Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supplementary data

Ligand Binding Effects on the Activation of EGFR Extracellular Domain

Qiang Shao[1,2]*, Weiliang Zhu[1,2,3]

[1]Drug Discovery and Design Center, CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai, 201203, China

[2]University of Chinese Academy of Sciences, Beijing, 100049, China

[3]Open Studio for Druggability Research of Marine Natural Products, Pilot National Laboratory for Marine Science and Technology, 1 Wenhai Road, Aoshanwei, Jimo, Qingdao, 266237, China

*To whom correspondence should be addressed. Qiang Shao, Tel: +86 21 50806600-1304, E-mail: qshao@simm.ac.cn.

Materials and Methods

NMA measurement

The iMod program has been often used to explore the transition path connecting two endpoint structures [1-3]. From a given starting structure, a bunch of iterations of iMod performance were run until the protein arrived at the predefined target structure. In the k^{th} iteration, the intermediate structure $R^{(k)}$ is generated by:

$$R^{(k)} = R^{(k-1)} + v^{(k)} = R^{(k-1)} + S^{(k)} \sum_{i}^{m^{(k)}} (d^{(k-1)} \cdot u_i^{(k)}) u_i^{(k)} \quad (1)$$

where $R^{(k-1)}$ is the intermediate structure obtained in $(k\text{-}1)^{th}$ iteration, $v^{(k)}$ is the combined displacement along $m^{(k)}$ low-frequency eigenmodes calculated by iMod. The displacement along the i^{th} eigenmode is proportional to the projection $d^{(k-1)} \cdot u_i^{(k)}$ of the instantaneous distance vector $d^{(k-1)}$ on $u_i^{(k)}$, and scaled by the step size $S^{(k)}$. The number of $m^{(k)}$ at k^{th} iteration is determined by the following function:

$$C(m^{(k)}) = \sum_{i}^{m^{(k)}} \cos^2(d^{(k-1)}, u_i^{(k)}) \quad (2)$$

The value of $m^{(k)}$ is the minimal number of modes which start from the lowest frequency mode and end until $C(m^{(k)})$ reaches the cutoff of 0.8 as defined in multiple pioneering studies [4, 5].

ITSMD enhanced sampling method

As pointed out in earlier literatures [6-8], in the ITSMD method, the conformational sampling is enhanced through modifying the potential energy by a summation over the integration of a desired temperature range:

$$V' = -\frac{1}{\beta} \ln \sum_{k} f(\beta_k) e^{-\beta_k V} \quad (k = 1, \ldots, N) \quad (3)$$

where $\beta_k = \frac{1}{k_B T_k}$ (k_B is the Boltzmann constant and T is temperature), V is the original unmodified potential energy in system. The form of $f(\beta')$ can be taken as a sum of delta functions:

$$f(\beta') = \sum_{k=1}^{N} n_k \delta(\beta' - \beta_k) \quad (4)$$

Then the integration in Eq. (3) can be written as a summation over discrete temperature values:

$$V' = -\frac{1}{\beta} \ln \sum_{k=1}^{N} n_k e^{-\beta_k V} \quad (5)$$

By controlling the range setting of β_k, the modified energy V' can be expanded in a large range in ITSMD simulation in comparison to the conventional MD simulation. To fulfill the thorough sampling on the configuration space, it is better for the potential energy to be sampled evenly in the expanded energy range. The potential energy distribution function becomes:

$$p(V') = e^{-\beta V'} = \sum_{k=1}^{N} n_k e^{-\beta_k V} \quad (6)$$

To achieve an even distribution in the energy range, the parameters n_k's need to be determined by the requirement that each term in Eq. (6) contributes equally to the total distribution. A preliminary iteration process is required for achieving converged values of discrete n_k's:

1) $P_k = n_k \int_r e^{-\beta_k V(r)} dr$ was defined to integrate over the entire configuration space, with the ratios between P_k's preselected for all k between 1 to N. Using the normalized quantities ($p_k = P_k / \sum_{k=1}^{N} P_k$), we then defined a set of fixed expectation values $\{p_k^0\}$. For the desired even distribution of P_k's, p_k should be equal to 1/N.

2) A set of initial values of n_k's (n_k^0) were set to run the preliminary simulation. For every 100 steps, p_k's were recalculated and a new set of n_k^i were obtained from $n_k^i = n_k^{i-1} p_k^0 / p_k$, $i = i + 1$, followed by normalization.

3) Until p_k satisfactorily converged to p_k^0, the preliminary iteration process was stopped and the values of n_k's were determined.

Once the converged values of n_k's were achieved, the long-time production runs of ITSMD simulation were eventually performed. The biased force in the simulation is:

$$F' = -\frac{\partial V'}{\partial r} = \frac{\sum_{k=1}^{N} \beta_k n_k e^{-\beta_k V}}{\beta \sum_{k=1}^{N} n_k e^{-\beta_k V}} F \quad (7)$$

where F is the original molecular force in the simulation system.

Weighted histogram analysis of ITSMD simulation data

After the data were collected in the production runs, the true ensemble average of a thermodynamic quantity A could be measured by: $\langle A \rangle = \langle A \times W \rangle / \langle W \rangle$, where W is a reweighting factor for achieving unbiased distribution function [6, 8, 9]:

$$W = e^{\beta(V'(r) - V(r))} = e^{-\beta V(r)} / p(V') = e^{-\beta V(r)} / \sum_{k=1}^{N} n_k e^{-\beta_k V} \quad (8)$$

REFERENCES

1. J. R. Lopez-Blanco, J. I. Garzon, P. Chacon, iMod: multipurpose normal mode analysis in internal coordinates, Bioinformatics 27 (2011) 2843-2850.
2. J. R. Lopez-Blanco, J. I. Aliaga, E. S. Quintana-Orti, P. Chacon, iMODS: internal coordinates normal mode analysis server, Nucleic Acids Res 42 (2014) W271-W276.
3. J. R. Lopez-Blanco, P. Chacon, New generation of elastic network models, Curr Opin Struct Biol 37 (2016) 46-53.
4. O. Marques, Y.-H. Sanejouand, Hinge-bending motion in citrate synthase arising from normal mode calculations, Proteins 23 (1995) 557-560.
5. Z. Yang, P. Majek, I. Bahar, Allosteric transitions of supramolecular systems explored by network models: application to chaperonin GroEL, Plos Comput Biol 5 (2009) e1000360.
6. Y. Q. Gao, An integrate-over-temperature approach for enhanced sampling, J Chem Phys 128 (2008) 064105.
7. Q. Shao, J. Y. Shi, W. L. Zhu, Enhanced sampling molecular dynamics simulation captures experimentally suggested intermediate and unfolded states in the folding pathway of Trp-

cage miniprotein, J Chem Phys 137 (2012) 125103.
8. L. Yang, C. W. Liu, Q. Shao, J. Zhang, Y. Q. Gao, From thermodynamics to kinetics: enhanced sampling of rare events, Acc Chem Res 48 (2015) 947-955.
9. Q. Shao, Z. Xu, J. Wang, J. Shi, W. Zhu, Energetics and structural characterization of the "DFG-flip" conformational transition of B-RAF kinase: a SITS molecular dynamics study, Phys Chem Chem Phys 19 (2017) 1257-1267.

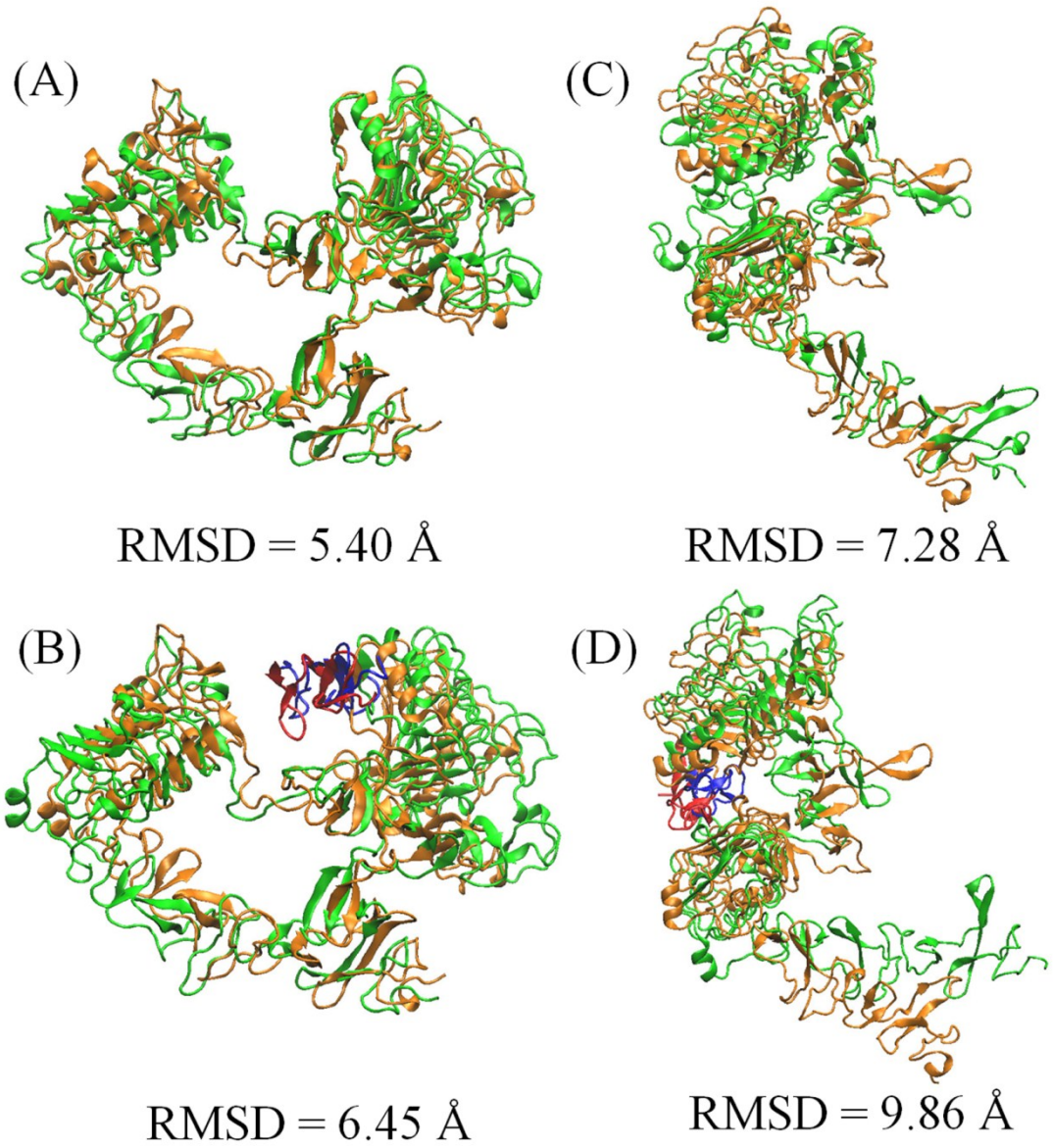

Fig. S1. Superposition of the representative simulated structures (green) to the experimental ones (orange). (A) Simulated and experimental inactive structures for unliganded EGFR-ECD. (B) Simulated and experimental inactive structures for liganded EGFR-ECD. (C) Simulated and experimental active structures for unliganded EGFR-ECD. (D) Simulated and experimental active structures for liganded EGFR-ECD. Under each structure is shown the backbone root-mean-square deviation (RMSD). The simulated EGF ligand is blue colored and the experimental one is red colored.

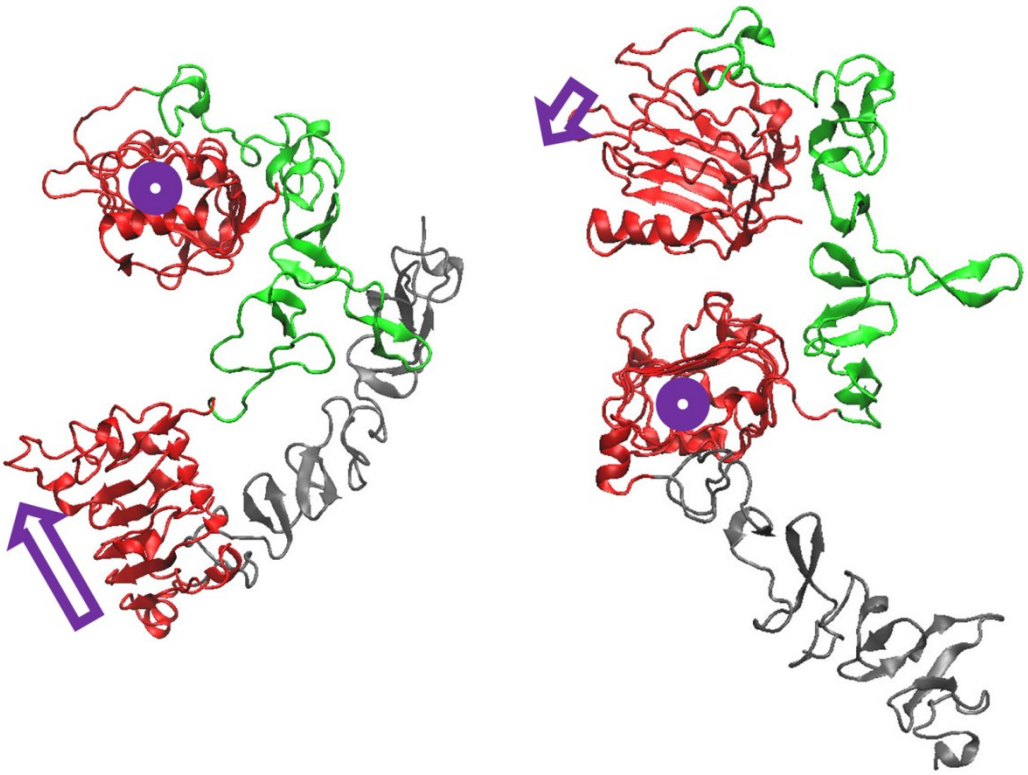

Fig. S2. Arrows showing the orientation of domains I and III in inactive (left) and active (right) states. A cycle indicates an arrow perpendicular to the page.

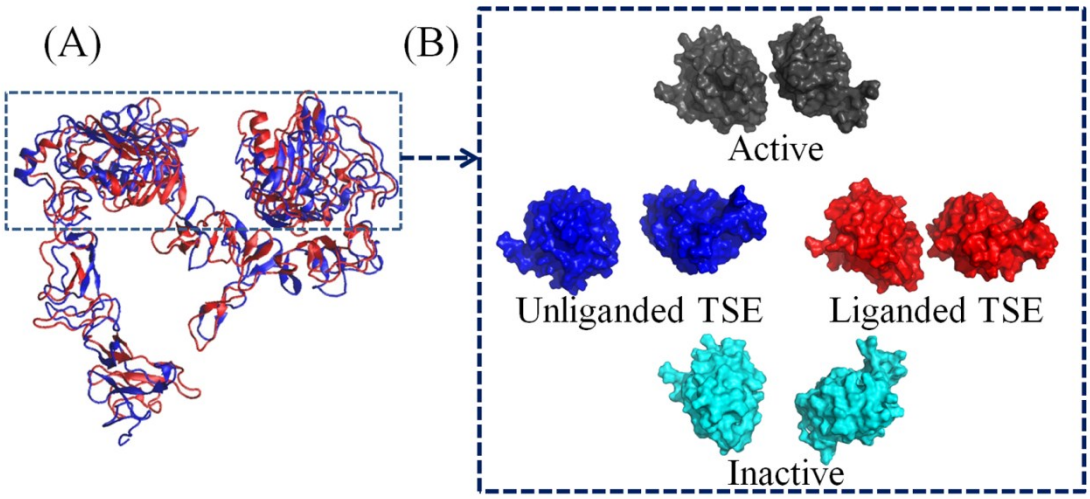

Fig. S3. (A) Superposition of the representative unliganded TSE structure (blue) to the liganded TSE structure (red) of EGFR-ECD. (B) Comparison of the position and orientation of domains I and III in simulated unliganded and liganded TSE states to those in the experimental inactive (cyan) and active (black) states.

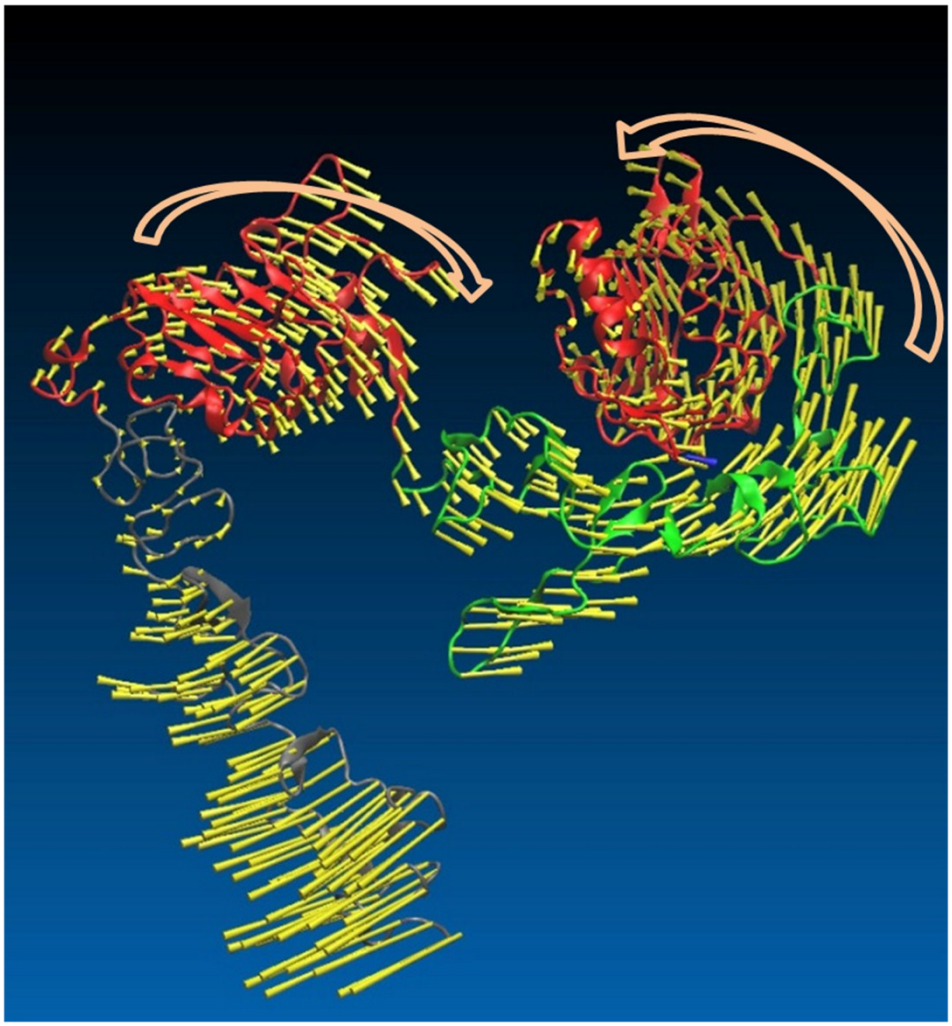

Fig. S4. Collective motions of the TSE state of unliganded EGFR-ECD obtained by iMod website service (http://imods.chaconlab.org).

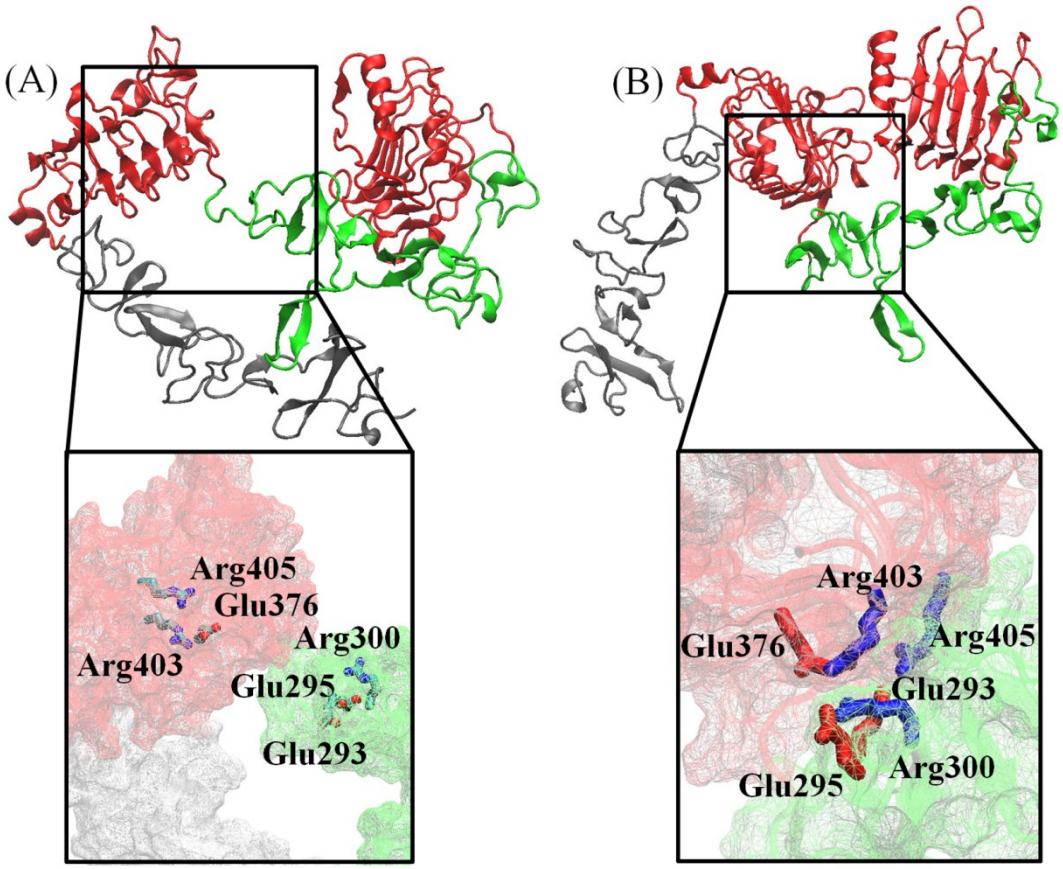

Fig. S5. The position of the residues consisting of the salt-bridge cluster in the active II-III domain interface in (A) inactive and (B) active states.

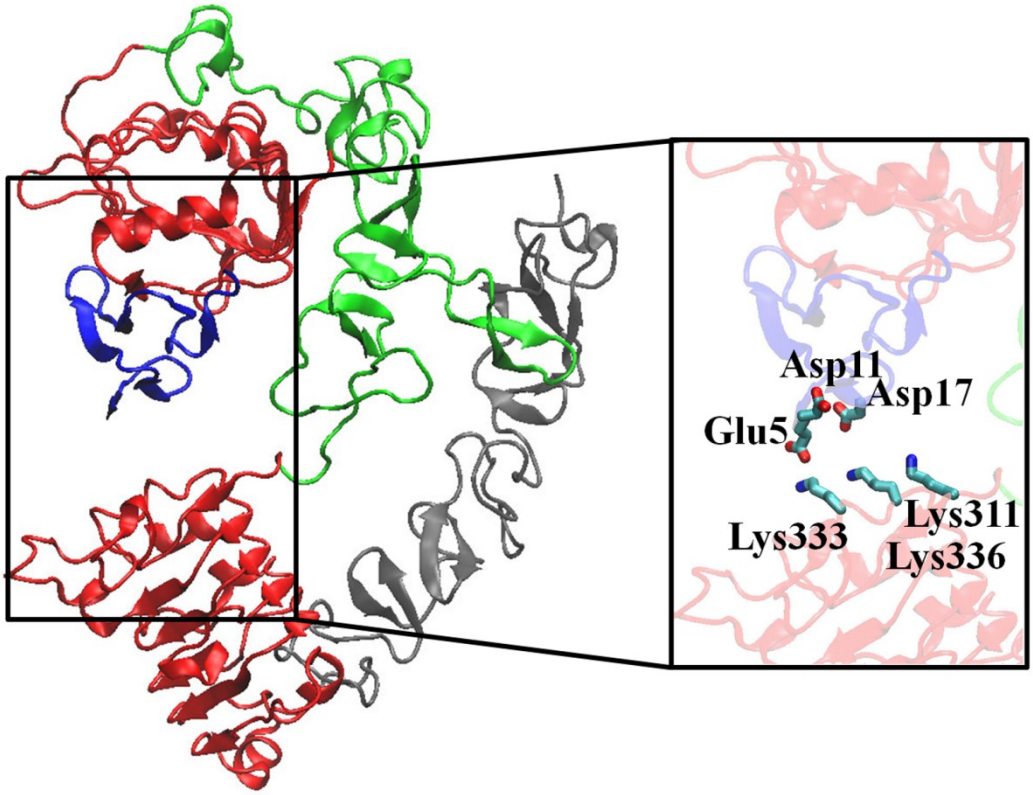

Fig. S6. The face-to-face orientation of the negatively charged residues (Glu5, Asp11, and Asp17) on EGF and the positively charged residues (Lys311, Lys333, and Lys336) on domain III in the inactive state of EGFR-ECD.