Supplementary Information

Mechanism of drug efficacy within the epidermal growth factor receptor revealed by microsecond molecular dynamics simulation

Shunzhou Wan, David W. Wright and Peter V. Coveney

Centre for Computational Science, Department of Chemistry, University College London, 20 Gordon St, London WC1H 0AJ, UK

Methods

System preparation. The molecular models we used have been described previously (1) and we briefly reprise the main features of the preparations here; the interested reader is referred to the original paper for full details. The x-ray structure (PDB id: 2GS6) (2) of wild-type EGFR tyrosine kinase domain (amino acids 679–959 in the pdf file, corresponding to 703–983 in the more common numbering system used in the main text) was used as the initial structure for the simulations. The ligand in the x-ray structure was removed to study EGFR dynamics in the apo state. The structure of L858R mutant EGFR was obtained by substituting leucine at 858 with arginine. The missing residues 748-750 (724-726 in the pdf file) at the N-terminal end of the αC-helix were modelled using ModLoop (3). Initial preparation of the systems was implemented using the binding affinity calculator (BAC) (4). The AMBER ff03 (5) all-atom force field was used to describe the protein parameters and TIP3P (6) for the water molecules. The Leap module in the AMBER (7) package was used for solvation and charge neutralization of the systems. The final systems consisted of approximately 50,000 atoms in rectangular boxes. The configurations of the systems after minimization and 2 ns NAMD (8) simulation were used as the initial structures for the current study. The structure files were then converted into Mae format using Maestro (9). Prior to production simulation, systems were equilibrated for another 10 ns using Desmond (10) on Ranger at the Texas Advanced Computing Center (TACC).

Simulation. Production runs were performed on Anton (11), a special-purpose machine for molecular dynamics (MD) simulation. The advances made in the Anton software and hardware make it possible to accelerate MD simulations by orders of magnitude relative to other existing parallel computing solutions. All production simulations were run in the NPT ensemble at 300 K and 1 bar using the Berendsen coupling scheme. Van der Waals and short-range electrostatic interactions were cut off at 9 Å. Long-range electrostatic interactions were computed using the *k*-space Gaussian split Ewald method (12), with a 64 × 64 × 64 grid. The multistep RESPA scheme was used for the integration of the equations of motion with timesteps of 2.0, 2.0, and 6.0 fs for the bonded, short-range nonbonded, and long-range nonbonded interactions, respectively. Finally, 10 μs production simulations were run for each molecular system and coordinates were recorded at every 250 ps.

Analyses. Principal component analysis (PCA) was used to identify the dominant low-frequency motions (1). The method allows us to turn the collective motions undergone by the system into a compact, low-dimensional representation. The trajectories of the backbone atoms were extracted from current 2 × 10 μs and our previous 3.2 μs (1) runs and combined into one single trajectory. The procedure generated a similar principal component space to that in our previous study (1), making direct comparison possible among different systems studied in our current and previous (1) simulations. Variations in the structure of a target protein influence the binding capacity of its ligand and drugs and therefore affect drug efficacy.

Non-overlap volume of a ligand was used to qualitatively measure the capacity of a protein to coordinate small molecules in the binding cleft. ATP or gefitinib, with their coordinates from x-ray structures, was fitted into the binding cleft, after properly aligning the residues around the cleft of each frame to the x-ray structures. The non-overlap volume was estimated as the volume of the ATP or ligand not shared by the protein. It was calculated as the difference in volume between the protein itself and the same conformer with the appropriate ligand fitted into the binding site. The binding capacity was then estimated by non-overlapped percentages of the ATP/ligand volumes ($V_{non\text{-}overlap}/V_{total}$). Although the flexibility of EGFR was represented by its ensemble conformations from MD simulations, each non-overlap estimation here was indeed the "lock-and-key" approximation in which fixed conformation was used for ATP/gefitinib from crystallography, and for protein from each snapshot of the simulations.

To take into account the flexibilities of ligand, we had applied a docking method to quantify the number of EGFR conformations capabale for ATP/gefitinib binding. The UCSF Dock 6.5 software (13) was used to predict the binding modes of ATP and gefitinib. To decease the computational cost of the post-analyses, the induced conformational change of the protein is not taken into account in current calculation. Docking method utilises very small molecules – fragments or functional groups of larger molecules – to generate efficient starting points. Here ATP and gefitinib were computationally deconstructed into fragments: an adenine ring, a ribose sugar and three phosphate groups for ATP, and an aniline, a quinazoline and a morpholine group for gefitinib. In the docking process, adenine ring of ATP and quinazoline of gefitinib first find their optimal orientations at the binding sites. These fragments act as anchors, on which layers of flexible regions were grown until ATP or gefitinib was fully built (13). For the successfully docked ligands, the conformations were clustered based on the distances between amino group of "hinge" residue M793 and nitrogen atoms N1, N3, N7 of ATP, and between the amino group and the nitrogen atoms N1, N3, NAS of gefitinib (Figures S2a and S2c). The clusters close to the x-ray conformations (PDB ids: 2GS6 (2) for ATP-EGFR and 2ITY (14) for gefitinib-EGFR) were regarded as the correct docking modes (Figures S2b and S2d), and the ratios of the numbers in the clusters to the total numbers of sampled snapshots were used to measure conformations capable for binding of ATP or gefitinib. The ratios were compared between the wild-type and mutant EGFRs, and between ATP and gefitinib.

Reference List

1. Wan S, Coveney PV. Molecular Dynamics Simulation Reveals Structural and Thermodynamic Features of Kinase Activation by Cancer Mutations Within the Epidermal Growth Factor Receptor. J Comput Chem 2011;32:2843-52.
2. Zhang XW, Gureasko J, Shen K, Cole PA, Kuriyan J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell 2006;125:1137-49.
3. Fiser A, Sali A. ModLoop: automated modeling of loops in protein structures. Bioinformatics 2003;19:2500-1.
4. Sadiq SK, Wright D, Watson SJ, Zasada SJ, Stoica I, Coveney PV. Automated molecular simulation based binding affinity calculator for ligand-bound HIV-1 proteases. J Chem Inf Model 2008;48:1909-19.
5. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong GM, Zhang W, *et al*. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 2003;24:1999-2012.
6. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. J Chem Phys 1983;79:926-35.
7. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, *et al*. The Amber biomolecular simulation programs. J Comput Chem 2005;26:1668-88.
8. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, *et al*. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781-802.
9. Maestro, version 9.2, Schrödinger, LLC, New York, NY [computer program]. 2011.
10. Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, *et al*. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, November 11-17 2006.
11. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, *et al*. Anton, a special-purpose machine for molecular dynamics simulation. CACM 2008;51:91-7.
12. Shan YB, Klepeis JL, Eastwood MP, Dror RO, Shaw DE. Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. J Chem Phys 2005;122:054101.
13. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, *et al*. DOCK 6: Combining techniques to model RNA-small molecule complexes. Rna-A Publication of the Rna Society 2009;15:1219-30.
14. Yun CH, Boggon TJ, Li YQ, Woo MS, Greulich H, Meyerson M, *et al*. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: Mechanism of activation and insights into differential inhibitor sensitivity. Cancer Cell 2007;11:217-27.