

Molecular dynamics of the asymmetric dimers of EGFR: Simulations on the active and inactive conformations of the kinase domain



Napat Songtawee^a, David R. Bevan^b, Kiattawee Choowongkomon^{a,c,*}

^a Department of Biochemistry, Faculty of Science, Kasetsart University, Chatuchak, Bangkok 10900, Thailand

^b Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, United States

^c Center for Advanced Studies in Tropical Natural Resources, National Research University-Kasetsart University, Kasetsart University, Chatuchak, Bangkok 10900, Thailand

ARTICLE INFO

Article history:

Accepted 5 March 2015

Available online 14 March 2015

Keywords:

EGFR
Kinase domain
Asymmetric dimers
Molecular dynamics
Essential dynamics
Intermolecular interactions

ABSTRACT

Abnormal activation of EGFR is associated with human cancer, and thus it is a key target for inhibition in cancer therapy. There is evidence suggesting that the activation mechanism of EGFR is based upon the formation of the asymmetric dimer of the kinase domains. Here, we performed MD simulations on the asymmetric dimer for both active and inactive conformations of EGFR kinase domain to investigate flexibility and intrinsic motions of the proteins. Simulations of the active conformation showed that the formation of the asymmetric dimer changes the dynamics of EGFR kinase domain by suppressing fluctuation of the protein and altering the direction of motion of the protein. In contrast, the asymmetric dimerization of the inactive conformation does not alter the overall fluctuation of the kinase domain and does not initiate destabilizing of the inactive structure. We also investigated the intermolecular interactions in the EGFR asymmetric dimers and found that in the active conformation the interactions are dominated by loop-loop contacts rather than those from the helix-helix interactions. In contrast, helix-helix interaction seemed to be more significant for the inactive kinase structure. This work helps us to better understand the conformational flexibility and dynamics of the EGFR kinase domain, as well as provides information that may be useful to develop newer classes of inhibitors that can block allosteric sites rather than the more traditional catalytic site.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Epidermal growth factor receptor (EGFR) is a transmembrane glycoprotein receptor tyrosine kinase that plays crucial roles in cellular responses to environments mediated by its natural ligands and growth factors. Aberrant activation or mutations of EGFR can promote tumor cell growth and survival of non-small cell lung cancer (NSCLC) as well as other types of cancer, e.g. liver, stomach, colorectal, breast and esophageal cancer [1,2]. Much effort on cancer research has been focused on designing drugs that inhibit EGFR activity [3]. EGFR can be divided into three structural domains: (i) an extracellular ligand-binding domain, (ii) a single transmembrane segment, and (iii) an intracellular tyrosine kinase domain flanked by a juxtamembrane (JM) and an extended C-terminal tail (CT) segments [4]. It is known that ligand-induced dimerization of the extracellular domain results in the activation of the kinase

domain, catalyzing the γ -phosphate transfer reaction from ATP to the conserved tyrosine residue at its C-terminal tail [5,6]. Since autophosphorylations at the C-terminal tail initiates several downstream signaling pathways [6]. As a result, EGFR kinase activity requires careful regulation.

Numerous crystal structures of the human EGFR kinase domain revealed the characteristics common to almost all protein kinases in that it is comprised of two lobes: (i) a smaller N-terminal lobe consisting mainly of β -strands with a single large α C helix; and (ii) a larger C-terminal lobe, which is almost exclusively α -helical with a long activation (A) loop. The ATP-binding site is located at the hinge region between two lobes, and thus motions of the lobes can determine dynamics of the active site of the enzyme (Fig. 1). Among the known crystal structures, it has been observed that the EGFR kinase domain exhibits at least two fundamentally different conformations. One adopts a catalytically active conformation in the presence of the anti-cancer drug erlotinib [7], and the other form is in an autoinhibited/inactive conformation bound to the drug lapatinib [8]. The pronounced structural features of the active conformation compared to that in the inactive one include: (i) orientation of the α C helix closer to the active site, resulting in the salt bridge formation of E738 on the helix and K721 on the

* Corresponding author at: Department of Biochemistry, Faculty of Science, Kasetsart University, Chatuchak, Bangkok 10900, Thailand. Tel.: +66 2 5625555x2051; fax: +66 2 5614627.

E-mail address: fsciktc@ku.ac.th (K. Choowongkomon).

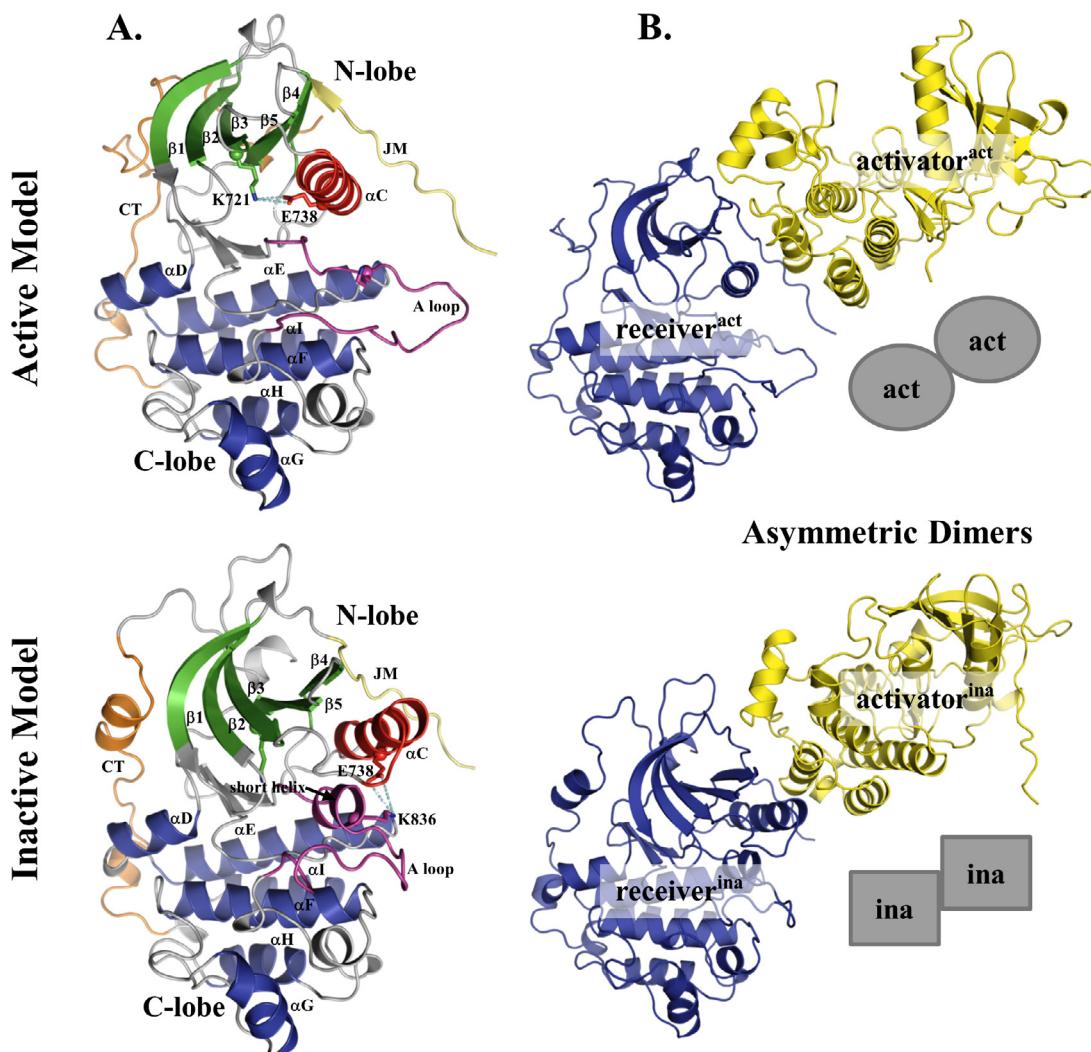


Fig. 1. Structural models for the monomers (A) and asymmetric dimers (B) of active and inactive EGFR kinase domain. The monomeric kinase domain with JM and CT segments are colored according to their structural regions; N-lobe β -strands, green; C-lobe α -helices, blue; α C-helix, red; activation loop, magenta. The K721–E738 and E738–K836 salt bridges are indicated in cyan dashed lines.

β 5-strand, which is essential for stabilizing the proper position of the ATP phosphates, and (ii) extension of the A loop away from the active site, which allows accessibility of a peptide substrate to the binding site. By contrast, the outward displacement of the α C helix coupled to the closed conformation of the A loop observed in the inactive conformation sterically hinder the substrate binding (Fig. 1A).

Knowledge of the autoinhibition, activation and feedback inhibition mechanisms of the EGFR kinase domain has been significantly advanced in recent years. It is thought that EGFR kinase domain is intrinsically autoinhibited, and its intermolecular interactions promote the kinase activation. Recent studies have demonstrated that activation of EGFR kinase results from the asymmetric dimer formation, in which the C-lobe of one kinase (so-called activator) interacts with the N-lobe of the other (so-called receiver), eventually leading to activation of the latter which is in the active-like conformation [9–11]. When the asymmetric dimer interface is disrupted, either by point mutations at the dimer interface, e.g. V924R of the activator and I682Q of the receiver, or by binding of cytosolic protein Mig6, that there is an intracellular negative feedback EGFR inhibitor, the kinase domain possess a low level of basal activity toward substrate peptides [9,10]. The crystal structure of Mig6-bound EGFR kinase domain

also demonstrated that the kinase structure adopts the autoinhibited/inactive conformation and Mig6 binds to the C-lobe of the kinase domain in which it contributes to EGFR inhibition through blocking the asymmetric dimer interface [10], emphasizing the important role of the asymmetric dimer formation in EGFR activation as well as a particularly intriguing perspective for the receptor inhibition.

A noticeable observation on the role of the asymmetric dimer in EGFR activation is that one crystal structure of EGFR and HER4 with their well-resolved juxtamembrane (JM) segment exhibits the inactive conformation but still possess the ability to form an asymmetric dimer that resembles the dimer of the active conformation [12,13]. Crystallographic analyses demonstrated that the C-terminal portion of the JM segment provides an additional interaction for the receiver to interact with the C-lobe of the activator in the context of the asymmetric dimer, and such interaction plays a critical role in stabilizing the dimerization. It could be hypothesized that the capability of the C-terminal portion of JM segment to dimerize kinase molecules even though they are in the inactive conformation might be essential for an early step of the receptor activation when two kinase molecules with inactive conformation are brought in the vicinity of each other and then interact in the context of the asymmetric dimer [14].

Protein conformations are actually expressed as a solution ensemble that is more dynamic rather than the single static snapshot of the actual state provided by crystallographic experiments, and thus it is limiting a broader understanding of the protein dynamics. Moreover, crystal contacts may stabilize non-native conformational states and thus introduce artifacts that may lead to inaccurate interpretations. Molecular dynamics (MD) simulation are an ideal method for analyzing conformational flexibility of protein molecules. The technique has been successfully used to study several biological systems, e.g. protein–protein, protein–lipid and protein–ligand interactions at the atomic-level detail. Several MD simulations have previously been applied for studying EGFR kinase domain and other family members in several contexts including the enzyme–drug interactions [15–19] and protein conformational studies [20–26] as well as the membrane-bound full length receptor [27]. In this study, we performed 100-ns MD simulations followed by principal component analysis (PCA) on the asymmetric dimer of both active and inactive conformations of EGFR kinase domain to investigate the structural flexibility and intrinsic motions of each kinase molecule in the dimers. We also investigated intermolecular interactions at the dimer interfaces and compared such interactions to those of the Mig6–EGFR complex observed in the crystal structure. Information gained from our simulations can give insights into better understanding in conformational flexibility and dynamics of EGFR in the context of the asymmetric dimer, as well as provide useful information for further development of a new class of inhibitors that will block the allosteric sites rather than targeting the catalytic sites of the enzyme.

2. Computational methods

2.1. Structural preparation

The atomic coordinates of human EGFR kinase domain determined for both active and inactive conformations were obtained from the Protein Data Bank (PDB). Missing and disordered residues in the retrieved structures were examined using Swiss-PDB-Viewer v4.1 [28]. The program was also used to reconstruct missing atoms of side chains and remove alternative conformations of some residues. The PDB codes that were used to create a crystallographic ensemble of human EGFR kinase domain include the active conformation as follows: 1M14, 1M17, 2GS6, 3VJO, 2GS2, 2ITW, 2ITX, 2J5F, 2J5E, 2J6M, 2ITY, 2ITN, 2ITV, 4I23 and inactive conformations: 3POZ, 2RGP, 3BEL, 1XKK, 2RF9, 3GOP, 2GS7, 3GT8, 3LZB, 4I22, 4I24. All of which the residues before L683 and after L955 were removed to reduce complexities. The residues from L683 to L955 are responsible for the actual kinase domain of human EGFR. Missing residues in those retrieved crystal structures were subsequently constructed using MODELLER 9v8 [29] offered by the ModLoop web server [30].

The MODELLER program was also used to construct the EGFR models in our study which include the C-terminal portion of JM (from G672 to I682), actual kinase domain (from L683 to L955), and extended CT segments (from V956 to I994). To construct the model of the active conformation, the PDB codes 1M17 and 2GS6 were used as the templates whereas 1XKK, 2GS7 and 2R4B (the crystal structure of human HER4) were used to construct that of the inactive conformation. Since evidence suggested that the JM and CT segments are essential for protein dynamics of the kinase domain, and biophysical studies showed that the deletion of these two segments significantly reduce the ability of EGFR kinase to form an asymmetric dimer [11,12,22]. For this reason, the 2GS6, 1XKK and 2GS7 were used to spatially restrain an orientation of the CT segments that should interact with both the C-lobe (via V956–P968) and the N-lobe (via E980–I994) of the kinase domain. Note that

the extended CT segment in the 1M17 distantly extends away from and loses any interaction with the kinase domain [7]. On the other hand, 2R4B was used to model the residues G672–N676 of the JM segment and residues E848–G850 of the A loop in the inactive conformation of EGFR kinase domain since those residues are not well resolved in all of the inactive EGFR structures. The stereochemical qualities of the resultant models were assessed using PROCHECK v3.5 [31].

A model of the asymmetric dimer of EGFR kinase domain for both active and inactive conformations was constructed according to the study of Zhang et al. [9] using the PDB codes 2GS6 and 2R4B as templates, respectively. A model of two EGFR monomers was created by removing the activator molecules from the dimeric models, and therefore they correspond to the receiver in the asymmetric dimer. We note that each molecule in our models is referred to here as the receiver^{act}, activator^{act}, monomer^{act} and the receiver^{ina}, activator^{ina}, monomer^{ina} for simulation of the active and inactive models, respectively. Prior to an MD simulation, the N- and C-termini of each amino acid chain were capped with the acetyl and amino groups, respectively, to give the neutral charged ends and mimic the full-length protein in which additional amino acids would be present. The GROMOS96-53A6 force field [32] was applied to all protein structures in simulated systems and the ionization state of amino acid residues was set according to the standard protocol. The SPC water model [33] was used for the solvent.

2.2. Molecular dynamics simulation

MD simulations for all EGFR models were carried out using explicit-solvent periodic boundary conditions using GROMACS v4.5 [34,35]. Each model was solvated in a rectangular box keeping a distance of 1.2 nm between the solutes and the sides of the solvent box. Each of the solvated systems was neutralized by adding enough sodium and chloride ions to give a concentration of 100 mM. All of the solvated systems were then energy-minimized using the steepest descent algorithm either until the maximum force was lower than 1000 kJ/mol/nm on any atom or until additional steps result in a potential energy change of less than 1 kJ/mol. All simulation systems were first equilibrated in three phases with position restraint (the force constant of 1000 kJ/mol/nm) applied to all heavy atoms of the proteins, allowing H atoms, and solvent and ion molecules to freely move. The first step was conducted under NVT conditions at 300 K of temperature using the modified Berendsen (velocity rescaling) thermostat [36,37]. The following step was conducted under NPT conditions at 1 bar of pressure using Parrinello–Rahman barostat [38,39]. In the last step, Nosé–Hoover thermostat was applied because it generates a correct kinetic ensemble and allows for fluctuations that produce more natural dynamics [40,41]. After equilibration, any position restraint of all heavy atoms in proteins was removed. The full dynamics production was subsequently performed under the NPT condition at 300 K and 1 bar of the system.

For all dynamics procedures, hydrogen bond lengths were constrained using LINCS algorithm which allows for a 2.0 fs-time step [42]. A cut-off distance for the short-range neighbor list was set to 1.0 and 1.4 nm for the electrostatic and van der Waals interactions, respectively. Long-range electrostatic interactions are approximated using the Particle Mesh Ewald (PME) method [43,44]. The atomic coordinates were recorded every 10 ps for the data collection.

2.3. Trajectory analysis

MD trajectories were analyzed using utilities in the GROMACS suite [45] that includes the backbone root mean square deviation, root mean square fluctuation, atomic distances and

residue contacts. Changes in secondary structure content during the simulations were assessed using DSSP [46]. An analysis of intermolecular interaction in the dimeric models (receiver vs. activator) includes interaction energies, hydrogen bonds and hydrophobic contacts. Hydrogen bonding is defined following the default values that are (i) donor–acceptor distance ≤ 0.35 nm, and (ii) acceptor–donor–hydrogen angle $\leq 30^\circ$. The Protein Interaction Calculator (PIC) web server [47] was used to determine hydrophobic interactions, by which each of the 100 dimeric structures extracted from the last 50 ns of simulation time was submitted one-by-one to the server. Hydrophobic interactions are identified using a distance cut-off of 0.5 nm between non-polar groups in the non-polar side chains of amino acid residues. It should be noted that no aromatic–aromatic, aromatic–sulfur and cation–π interaction was observed at the interface of both EGFR dimeric models. The Grace v5.1 [48] was used to plot the 2D data and the 3D images were created and rendered with PyMOL v1.3 [49].

2.4. Principal component analysis

The protein ensembles include (i) a set of twenty-five crystal structures of human EGFR kinase domain (as mentioned above) and (ii) six separate MD sets that differ in terms of both molecule type (monomer, receiver and activator) and protein conformation (active and inactive) of 5000 structures, which were extracted from the last 50 ns of simulation. Each protein ensemble was separately subjected to performing a principal component analysis (PCA) using the GROMACS utilities [45]. Mass-weighted covariance matrices were constructed based on backbone fluctuations from their average position after least-squares fitting to remove overall rotational and translational motions. Diagonalization of each covariance matrix then produced a set of eigenvectors and corresponding eigenvalues, which represent the direction and amplitude of a motion, respectively. Eigenvectors were subsequently ranked by decreasing eigenvalues, and such that the first eigenvector, which is also referred to as the first principal component (PC1), represents the largest contribution to the total fluctuation of the protein [50,51]. A motion represented by each PC was visualized by projecting the structures from ensembles (or trajectories) onto each eigenvector of interest and then transforming them back into atomic coordinates. Two extreme projections along that eigenvector were then be interpolated to create an illustration using PyMOL v1.3 [49].

3. Results and discussion

We used a homology-based approach to construct 3D models for both active and inactive conformations of the entire EGFR kinase with its flanked JM and extended CT segments (the residues G672 to I994). Our homology models were stereochemically evaluated according to Ramachandran plots. No residue is located in disallowed region of the plot for both active and inactive models, and more than 90% of the residues are in most favored regions (data not shown) suggesting the reliability of our resultant models. The structural model for the EGFR asymmetric dimer of both active and inactive conformations were then constructed according to the study of Zhang et al. [9] (Fig. 1B). It should be noted that although the relative orientation between receiver and activator in both dimeric models seems identical, the interface area of the active model was slightly larger than that of the inactive one (i.e. 10.4 vs. 9.6 nm², respectively) as estimated from PDBePISA server [52].

The 100-ns MD simulations were carried out on the asymmetric dimers as well as their separate monomer to investigate how dimerization affects the flexibility and dynamics of the EGFR kinase

domain. The backbone-atom root mean square deviation (RMSD) was first measured to observe stability of the simulations and the degree of similarity between simulated and reference structures over the simulation time (Fig. 2). With respect to their initial structure, the backbone RMSD values increased more than 0.3 nm and then fluctuated around 0.3–0.4 nm; nevertheless, the values of the receiver^{act} molecule fluctuated around 0.2–0.3 nm. The RMSD of the receiver in both of the dimeric models remained stable after 50 ns when compared to those of the monomer and activator models which slightly increased. In the following sections: we discuss the results of the MD simulations as follows: (i) the difference in conformational flexibility among the receiver, activator and monomeric molecules observed from the simulations, (ii) essential dynamics of the MD trajectories with and without respect to the crystallographic ensembles, and (iii) intermolecular interactions between receiver and activator from the simulations of the asymmetric dimers.

3.1. Conformational flexibility of EGFR kinase domains

To give insight into conformational flexibility with respect to each structural region of the EGFR kinase domain, we then measured a backbone root mean square fluctuation (RMSF) as a function of the residue number (Fig. 3). The reference structures were averaged over the last 50 ns simulation time. The RMSF was also qualitatively illustrated in the backbone traces and tubes of the 50 snapshots from the simulations (supplementary data S1 and S2, respectively).

3.1.1. Flexibility of the active conformation

As shown in Fig. 3A, in the active models the RMSF of the receiver were reduced in the N-lobe and CT segment when compared to those of the monomer and activator, in particular the loops between β1–β2 and β3–αC and the CT residues D970–D990. This result indicated that the dimerization process suppresses the conformational flexibility of the N-lobe of the kinase domain as well as of the CT segment. The reduced fluctuation in the β3–αC loop is consistent with the other published work in which this region in the EGFR monomer is intrinsically disordered and becomes ordered when dimerization [24]. The β1–β2 loop is important for organizing the ATP-binding since it makes contacts with the γ-phosphate of ATP. Slightly lower RMSF values were also observed in the hinge region of the receiver (the loop following β5 strand), which makes contacts with the ATP-adenine ring. Furthermore, the K721–E738 salt bridge seemed to be more stable in the receiver than in the monomer and activator (supplementary data S3), and thus consistent with the others [20,24,53]. The salt bridge allows the side chain of K721 to bind to the α- and β-phosphates of ATP, and reorients the αC helix closer to the ATP binding site. These results indicated that the asymmetric dimerization helps stabilizing the active conformation, presumably involving the organization of the ATP-binding site.

The significant difference in the RMSF values was not observed in the JM segment (Fig. 3A). In the asymmetric dimer, this segment of the receiver obviously interacts with the C-lobe of the activator. The RMSF values of the receiver were slightly lower than those of the monomer and activator in contrast to the previous work, in which more fluctuations of the JM segment were observed the monomeric kinase [22]. It might be because in the monomer and activator the JM segment could make contacts with the A loop rather than left flexibly (Supplementary data S2). The reduced fluctuation of the receiver was also observed in various regions of the C-lobe of the kinase domain, for examples the distal part of the A loop, αG–αH loop, αI helix and the loop between αI and CT segment. These results indicated that the asymmetric dimerization could also alter the flexibility in regions that are distant from the dimerization interface.

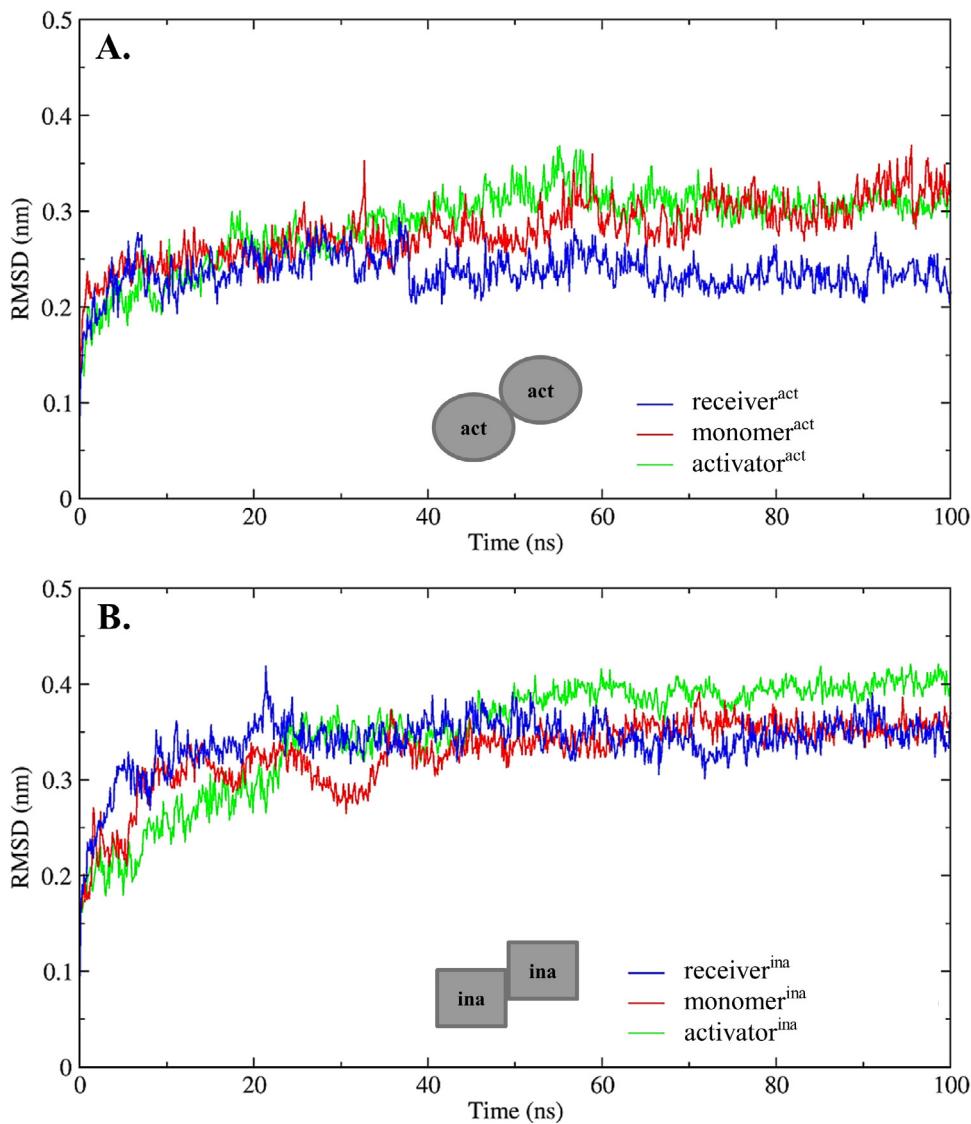


Fig. 2. Backbone RMSD plots as a function of time from the active (A) and inactive (B) models with respect to their initial structure.

3.1.2. Flexibility of the inactive conformation

In contrast to the active conformation, the flexibility of the receiver and activator in the inactive model was similar. The dimerization did not alter fluctuations in the N-lobe of the receiver; except for the α C helix and β 4– β 5 loop (Fig. 3B), which are located in the dimer interface. Interestingly, the RMSF of the β 2– β 3 loop in the monomer was significantly lower than those of both the receiver and the activator. This loop segment is not a part of the dimer interface and is disordered in at least four crystal structures of the EGFR kinase domain. The β 1– β 2 loop also showed more fluctuations in all three inactive molecules. In our previous study, the loop exhibited less fluctuation in the inactive structure, but became more flexible when binding to the drug lapatinib [25]. In addition to the N-lobe, the simulations also showed that the distal portion of the A loop (residues E844–K851) exhibited lower fluctuations in the receiver than that of activator and monomer, even slightly (Fig. 3B). This portion was disordered in all crystal structures of the inactive conformation of EGFR. One possible explanation for this is that this portion in the receiver seemed to make contacts with the C-lobe (the loops preceding α F and α G helix) whereas it was solvent-exposed in the activator (Supplementary data S2).

Since it has been thought that the asymmetric dimer of the inactive conformation might be an early step of the activation of EGFR kinase domain [14], we therefore followed the RMSD values of (i) the active models with respect to the inactive structures, and (ii) the inactive models with respect to the active structure during the 100-ns simulation time (supplementary data S4). A global switch from the active to inactive conformations and vice versa was not observed. In the inactive models, the RMSD at the end of simulation with respect to the active structure of the receiver^{ina}, which would be expected to close to the active conformation when formed the asymmetric dimer, is 0.66 nm. Neither active nor inactive monomer conformation closes to the inactive and active conformations at the end of simulation (RMSD values of 0.57 nm and 0.62 nm, respectively). These results indicated that the 100-ns time scale was still too short to observe either the transition between two distinct conformations or to sample an overlapping intermediate structure. A recent MD study has demonstrated that no active-to-inactive (and vice versa) conformational transition of the wild type monomeric EGFR kinase domain was observed during the 200-ns simulation time [23]. Nevertheless, another study of the other has demonstrated that an intermediate state between

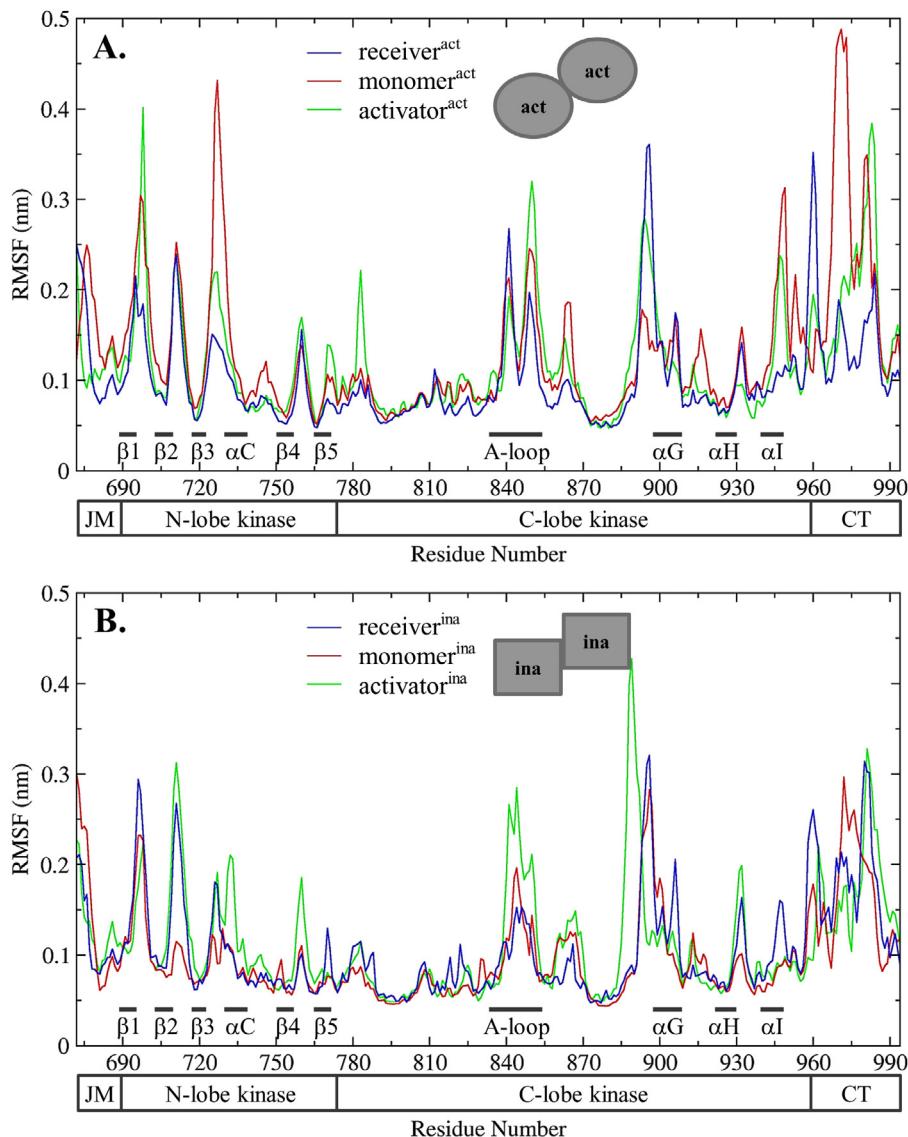


Fig. 3. Backbone RMSF plots as a function of residues from the active (A) and inactive (B) models. The various structural regions the EGFR kinase domain and the JM and CT segments are indicated. The last 50 ns in each MD trajectory were taken to calculate in all cases.

active and inactive conformations of EGFR kinase domain can be captured in a microsecond time scale MD simulation [24].

We were subsequently interested in examining the destabilization effects on the structure of the receiver^{ina} molecule upon the dimerization process. There are several factors stabilizing the inactive state of EGFR, e.g. a short helix of residues L834–L838 located on the A loop that makes hydrophobic contacts with the α C helix. The interaction results in the obstruction of the K721–E738 salt bridge, but the formation of the E738–K836 salt bridge instead [9,20]. Destabilization of this hydrophobic interaction has been thought to be one of the mechanisms for EGFR kinase activation, e.g. L834R mutation [20]. In addition, an α helix formed at the CT residues F973 to M978 partially blocks the ATP-binding site of the 1XKK structure [8,22]. In our inactive models, we found that the helix of L834 to L838 of the receiver remained stable, in contrast to the monomer of which the helix unwound (supplementary data S5). The results were also in accordance with the critical E738–K836 salt bridge that is more stable in the receiver than observed in the monomer (Supplementary data S3). In combination with the results of RMSD as mentioned above it would be suggested that the asymmetric dimerization of EGFR, at least 100-ns

time scale simulations could not even disrupt interactions stabilizing the inactive conformation, in which it has been thought to be an early step required for conformational changes to the activated kinase.

3.2. Essential dynamics of EGFR kinase domains

Principal component analysis (PCA) can be used to identify and compare the principal modes of motion of the EGFR kinase domain using either crystallographic or MD ensembles. The first few principal components (so-called essential dynamics, ED) describe the large-scale collective motions that represent the functionally important global movements, in contrast to the remaining principal components that capture the smaller, localized and often insignificant fluctuations that should not influence protein function [50,51]. In this study, PCA was applied to investigate the ability of our MD simulations to sample the experimentally known motions, which were identified from the crystallographic experimental data set of EGFR kinase domain. In addition, the quantity and direction represented by the first principal component (PC1), which

was individually generated from the MD simulations, were visually compared.

3.2.1. Principal components of the protein ensemble from crystal structures

We established the conformational changes captured in the crystallographic ensembles as an experimental reference. The crystallographic ensemble is a set of twenty-five crystal structures of

EGFR kinase domain (i.e. 12 wild type/active, 2 mutant/active, 5 wild type/inactive and 6 mutant/inactive structures), and can produce a common set of PCs for the active and inactive conformations (supplementary data S6A), and therefore making direct comparison possible among different protein models. This approach has been previously used on other proteins for which substantial crystallographic data exists [54–56]. A projection of the crystal structures onto the two-dimensional (2D) essential subspace defined by the

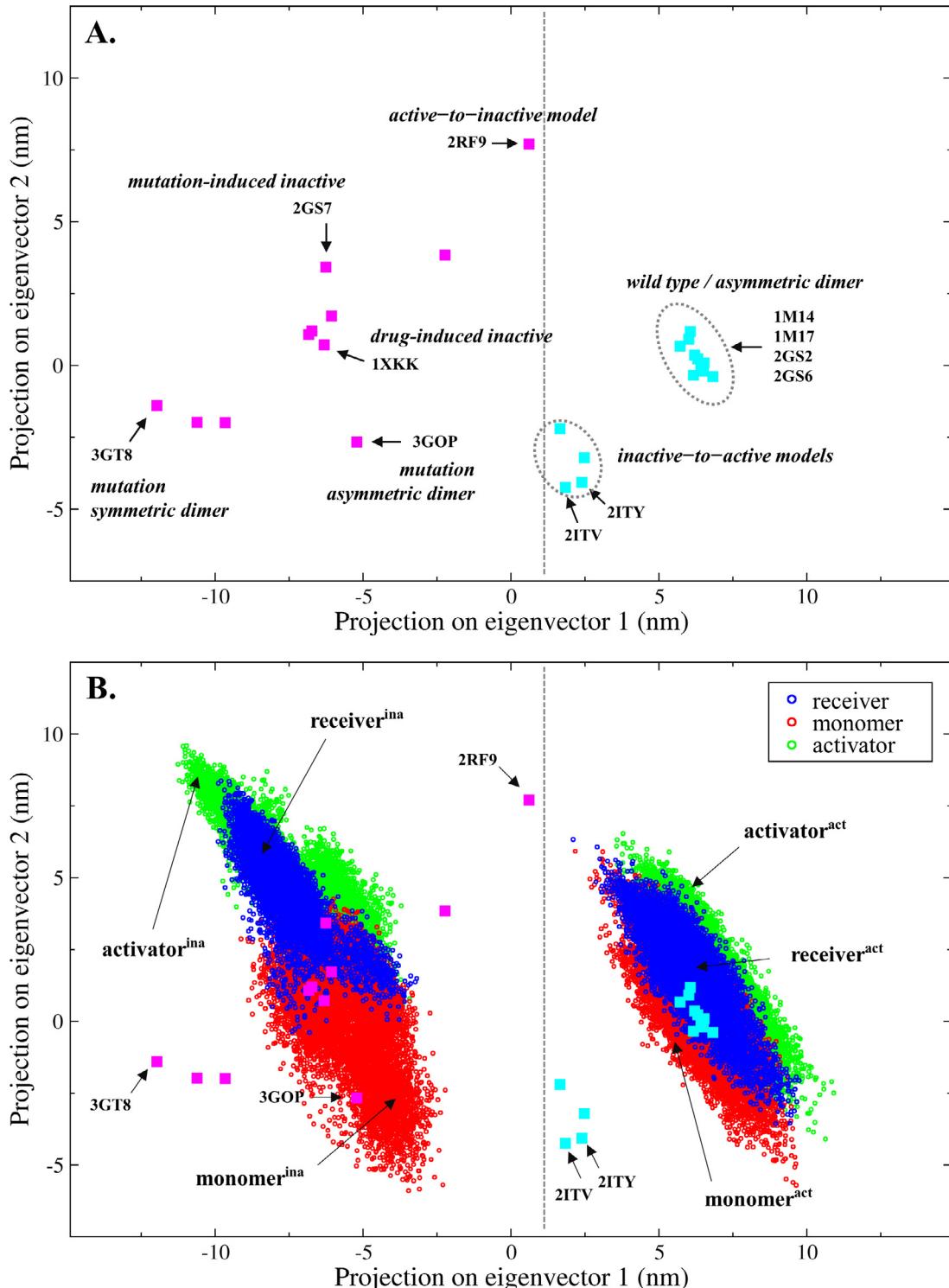


Fig. 4. Principal components (PC) identified from crystallographic ensemble of EGFR kinase domain. Crystal structures (A) and MD structures (B) were projected onto the plane defined by PC1 and PC2. Active and inactive conformations of the crystal structures are shown in cyan and magenta filled rectangles, respectively. Selected PDB codes (discussed in the text) as well as the areas for the simulated structures of the last 10 ns are indicated.

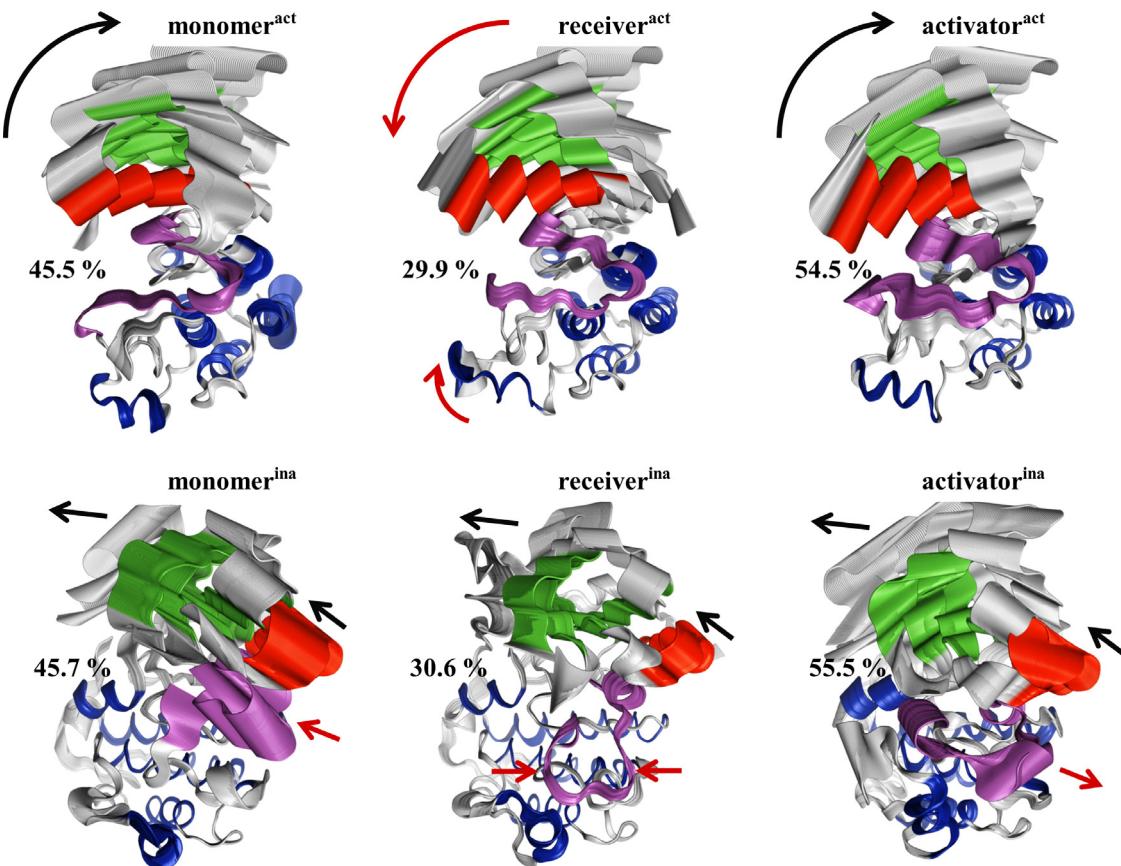


Fig. 5. Comparison of the domain motion described by PC1 from the simulations of the active and inactive models. The two extreme projections for PC1 of the monomer, receiver and activator are illustrated statically by showing all frames simultaneously as backbone trace representation. The structural regions are separately colored as follows: N-lobe β -strands, green; C-lobe α -helices, blue; α C-helix, red; activation loop, magenta. Arrows indicate the approximate direction of the motion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

first two principal components of the crystallographic ensemble was shown in Fig. 4A. The 2D projection could define the clusters that correlate well with two distinct conformations of EGFR kinase domain, in which the PC1 mode was able to differentiate between the active and inactive conformations with positive and negative values, respectively. Visualization of a protein motion was illustrated as multi-frames of the two extreme structures (supplementary S6B). The PC1, which represented approximately 71% of the total motion of the ensemble, involved the large movement of the N-lobe of the kinase domain, e.g. α C helix and almost β -strands, as well as extension of the A loop to be solvent-exposed. The functional significance of this motion presumably modulates the shape of the ATP binding site of EGFR by rearrangement of structural elements that are important for catalysis. The PC2, which comprised only 11% of the total motion of the ensemble, also involved the movement of the N-lobe of the kinase domain, but in the opposite direction observed in PC1. This N-lobe displacement might be related to the inactive-to-active conformational transition of EGFR kinase domain.

We then projected the protein ensemble obtained from all MD trajectories onto the 2D essential subspace of the crystallographic ensemble to investigate conformational sampling of the MD simulations with respect to the experiments. As shown in Fig. 4B, the simulations of the active and inactive models were remarkably distinguished using the PC1 defined from the experiments. Almost the active models largely occupied the unique group of the wild type active EGFR kinase structures (cyan rectangles) and none of them was sampled in the region of the minority. Unlike the crystal structures of inactive EGFR kinase domain (magenta rectangles), the

cluster of the inactive models was less distributed throughout the negative value of PC1. Note that only the monomer^{ina} could sample a region of the asymmetric dimerized-inactive EGFR kinase domain (PDB code 3GOP) [12]. The structure has been thought to be an early step of the kinase activation. Furthermore, none of all inactive models was sampled in the area of the symmetric dimerized-autoinhibited/inactive EGFR kinase domain (PDB code 3GT8) [11]. From the results of PCA, it could be implied that the asymmetric dimer constructed in our EGFR inactive model is unlikely to be in either the autoinhibition or activation form, and thus it would suggest an alternative role of the inactive conformation of EGFR in the context of the asymmetric dimer. In addition, since there was no overlap between the MD ensembles from the active and inactive models, in accordance with the result of RMSD analysis, this indicated that the conformational transitions of the EGFR kinase domain were not observed, at least during the 100-ns simulation time.

No crystal structures have been observed in the intermediate conformational region between active and inactive clusters to date, e.g. the Mig6-bound (PDB code 2RF9) and the L834R oncogenic mutant (2ITN) structures. A recent study suggested that the Mig6-bound inactive conformation of EGFR kinase domain [10] is highly consistent with the intrinsically disordered intermediate sampled from the active to inactive transition of EGFR kinase domain [24]. On the other hand, the structure of the L834R oncogenic mutant protein may represent the inactive-to-active transition since there is evidence that induction of this mutation in the EGFR kinase domain promotes either destabilization of the inactive conformation [20] or the shift the active-inactive equilibrium to prefer the active state

Table 1

List of H-bonds observed at the dimer interface from the simulations of the active and inactive models. The last 50 ns of the trajectory were taken in all cases.

Activator		Receiver		% Occupancy
Residue	Location	Residue	Location	
<i>Active model</i>				
G906-NH	α G- α H loop	L758-CO	β 4 strand	93
G906-CO	α G- α H loop	L758-NH	β 4 strand	58
Q911-OE	α G- α H loop	L680-NH	JM segment	98
Q911-NE	α G- α H loop	A678-CO	JM segment	97
T916-OG	α G- α H loop	N676-CO	JM segment	72
T916-OC	α G- α H loop	A678-NH	JM segment	79
I917-NH	α H helix	A678-CO	JM segment	21
D918-OD1	α H helix	Y740-OH	α C helix	37
D918-OD2	α H helix	Y740-OH	α C helix	81
D930-OD1	α H- α l loop	T727-OG	β 3- α C loop	44
D930-OD2	α H- α l loop	T727-OG	β 3- α C loop	54
D950-OD1	α l-CT loop	G672-NH	JM segment	68
D950-OD1	α l-CT loop	E673-NH	JM segment	36
D950-OD2	α l-CT loop	G672-NH	JM segment	56
D950-OD2	α l-CT loop	E673-NH	JM segment	58
R953-CO	α l-CT loop	N676-NH	JM segment	18
<i>Inactive model</i>				
R908-N1	α G- α H loop	L758-CO	β 4 strand	17
R908-N2	α G- α H loop	L758-CO	β 4 strand	12
Q911-OE	α G- α H loop	L680-NH	JM segment	93
Q911-NE	α G- α H loop	A678-CO	JM segment	88
I917-NH	α H helix	N676-CO	JM segment	37
D918-OD1	α H helix	Y740-OH	α C helix	46
D918-OD2	α H helix	Y740-OH	α C helix	47
S933-OG	α H- α l loop	K733-NH	α C helix	16

[23]. In addition, it should be noted that the drug gefitinib-bound structure (PDB code 2ITY) was sampled in the same area of the L834R mutant structure, suggesting the similar effect of gefitinib does to the inactive conformation. Our previous simulations have shown that gefitinib could destabilize both active and inactive conformations of EGFR kinase domain [25].

3.2.2. Principal component of the active conformation from MD trajectories

PCA was also performed on the individual MD trajectories to characterize the dominant motion captured by the first principal component (PC1). In the active models, PC1 involved the large displacement in the N-lobe with respect to the C-lobe of the kinase domain in all three active models (Fig. 5, upper panel). Nevertheless, the direction of this motion in the receiver seemed opposite to that observed in the activator and monomer, and the contribution of the motion represented approximately 29.9%, 54.5% and 45.5% of the total motion in the protein, respectively. These results suggested that the asymmetric dimer can alter the global dynamics and suppress fluctuations of the active EGFR kinase domain. The interlobal motion of the active conformation is likely to be related to the opening and the closing of the binding cleft around the hinge region that is known to be important for ATP binding and catalysis in other protein kinases, e.g. PKA [57]. This opening and closing motion of the active conformation of EGFR was also identified in the another MD study in which the motion in the kinase domain is closely coupled with the conformational changes of the JM and CT segments [22], implying that the opening and closing motion in the kinase domain is the hallmark for protein kinases in the active conformational state, presumably for the tight regulation of a catalytic activity.

3.2.3. Principal component of the inactive conformation from MD trajectories

In all three inactive models, PC1 exhibited a similar motion that involves the rigid-body displacement of the N-lobe with respect to the C-lobe of the kinase domain (Fig. 5, lower panel). This motion

Table 2

List of hydrophobic interactions observed at the dimer interface from the simulations of the active and inactive models. The last 50 ns of the trajectory were taken in all cases.

Activator		Receiver		% Occupancy
Residue	Location	Residue	Location	
<i>Active model</i>				
P913	α G- α H loop	L679	JM segment	10
I917	α H helix	A678	JM segment	100
I917	α H helix	L680	JM segment	100
I917	α H helix	Y740	α C helix	100
I917	α H helix	A743	α C helix	100
Y920	α H helix	L680	JM segment	100
Y920	α H helix	I682	JM segment	88
M921	α H helix	L680	JM segment	76
M921	α H helix	L736	α C helix	100
M921	α H helix	A739	α C helix	44
M921	α H helix	Y740	α C helix	100
V924	α H helix	L680	JM segment	84
V924	α H helix	I682	JM segment	90
V924	α H helix	L736	α C helix	92
V924	α H helix	I756	β 4 strand	48
V924	α H helix	L758	β 4 strand	100
W927	α H helix	I682	JM segment	18
W927	α H helix	L758	β 4 strand	68
M928	α H- α l loop	L736	α C helix	98
M928	α H- α l loop	L758	β 4 strand	100
M928	α H- α l loop	V762	β 5 strand	94
M947	α I helix	Y740	α C helix	12
V956	CT tail	P675	JM segment	98
<i>Inactive model</i>				
P913	α G- α H loop	L679	JM segment	88
I917	α H helix	A678	JM segment	86
I917	α H helix	L680	JM segment	98
I917	α H helix	Y740	α C helix	100
I917	α H helix	A743	α C helix	96
Y920	α H helix	L680	JM segment	84
Y920	α H helix	I682	JM segment	94
M921	α H helix	L680	JM segment	24
M921	α H helix	L736	α C helix	38
M921	α H helix	Y740	α C helix	100
V924	α H helix	L680	JM segment	42
V924	α H helix	I682	JM segment	90
V924	α H helix	L736	α C helix	100
V924	α H helix	I756	β 4 strand	20
V924	α H helix	L758	β 4 strand	58
M928	α H- α l loop	L736	α C helix	96
M928	α H- α l loop	L758	β 4 strand	92
V956	CT tail	A674	JM segment	12
V956	CT tail	P675	JM segment	92

of the N-lobe looked roughly orthogonal to that of identified in the active models. The fluctuation of the motion is weaker in the receiver than in the activator and monomer, which represented approximately 30.6%, 55.5% and 45.7% of the total motion in the protein, respectively. These results suggested that, unlike the active conformation, the asymmetric dimer formation in the inactive conformation did not drastically change the dynamics of the N-lobe of EGFR kinase domain, but decreased the fluctuation of the protein instead. It would be noted that the motion of the A loop among three inactive models was found to be significantly different. The A loop of the receiver was less mobile. While the A loop of the activator moved toward the opposite direction with respect to the N-lobe, the motion of the N-lobe, α C helix and A loop in the monomer were highly correlated. Functional significance of PC1 observed from the inactive models is unclear. The motion of the N-lobe with the α C helix movement is common in three inactive models, and these suggested that it would not be specifically responsible for conformational changes of the receiver^{ina} molecule upon the dimerization. Moreover, the different motions in the A loop of the three inactive molecules might suggest the intrinsic flexibility of the loop, presumably as a result of various populations of the A

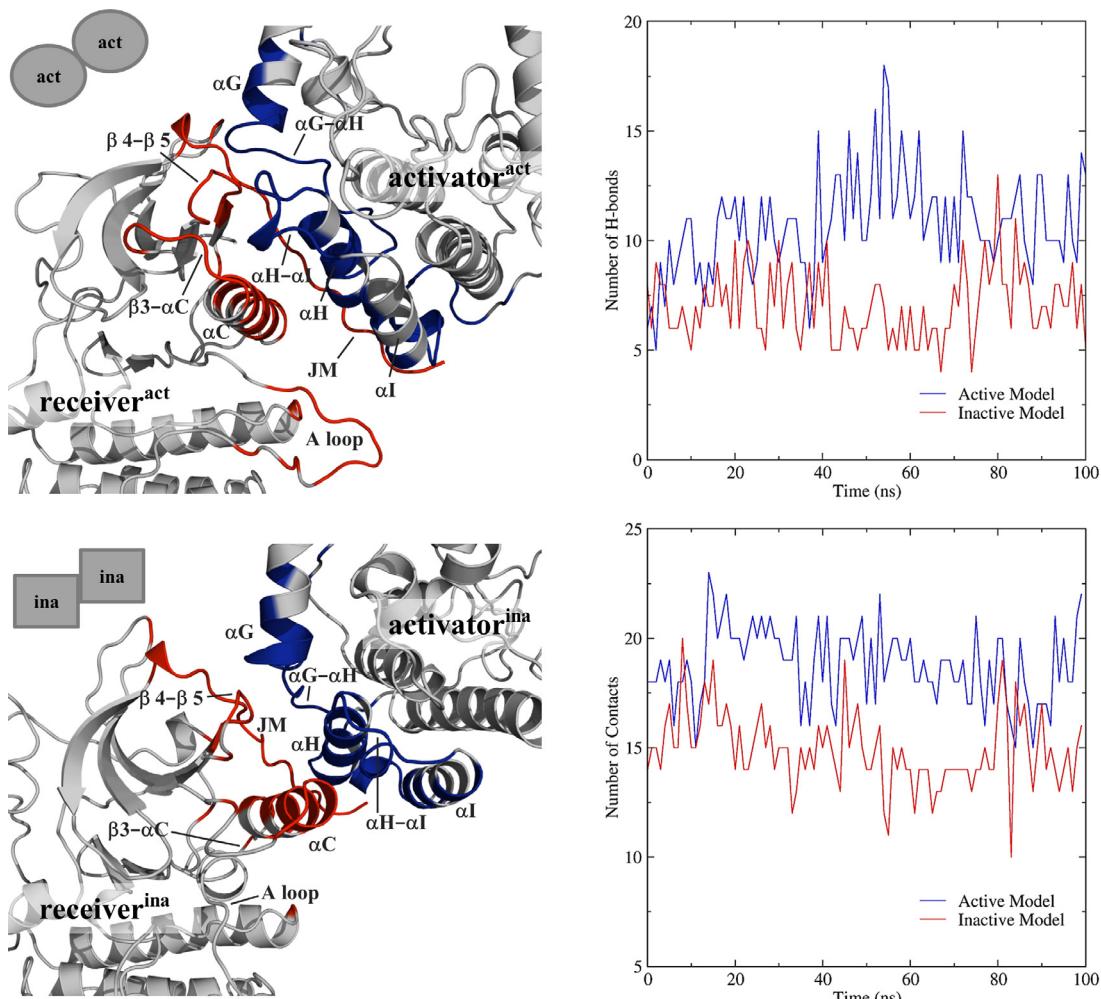


Fig. 6. Ribbon representation of the dimer interface (left images) and number of intermolecular contacts (H-bonds and hydrophobic interactions) as a function of time between the receiver and activator (right images) were observed from the simulations of the active and inactive models. Structural regions responsible for the dimer interface in the activator and receiver are colored in blue and red, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

loop conformation sampled in the MD simulations as well as the crystal structures, all of which are disordered in the distal portion of the A loop (residues E842–K851).

It has been stated that the formation of an asymmetric dimer is essential for the first step of EGFR activation when two inactive kinase molecules are brought into close proximity and interact with each other [14]. Our inactive model in the context of the asymmetric dimer would serve as a starting point to explore an early step of EGFR activation. However, the destabilizing effects on the inactive conformation upon the asymmetric dimerization could not be observed. These results suggested that the functional relevance of the asymmetric dimer formation of the inactive conformation of EGFR kinase domain would not be so evident; or it might play a subtle role involved in regulation rather than a direct mechanism on EGFR activation. Another mechanism involved in regulation of EGFR subpopulations could be suggested. A dynamic equilibrium has been proposed to exist between the two distinct EGFR kinase conformations, and it is maintained by their thermodynamic stabilities [58]. Biochemical experiments demonstrated that the isolated kinase has a low catalytic activity in solution, and when its local concentration was increased by attaching the isolated protein to the surface of lipid vesicles, the EGFR activity dramatically increased, indicating intermolecular interaction between isolated molecules [9]. It would be therefore possible

that the dimerization by the activator to the active monomeric kinase molecule may help stabilize the active conformation of such existing monomeric populations, eventually leading to shifting the equilibrium between active and inactive conformations, and such intermolecular interactions preferably EGFR kinase populations of the more stable active conformational states.

3.3. Intermolecular interactions at EGFR dimer interfaces

We next tried to study the nonbonded factors stabilizing the activator–receiver interactions in the asymmetric dimers of EGFR kinase domain. The possibility of the asymmetric dimer formation for the inactive conformations was not excluded in these analyses, and therefore this can be compared to interactions observed from the active conformation. Pairs of hydrogen (H-) bonds and hydrophobic interactions observed between the activator and receiver with respect to the domains in which each amino acid belongs to are listed in Tables 1 and 2. Each of the interaction pairs are maintained at least 10% of the time during the last 50 ns of the simulations. In addition, the potential energies that contribute to such interactions (referred to here as interaction energies) were calculated as the sum of the electrostatic and van der Waals (vdW) energies over the last 50 ns of simulation. The interaction energies

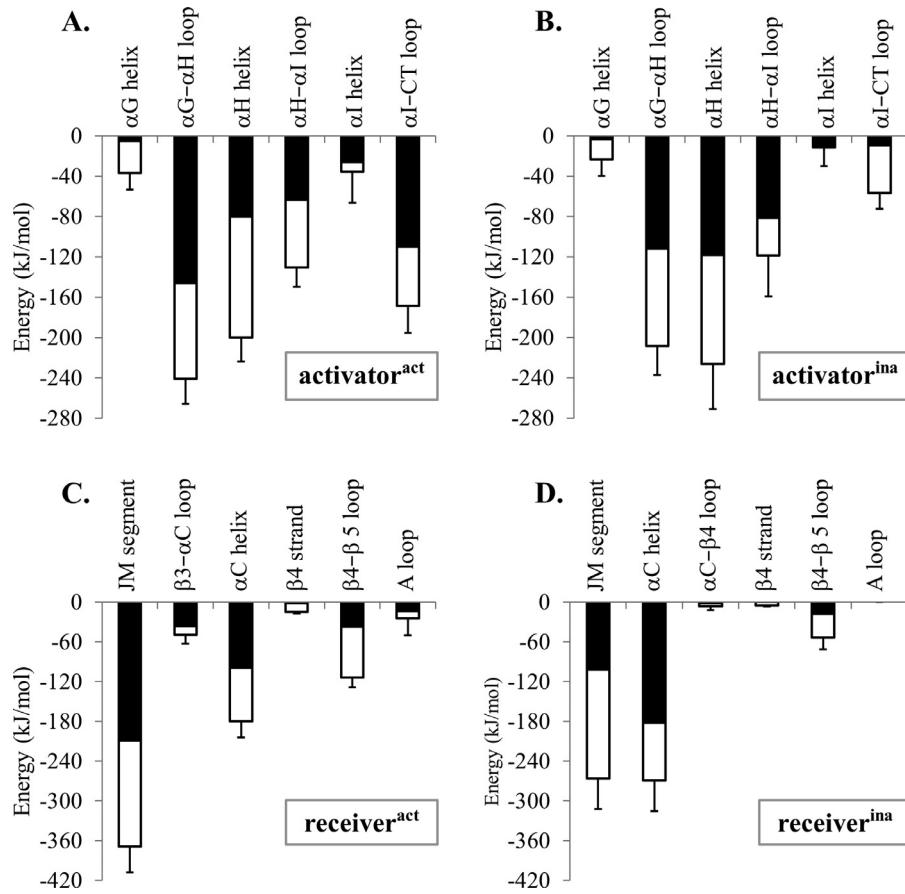


Fig. 7. Potential energies for interaction between the activator and receiver observed from the simulations of the active (A, C) and inactive (B, D) models. The interaction energies are represented as region-based decomposition of the C-lobe of the activator and the N-lobe of the receiver. The electrostatic and van der Waals contributions are shown in black and white bars, respectively. Error bars represent the standard deviation of the total interactions. The last 50 ns of each trajectory were taken in all cases.

with respect to various structural regions of EGFR kinase domain for each dimeric model could be quantitatively compared (Fig. 7).

The total interaction energy observed from the inactive model was estimated as being 4/5 weaker than that observed from the simulation of the active model. This result would be expected because the number of atomic contacts (i.e. H-bonds and hydrophobic interactions) between the receiver and activator of the active model were more than those observed in the inactive models throughout the simulations (Fig. 6, right panel). It should also be noted that vdW energies were found to be weaker than electrostatic interactions in both cases, presumably indicating the more contributions mediated by polar interactions (e.g. H-bonds). Analyzing the residue contacts within a distance of ≤ 0.5 nm during the last 50 ns revealed that the structural regions of the activator involving the dimer interface are similar between the two dimeric models, which range from the α_F - α_G loop, α_G helix, α_G - α_H loop, α_H helix, α_H - α_I loop, α_I helix, α_I -CT loop and the N-terminal portion of CT segment (Fig. 6, left panel). On the other hand, the structural regions of the receiver responsible for the dimer interfaces include the JM segment, α_C helix, α_C - β_4 loop, β_4 strand, β_4 - β_5 loop and β_5 strand (Fig. 6, left panel). As it is expected owing to the in-orientation of the α_C helix in the active conformation, the loop preceding α_C helix (β_3 - α_C loop) of the receiver could make contacts with the α_H - α_I loop of the activator. Several polar residues located on the extended portion of the A loop of the receiver^{act} could also make interactions with the α_I -CT loop of the activator. The A loop of the receiver^{ina} was found to move down to make contacts with the C-lobe surface of the same molecule instead.

3.3.1. Dimer interface of the active conformation

In the active model, it can be seen that the loop between α_G and α_H (α_G - α_H loop) of the activator exhibited the largest negative interaction energies (Fig. 7A). The loop made interactions with the JM segment and β_4 strand of the receiver. Several H-bonds contributing such interactions include (i) the main chain-main chain H-bond of G906-L758, (ii) side chain-main chain H-bond of Q911-L680 and Q911-A678, and T916-N676 (Table 1). On the other hand, several residues in the α_H helix of the activator (I917, Y920, M921, V924 and W927), as well as M928 on the α_H - α_I loop and V956 on the CT formed the hydrophobic interactions with a number of nonpolar residues on the JM segment, α_C helix and β_4 strand of the receiver (Table 2). Such hydrophobic contributions are consistent with experiments, which demonstrated that those residues are critical for the asymmetric dimer formation, and their mutations abolished the EGFR activity in cells [9,11]. The large negative interaction energies of the JM segment of the receiver also indicated the importance of this segment in the formation of the asymmetric dimer (Fig. 7B), in accord with several experiments [9–12].

From the crystal structures of active EGFR kinase domain (the PDB codes 2GS2 and 2GS6) [9], it can be seen that the hydrophobic residues located on the JM segment, α_C helix, β_4 and β_5 strand (listed in Table 2) form a large hydrophobic surface area. The so-called hydrophobic patch is buried in the inactive structure, but becomes solvent-exposed in the active conformation. It has been stated that the asymmetric dimer stabilizes the active conformation by compensation for the thermodynamic penalty associated with the solvent exposure of the patch in the active conformation

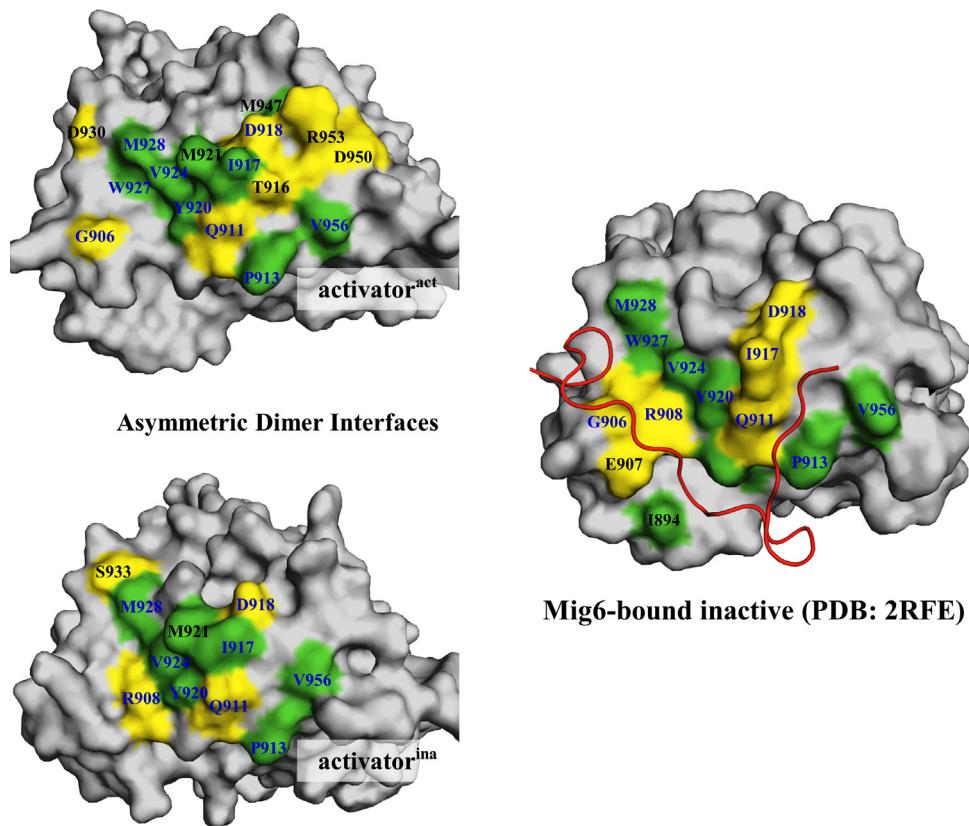


Fig. 8. Comparison of the binding interface at the C-lobe of the activator between the active and inactive models (left images) as well as the crystal structure of Mig6-bound inactive conformation (PDB code 2RFE). Mig6 peptide is shown in the red ribbon. The binding interfaces are shown in a surface representation and the amino acid residues contributing to H-bonds and hydrophobic interactions are colored in yellow and green, respectively. The shared interacting residues are labeled in blue letters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

[59]. Although it has been suggested that the core of the EGFR asymmetric dimer interface is dominated by hydrophobic interactions, our analyses on the estimated interaction energies suggested that in addition to the helix–helix packing (i.e. α H vs. α C), several interactions on the loop regions in particular the H-bond formation of both α G– α H and α I–CT loops of the activator with the JM segment of the receiver could be major contributing factors that stabilize the asymmetric dimer of EGFR kinase domain.

3.3.2. Dimer interface of the inactive conformation

Unlike the dimer interface of the active model, the helical interaction between the α H of the activator and α C of the receiver could be the major contribution in the asymmetric dimer of the inactive conformation (Fig. 7B and D). The α H helix exhibited more negative interaction energies than those from the α G– α H loop. Several hydrophobic interactions in the α H helix found in the active conformation were also observed in the dimer interface of the inactive conformation. These residues include I917, Y920, M921 and V924 on the α H helix, which made contacts with A678, L680 and I682 on the JM segment as well as L736, Y740 and A743 in the α C helix (Table 2). The strong hydrophobic interactions on the loop segments observed in the active conformation were also found (i.e. the M928–L758 and V956–P675). Although the H-bond interactions seemed to be contributing less in the asymmetric dimer of the inactive conformation since the number of interactions were less than those found in the active model, strong H-bond interactions between the α G– α H loop and the JM segment (i.e. Q911–L680 and Q911–A678) were observed (Table 1).

The precise location of the asymmetric dimer interface for the EGFR inactive conformation has not been proven to date. However, evidence from crystallographic analyses of PDB codes 2R4B and

3GOP suggested the possibility of the asymmetric dimer formation for the inactive conformation of EGFR kinase domain [12,13]. It should be noted that the crystal structure of the V924R-mutated EGFR kinase domain, which adopts the inactive conformation, cannot form the asymmetric dimer (PDB code 2GS7); nevertheless, it demonstrated extensive contacts in the context of the symmetric dimer instead (PDB code 3GT8 [11]). This finding indicates that the residue V924 of the activator is critical for the asymmetric dimer formation for both active and inactive conformations of EGFR kinase domain.

3.3.3. Comparison with the Mig6-kinase binding interface

The crystal structure of the complex between the EGFR kinase domain and a fragment of Mig6, which is a negative feedback inhibitor of EGFR, shows that residues K336–K362 of the Mig6 protein can bind to the C-lobe of EGFR kinase domain [10]. The crystal structure of such complex revealed that Mig6 peptide adopts an extended conformation and places on the superficial surface of the C-lobe of the kinase. The interface is dominated by the α G and α H helices, and the α F– α G, α G– α H, α H– α I loops, which buries the area of $\approx 18.00\text{ nm}^2$ [10]. Such interactions are largely polar; nonetheless, a few hydrophobic residues located on the α H helix contribute to the interface.

We then compared the interacting residues located on the C-lobe kinase domain between the asymmetric dimer interfaces and the Mig6-binding interface (Fig. 8). It was found that the Mig6-bound interface overlaps with the asymmetric dimer interfaces. The amino acid residues that contribute to the shared interface include (i) the H-bond contributions via G906, R908, Q911 on the α G– α H loop; I917, D918 on the α H helix, and (ii) hydrophobic interactions via P913 on the α G– α H loop; Y920, V921, V924, W927

on the α H helix; M928 on the α H– α I loop as well as V956 on the CT segment. The biochemical and biophysical studies on the Mig6 binding and EGFR inhibition suggested that introducing V924R mutation to the kinase domain can abolish the peptide binding. On the other hand, the kinase domain with the I682Q mutation has the same binding affinity as for the wild type protein [10]. Residue I682 is located on the JM segment on the N-lobe of EGFR kinase domain, and this mutant form is not activated because it is unable to form the asymmetric dimer, indicating that the Mig6 peptide can inhibit the EGFR activity in the context of the asymmetric dimer formation, but not in the basal activity of the monomeric kinase domain. The blockage of the asymmetric dimer interface for EGFR inhibition by the cytoplasmic protein Mig6 suggests one of the alternative approaches for development of a new class of inhibitors that can bind to the dimer interface on the C-lobe of EGFR kinase domain.

4. Conclusions

We have performed explicit-solvent MD simulations on two different conformations of EGFR kinase domain in the context of the asymmetric dimer. Although the active-to-inactive conformational transition (and vice versa) in both receiver and monomeric molecules was not observed within the 100-ns simulations, the long MD simulations in our study do allow us to assess relevant protein dynamics observed from the multiple ensemble simulations [23,56]. Our simulations on the active conformation showed that the asymmetric dimerization changed the global motion as well as suppressed the fluctuation of the receiver molecule. The reduced fluctuation in the intrinsically disordered regions and the dynamics alteration of EGFR kinase domain are consistent with the other works [20,22,24]. On the other hand, the asymmetric dimerization of the inactive conformation did not promote a destabilizing effect on the receiver molecule. The dimerization process suppressed its fluctuation to a greater extent and significantly changed its global dynamics; nevertheless, different kinds of motion in the A loop of all three inactive molecules would reflect intrinsic flexibility of the loop of EGFR inactive conformation.

We have investigated intermolecular interactions (i.e. H-bonds and hydrophobic interactions) at the asymmetric dimer interface of the active and inactive conformations of EGFR. The structural regions and key amino acid residues on the C-lobe of the activator as well as the N-lobe of the receiver responsible for such interactions in the asymmetric dimers were identified. The larger contributions by electrostatics rather than vdW interactions may correspond to presence of multiple H-bonds between several loops on the C-lobe of the activator and the JM segment of the N-lobe of the receiver. Our simulations also suggests that the interactions between the loop segments could play a role in the stabilization of the asymmetric dimers in addition to the hydrophobic contact of the helix–helix interaction, as previously demonstrated by crystallographic structures [9]. Furthermore, when compared with the interface of the Mig6–EGFR complex [10], several residues on the C-lobe of the kinase domain responsible for such peptide–protein interactions are also common to the asymmetric dimer interfaces. Since obstruction of the asymmetric dimer interface leads to the inhibition of EGFR activity by Mig6, this suggests another approach for development of novel inhibitors that bind to the C-lobe of the kinase domain. Information gained from our simulations should help us to locate key interacting motifs at the dimer interface on the C-lobe of EGFR kinase domain. This is the first step of a multi-step journey toward the design of peptide inhibitors to EGFR that can prevent the activator from either activating the isolated inactive receiver at an early stage of activation or to help stabilize the active monomeric populations in equilibrium.

Acknowledgments

This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission and the Thailand Research Fund grants RMU5180032 (KC). The authors are grateful for the support of the Faculty of Science, Kasetsart University and Professor Dr. Supa Hannongbua in particular. NS was supported by a grant under the program Strategic Scholarships for Frontier Research Network for the Join PhD Program Thai Doctoral degree from the Office of the Higher Education Commission, Ministry of Education, Thailand. KC is supported by Faculty of Science, Kasetsart University, and Thailand Research Fund (RSA5780066). All MD simulations were performed on the Virginia Tech's System X Supercomputer of the Terascale Computing Facility at Virginia Tech, United States. The authors also thank Dr. Justin A. Lemkul for his useful discussions about the results of this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jmgm.2015.03.002>.

References

- [1] F.R. Hirsch, M. Varella-Garcia, P.A. Bunn Jr., M.V. Di Maria, R. Veve, R.M. Bremmes, et al., Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis, *J. Clin. Oncol.* 21 (2003) 3798–3807.
- [2] M.A. Lemmon, J. Schlessinger, Cell signaling by receptor tyrosine kinases, *Cell* 141 (2010) 1117–1134.
- [3] N.E. Hynes, H.A. Lane, ERBB receptors and cancer: the complexity of targeted inhibitors, *Nat. Rev. Cancer* 5 (2005) 341–354.
- [4] K.M. Ferguson, Structure-based view of epidermal growth factor receptor regulation, *Annu. Rev. Biophys.* 37 (2008) 353–373.
- [5] G.M. Walton, W.S. Chen, M.G. Rosenfeld, G.N. Gill, Analysis of deletions of the carboxyl terminus of the epidermal growth factor receptor reveals self-phosphorylation at tyrosine 992 and enhanced *in vivo* tyrosine phosphorylation of cell substrates, *J. Biol. Chem.* 265 (1990) 1750–1754.
- [6] Y. Yarden, M.X. Sliwkowski, Untangling the ErbB signalling network, *Nat. Rev. Mol. Cell Biol.* 2 (2001) 127–137.
- [7] J. Stamos, M.X. Sliwkowski, C. Eigenbrot, Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor, *J. Biol. Chem.* 277 (2002) 46265–46272.
- [8] E.R. Wood, A.T. Truesdale, O.B. McDonald, D. Yuan, A. Hassell, S.H. Dickerson, 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.* 64 (2004) 6652–6659.
- [9] X. Zhang, J. Gureasko, K. Shen, P.A. Cole, J. Kurian, An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor, *Cell* 125 (2006) 1137–1149.
- [10] X. Zhang, K.A. Pickin, R. Bose, N. Jura, P.A. Cole, J. Kurian, Inhibition of the EGF receptor by binding of MIG6 to an activating kinase domain interface, *Nature* 450 (2007) 741–744.
- [11] N. Jura, N.F. Endres, K. Engel, S. Deindl, R. Das, M.H. Lamers, et al., Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment, *Cell* 137 (2009) 1293–1307.
- [12] M. Red Brewer, S.H. Choi, D. Alvarado, K. Moravcevic, A. Pozzi, M.A. Lemmon, et al., The juxtamembrane region of the EGF receptor functions as an activation domain, *Mol. Cell* 34 (2009) 641–651.
- [13] E.R. Wood, L.M. Shewchuk, B. Ellis, P. Brignola, R.L. Brashear, T.R. Caferro, et al., 6-Ethynylthieno[3,2-d]- and 6-ethynylthieno[2,3-d]pyrimidin-4-anilines as tunable covalent modifiers of ErbB kinases, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 2773–2778.
- [14] N. Jura, X. Zhang, N.F. Endres, M.A. Seeliger, T. Schindler, J. Kurian, Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms, *Mol. Cell* 42 (2011) 9–22.
- [15] B. Liu, B. Bernard, J.H. Wu, Impact of EGFR point mutations on the sensitivity to gefitinib: insights from comparative structural analyses and molecular dynamics simulations, *Proteins* 65 (2006) 331–346.
- [16] T.E. Balias, R.C. Rizzo, Quantitative prediction of fold resistance for inhibitors of EGFR, *Biochemistry*, 48 (2009) 8435–8448.
- [17] A. Dixit, G.M. Verkhiyker, Hierarchical modeling of activation mechanisms in the ABL and EGFR kinase domains: thermodynamic and mechanistic catalysts of kinase activation by cancer mutations, *PLoS Comput. Biol.* 5 (2009), e1000487.

- [18] S. Wan, P.V. Coveney, Rapid and accurate ranking of binding affinities of epidermal growth factor receptor sequences with selected lung cancer drugs, *J. R. Soc. Interface.* 8 (2011) 1114–1127.
- [19] O. Sawatdichaikul, S. Hannongbua, C. Sangma, P. Wolschann, K. Choowongkamon, In silico screening of epidermal growth factor receptor (EGFR) in the tyrosine kinase domain through a medicinal plant compound database, *J. Mol. Model.* 18 (2012) 1241–1254.
- [20] A. Papakyriakou, D. Vourloumis, F. Tzortzatou-Stathopoulou, M. Karpusas, Conformational dynamics of the EGFR kinase domain reveals structural features involved in activation, *Proteins* 76 (2009) 375–386.
- [21] S.E. Telesco, R. Radhakrishnan, Atomistic insights into regulatory mechanisms of the HER2 tyrosine kinase domain: a molecular dynamics study, *Biophys. J.* 96 (2009) 2321–2334.
- [22] M. Mustafa, A. Mirza, N. Kannan, Conformational regulation of the EGFR kinase core by the juxtamembrane and C-terminal tail: a molecular dynamics study, *Proteins* 79 (2011) 99–114.
- [23] S. Wan, P.V. Coveney, Molecular dynamics simulation reveals structural and thermodynamic features of kinase activation by cancer mutations within the epidermal growth factor receptor, *J. Comput. Chem.* 32 (2011) 2843–2852.
- [24] Y. Shan, M.P. Eastwood, X. Zhang, E.T. Kim, A. Arkhipov, R.O. Dror, et al., Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization, *Cell* 149 (2012) 860–870.
- [25] N. Songtawee, M.P. Gleeson, K. Choowongkamon, Computational study of EGFR inhibition: molecular dynamics studies on the active and inactive protein conformations, *J. Mol. Model.* 19 (2013) 497–509.
- [26] Y. Shan, A. Arkhipov, E.T. Kim, A.C. Pan, D.E. Shaw, Transitions to catalytically inactive conformations in EGFR kinase, *Proc. Natl. Acad. Sci. U.S.A.* 110 (2013) 7270–7275.
- [27] A. Arkhipov, Y. Shan, R. Das, N.F. Endres, M.P. Eastwood, D.E. Wemmer, et al., Architecture and membrane interactions of the EGF receptor, *Cell* 152 (2013) 557–569.
- [28] N. Guex, M.C. Peitsch, SWISS-MODEL the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis* 18 (1997) 2714–2723.
- [29] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.* 234 (1993) 779–815.
- [30] A. Fiser, A. Sali, ModLoop: automated modeling of loops in protein structures, *Bioinformatics* 19 (2003) 2500–2501.
- [31] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291.
- [32] C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, *J. Comput. Chem.* 25 (2004) 1656–1676.
- [33] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration, in: B. Pullman (Ed.), *Intermolecular Forces*, Reidel, Dordrecht, The Netherlands, 1981, pp. 331–342.
- [34] D. Van Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, *J. Comput. Chem.* 26 (2005) 1701–1718.
- [35] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, et al., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics* 29 (2013) 845–854.
- [36] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* 81 (1984) 3684–3690.
- [37] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* 126 (2007) 014101.
- [38] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, *J. Appl. Phys.* 52 (1981) 7182–7190.
- [39] S. Nosé, M.L. Klein, Constant pressure molecular dynamics for molecular systems, *Mol. Phys.* 50 (1983) 1055–1076.
- [40] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, *Mol. Phys.* 52 (1984) 255–268.
- [41] W.G. Hoover, Canonical dynamics: equilibrium phase-space distributions, *Phys. Rev. A* 31 (1985) 1695–1697.
- [42] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472.
- [43] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an $N \bullet \log(N)$ method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092.
- [44] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.* 103 (1995) 8577–8593.
- [45] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, et al., Gromacs User Manual Version 4.5.4, 2010 www.gromacs.org
- [46] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* 22 (1983) 2577–2637.
- [47] K.G. Tina, R. Bhadra, N. Srinivasan, PIC: protein interactions calculator, *Nucleic Acids Res.* 35 (2007) W473–W476.
- [48] P.J. Turner, Grace, Center for Coastal and Land-Margin Research Oregon Graduate Institute of Science and Technology, Beaverton, OR, USA, 2008.
- [49] W.L. DeLano, The PyMOL Molecular Graphics System, Version 1.3r1, Schrödinger, LLC, 2010.
- [50] A. Amadei, A.B. Linssen, H.J. Berendsen, Essential dynamics of proteins, *Proteins* 17 (1993) 412–425.
- [51] S. Hayward, B.L. de Groot, Normal modes and essential dynamics, *Methods Mol. Biol.* 443 (2008) 89–106.
- [52] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol.* 372 (2007) 774–797.
- [53] A.J. Shih, S.E. Telesco, S.H. Choi, M.A. Lemmon, R. Radhakrishnan, Molecular dynamics analysis of conserved hydrophobic and hydrophilic bond-interaction networks in ErbB family kinases, *Biochem. J.* 436 (2011) 241–251.
- [54] D.M. van Aalten, D.A. Conn, B.L. de Groot, H.J. Berendsen, J.B. Findlay, A. Amadei, Protein dynamics derived from clusters of crystal structures, *Biophys. J.* 73 (1997) 2891–2896.
- [55] B.L. de Groot, S. Hayward, D.M. van Aalten, A. Amadei, H.J. Berendsen, Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data, *Proteins* 31 (1998) 116–127.
- [56] A. Ivetac, J.A. McCammon, Elucidating the inhibition mechanism of HIV-1 non-nucleoside reverse transcriptase inhibitors through multicopy molecular dynamics simulations, *J. Mol. Biol.* 388 (2009) 644–658.
- [57] N. Narayana, S. Cox, X. Nguyen-huu, L.F. Ten Eyck, S.S. Taylor, A binary complex of the catalytic subunit of cAMP-dependent protein kinase and adenosine further defines conformational flexibility, *Structure* 5 (1997) 921–935.
- [58] M. Landau, N. Ben-Tal, Dynamic equilibrium between multiple active and inactive conformations explains regulation and oncogenic mutations in ErbB receptors, *Biochim. Biophys. Acta* 1785 (2008) 12–31.
- [59] R. Bose, X. Zhang, The ErbB kinase domain: structural perspectives into kinase activation and inhibition, *Exp. Cell Res.* 315 (2009) 649–658.