Supporting Information for "The Effect of a Wide-spread Cancer Causing Mutation on the Inactive to Active Dynamics of the B-Raf Kinase"

Kristen A. Marino, Ludovico Sutto, and Francesco L. Gervasio
Department of Chemistry, UCL, London, UK

Methods

Protein structures

The crystal structures for the active, αC-helix out, and DFG-motif out structures of the B-Raf kinase domain were taken from PDB files 4E26 [1], 3SKC [2], and 4DBN [3], the crystal structures of each state with the highest resolution. Since at the time the project was started no crystal structure of BRaf as a monomer was available, the protomer with the fewest missing residues in the A-loop was taken from the dimer. In the case of the αC-helix out inactive state (3SKC), one chain of the dimer had no missing residues in the A-loop, but the other two structures were missing residues in the A-Loop. The residues missing in crystal structures 4E26 and 4DBN were added using MODELLER [4] using other kinases as a template. MODELLER was also used to add residues to the termini of the structures to ensure all structures were the same length. To create the mutant structures, residue 600 was replaced in each of the three WT structures with the completed A-loop. The wild-type protein has 4459 atoms while the V600E mutant has 4458.

Simulation set-up

The GROMACS 4.6 [5] MD engine compiled with the plumed 1.3 plug-in [6] was used for all simulations. All simulations used the AMBER99sb*-ildn force field [7] to describe the protein and the TIP3P water model [8]. First the pdb files created by MODELLER were subject to 10,000 steps of steepest descent energy minimization. The proteins were then placed in a dodecahedron box and solvated with 13,610 water molecules, which was sufficient to ensure a minimum distance of 1 nm between the protein and the box. After charge neutralization with 8 (wild-type) or 7 (mutant) Cl- ions, energy minimization was performed again. NPT equilibration at 300 K and 1 bar was performed using Berendsen temperature and pressure coupling [9]. A 50 ns NVT run using v-rescale temperature coupling [10] with τ_T=0.2 was used to fully equilibrate the system. Long-range electrostatics were treated using PME [11] with a grid spacing of 0.12 Å. In real space, a switch function ensured that the potential decayed to zero between 0.8 and 1.0 Å. A switch function was also used for the van der Waals interactions between 0.8 and 1.0 Å. All bonds were constrained using LINCS allowing for a time-step of 2 fs.

Molecular dynamics and enhanced sampling

Two 500 ns runs for each of the six systems (three wild-type, three mutant) were performed starting from the equilibrated structures. Since 500 ns is too short to sample large conformational changes, we use parallel tempering metadynamics (PT-metaD) to increase sampling. Metadynamics achieves this by adding a history dependent biasing potential, *V(s,t)*, which acts on a small set of collective variables (CVs). In well-tempered metadynamics, the

bias is constructed from Gaussians that decrease in height as the simulation proceeds. The height of the gaussians added at time t is $W = W_0 e^{(-V(s,t)/(f-1)T)}$, where W_0 is the initial height, s is the value of the CV at time t, T is the replica temperature, and f is the bias factor. The bias encourages the system to sample regions of CV space not normally sampled in straightforward MD. The free energy surface is constructed as a function of the CVs. The CVs must distinguish between the different states of the system and represent all slow modes of the system making the selection of appropriate CVs crucial. This is difficult for systems which have a large number of degrees of freedom such as proteins. PT-metaD reduces this by sampling at increasing temperatures where free energy barriers are easier to overcome. In PT-metaD several independent metadynamics simulations (or replicas) at a range of temperatures are run in parallel and an exchange between adjacent replicas is attempted every 2 ps and accepted according to a Metropolis rule. Integrating metadynamics with parallel tempering allows the system to cross barriers on all degrees of freedom, including the slow degrees of freedom not biased in the metadynamics runs [12]. Two PT-metaD simulations were then performed, one for the wild-type and one for the mutant. Both simulations started from the equilibrated active state structure. Thirty-six replicas over a temperature range of 290-390 K were used to achieve an exchange rate of 20%. The intermediate temperatures were determined from an exponential fit. The initial Gaussians height, W_0, was 5 kJ/mol, the bias factor was 10, and the width of the Gaussians, σ, was 0.005. The PT-metaD simulations were started from the equilibrated active state conformation since it was unclear which crystallized inactive state was the primary inactive state.

Selection of CVs is always an important but difficult choice (see Ref. 15). The contact map CVs were successfully used in combination with PT-metaD to obtain converged conformational FESs associated with the activation of the EGFR and FGFR kinases [16,17]. Further, since two inactive states were crystallized, and there was no experimental evidence pointing to which one was the preferred inactive state or if both were equally probable, the contact map CVs were chosen because they were very general requiring no insight into the preferred active to inactive pathway (eg. using the DFG-motif dihedral angles to capture the DFG-flip). Three contact map collective variables were used for the metadynamics. The contacts were determined by first performing energy minimization of the equilibrated structures in vacuum. Then the unique contacts were determined for each of the three structures. A contact was considered unique if the distance between the atoms was less than 5 Å in a structure and the distance was greater than 5.5 Å in the other two structures. All contacts between the C_α, C_β, and O atoms of nonadjacent residues were considered. The contact map, *CM*, of structure i is defined as $CM_i = \frac{1}{N} \sum_{\gamma \in \Gamma}^{N} (D_\gamma(R) - D_\gamma(R_i))^2$.

Here, i refers to the reference structure (semi-active, αC-helix out, or DFG-motif out), N is a normalization constant and the formation of a contact γ in structure R is defined by a sigmoidal function, $D_\gamma(R)$.

$$D_\gamma(R) = w_\gamma \frac{1 - (r_\gamma/r_\gamma^0)^n}{1 - (r_\gamma/r_\gamma^0)^m}$$

The contact weight w_γ is 1 for all contacts and the exponents n and m are 6 and 10 respectively. For a contact, γ, r_γ is the contact distance in the structure R, while r_γ^0 is the contact distance in the reference structure in which the contact is unique, either semi-active, αC-helix out, or DFG-motif out.

Analysis
The FESs as a function of the three CVs biased in the PT-metaD simulations were constructed using the sum_hills utility of PLUMED 1.3. The FESs with CVs which were not biased in the simulation were constructed using the reweighting algorithm of Tiwary and Parrinello [13].
The rmsf of the unbiased simulations was calculated using the g_rmsf utility of GROMACS on the C_α atoms. To eliminate the role of large-scale conformational changes, the rmsf was calculated using overlapping windows of 50ns. The value for the trajectory was taken as the average of the rmsf calculated for the windows. The GROMACS g_helix tool was used to calculate the rmsd of the αG-helix with respect to an ideal α-helix. The salt-bridge distances were calculated using the N_ζ atoms of Lys483, Lys507 & Lys578, the C_ζ of Arg 603, the C_γ of Asp594 & Asp576, and the C_δ of Glu501 & Glu600. The volume of the ATP-binding site was estimated using mdpocket [14]. The amino acids making up the pocket were first determined from the unbiased MD trajectories of the wild-type and mutant. These residues were then used to calculate the volume in the structures sampled by the PT-metaD simulations at 300 K. K-medoid clustering was used to select representative structures in the minima. All structures from the 300 K replica which had an energy of less than 20 kJ/mol (relative to the minimum energy) were clustered based on their C_α rmsd.

Convergence
Several checks were made to ensure convergence of the simulations.
In Figure S1 we report the mono-dimensional projections of the 3D FES for V600E B-Raf reconstructed as a function of time in the last 500 ns with time intervals of 50 ns. It can be seen that all the relevant minima are present from the beginning that the relative free energy differences change less than 1 kJ/mol for the main minima and 5 kJ/mol for the secondary minima.

In Fig. S2 an independent estimation of the error on the one-dimensional free energy profiles for WT and V600E B-Raf is shown. The error, shown as a filled curve around the average free energy, is obtained as the standard deviation between the usual free energy obtained integrating the bias deposited along the PT-MetaD trajectory and the time-independent free energy estimate obtained using Tiwary and Parrinello reweighting algorithm[13]. The larger errors (always smaller than 10 kJ/mol) are mainly located at the transition barriers. The sampling error at the global and local minima is negligible (less than 1 kJ/mol).
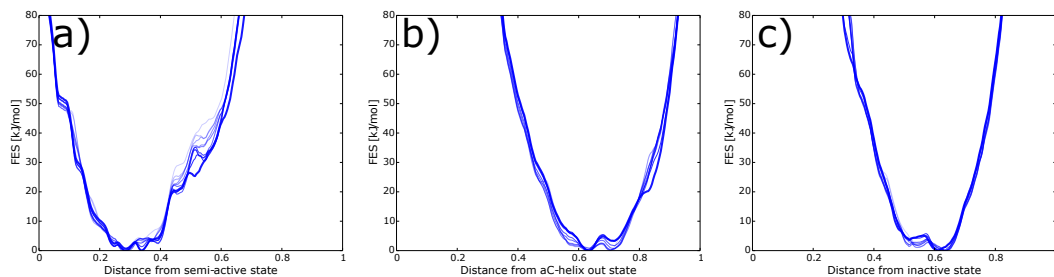
Figure S1. Free energy profiles at T=300 K of V600E B-Raf as a function of the 3 CVs. Ten curves with increasing color intensity are shown, separated by 50 ns of simulation time and corresponding to the last 500 ns of the 1900 ns simulation. The final converged FES is shown with a thicker blue line.
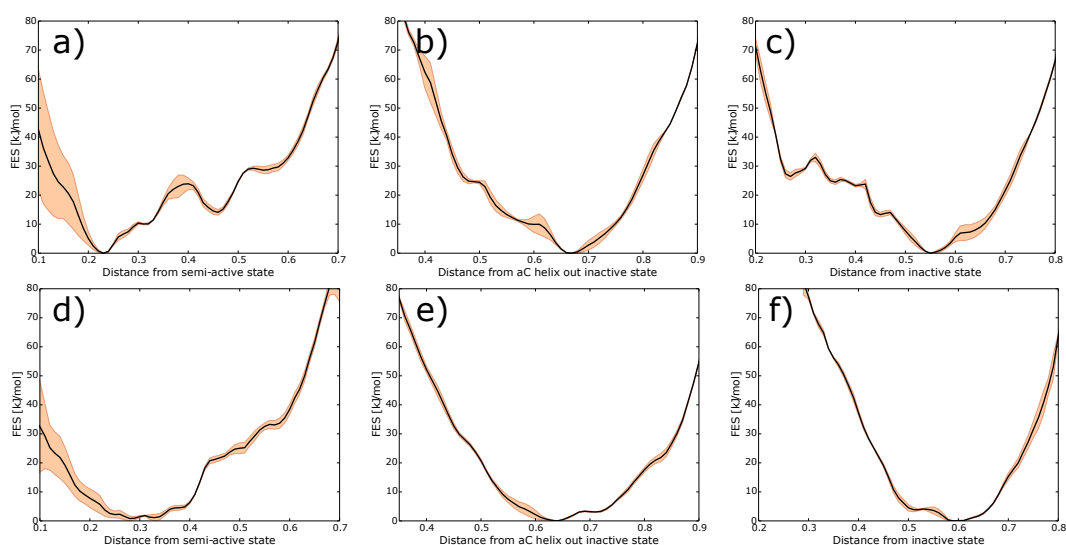

Figure S2. Mono-dimensional free energy profiles of WT B-RAF (top row) and V600E B-Raf (bottom row) as a function of the 3 CVs: a) and d) the distance from semi-active state, b) and e) the distance from αC-helix out inactive state, c) and f) the distance from the inactive state. The average free energy (black curve) and its standard deviation (colored) is calculated from 2 independent free energy profiles: one calculated by integrating the metadynamics bias deposited during the PT-MetaD run and the other by the time-independent free energy estimate obtained using Tiwary and Parrinello reweighting algorithm[13].

In Figure S3 the typical trajectories resulting from unbiased MD runs of the V600E mutant starting from the active structure with a fully extended A-loop and from the inactive structure are shown. The projection on the FES make evident that they not stable minima on the FES. Starting a 100 ns long unbiased simulation in one or the other structure and projecting its trajectory over the FES, shows that the trajectories depart from their initial structure. In the case of the active one (a) the trajectory immediately falls in the semi-active basin and stays there, while in the case of the inactive conformation (b) it diffuses towards a free energy plateau corresponding to the "intermediate" state of Fig. 4a. Three independent runs from each starting structure showed similar behavior (not shown).
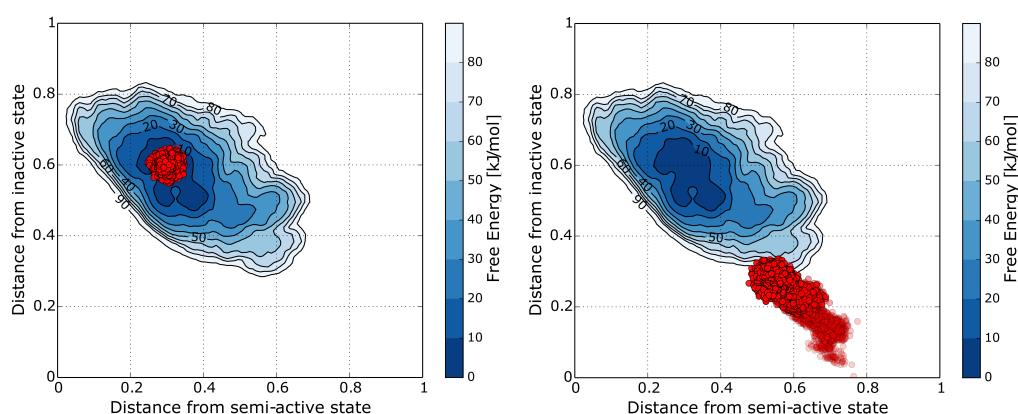

Figure S3. Two independent 100 ns-long unbiased simulations of V600E B-Raf are shown with red dots overlaid to the FES. A color-intensity code is used to show the progression in time, the most faded points corresponding to the initial structure. On the left hand-side the initial structure was an active conformation with a fully-extended A-loop while on the right hand side the initial structure was an inactive structure.
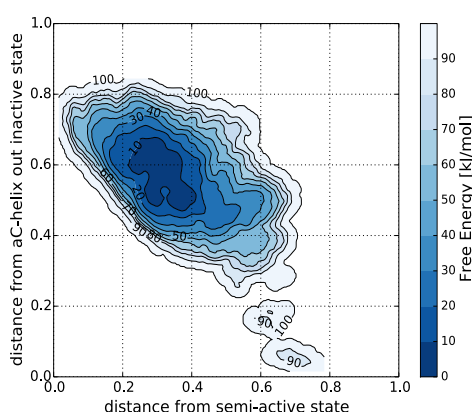

Figure S4. FES as a function of distance from semi-active state (x-axis) and distance from inactive state (y-axis) of the mutant B-Raf. The maximum energy threshold has been increased with respect to Fig. 4b to show the relative free energy difference between the inactive state (lower right corner) and the semi-active minimum. The large energy difference explains the behavior reported in Fig. S3.
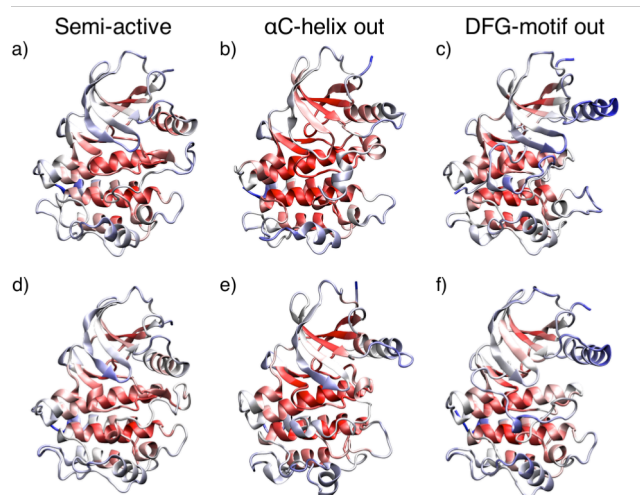
Semi-active   αC-helix out   DFG-motif out

Figure S5. RMSF of the Cα atoms during the 500 ns unbiased MD trajectories for wild-type (a,b,c) and mutant (d,e,f) B-Raf. The residues are colored by RMSF value with red having the lowest rmsf and dark blue having the highest value of 0.3 nm.

References

1. Qin, J.; Xie, P.; Ventocilla, C.; Zhou, G.; Vultur, A.; Chen, Q.; Liu, Q.; Herlyn, M.; Winkler, J.; Marmorstein, R. *J. Med. Chem.* **2012**, *55*, 5220.

2. Wenglowsky, S. et al. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 5533.

3. Okaniwa, M. et al. *J. Med. Chem.* **2012**, *55*, 3452.

4. Sali A. and Blundell T. J. *Journal of Molecular Biology* **1993**, *234***,** 779.

5. Hess, B.; Kutzner, C.; Spoel, D. V. D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.

6. Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. a.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961.

7. Lindor-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950.

8. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

9. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.*, **1984**, *81*, 3684.

10. Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

11. Laio, A; Gervasio, F. L. *Rep. Prog. Phys.*, **2008**, *71*, 126601; Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435.

12. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

13. Tiwary, P.; Parrinello, M. *J. Phys. Chem. B* **2014**, *119*, 736

14. Schmidtke, P.; Le Guilloux, V.; Maupetit, J.; Tuffery, P. *Nucleic Acids Res.* **2010**, *38*, 582.

15. Cavalli, A.; Spitaleri, A.; Saladino, G.; Gervasio, F. L. Acc. Chem. Res. **2015**, 48, 277–285.

16. Sutto, L.; Gervasio, F. L. Proc. Natl. Acad. Sci. U.S.A. **2013**, 110, 10616.

17. Bunney, T. D.; Wan, S.; Thiyagarajan, N.; Sutto, L.; Williams, S. V.; Ashford, P.; Koss, H.; Knowles, M. A.; Gervasio, F. L.; Coveney, P. V.; Katan, M. EBioMedicine **2015**, 2, 194–204.