Supplemental File (Screenshot 2)

Click here to access/download;Supplemental File (Figures, Permissions, etc.);JoVE61125_screenfile2.tif