Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Supporting Information for "Predicting the Conformational Variability of Oncogenic GTP-bound G12D Mutated KRas-4B Proteins at Zwitterionic Model Cell Membranes"

Huixia Lu[a] and Jordi Martí[*b]

1 Characteristics and convergence of Molecular Dynamics simulations

The main components of the system considered in this work are reported in Fig. 1. Species GTP, DOPC, DOPS and cholesterol are sketched. The sequence of aminoacids forming KRas-4B is indicated and the detailed structure of the farnesyl tail is also shown, including the methyl group introduced to KRas in the methylation process. The sites indicated in red correspond to relevant species that will be referred to in the text and the figures. The component aminoacids of KRas-4B are listed in Table 1. In addition, we show the secondary structures of KRas-4B in Fig. 2, as they will be referred in the main text.

The system was first energy minimised and then equilibrated in Molecular Dynamics (MD) runs of 10 ns. Production runs were performed within the NPT ensemble for 1 μs. The GROMACS/2018.3 package was employed[1] for the MD simulations. Time steps of 2 fs were used in all production simulations and the particle mesh Ewald method with Coulomb radius of 1.2 nm was employed to compute long ranged electrostatic interactions. The cutoff for Lennard-Jones interactions was set to 1.2 nm. Pressure was controlled by a Parrinello-Rahman piston with damping coefficient of 5 ps^{-1} whereas temperature was controlled by a Nosé-Hoover thermostat with a damping coefficient of 1 ps^{-1}. Finally, periodic boundary conditions in three directions of space have been taken.

In order to have additional information on the radius of gyration Rg of the protein (given by Eq. 1) of a selected group of residues forming a hydrophobic core (see section 3.2 of the main text) we report Rg as a function of the simulation time in Fig. 3. The equilibration regime is observed after 100 ns.

$$R_g = \sqrt{\frac{\sum_i \|r_i\|^2 m_i}{\sum_i m_i}}, \quad (1)$$

where m_i is the mass of atom i and r_i the position of the same atom with respect to the centre of mass of the selected group.

On the convergence of MD runs, we have computed the thickness of the membrane, its area per lipid, the radius of gyration (according to the definition given in Eq. 1) of different parts of KRas-4B (catalytic domain, allosteric domain and effector domain) along the last 500 ns trajectory of the standard MD simulations of total length 1000 ns. We have also calculated a solvent-accessible surface area (SASA) for the same domains considered above and for the contact area between the allosteric and effector domains of KRas-4B, using in-house VMD TCL scripts along the whole trajectory. We can observe all four computed properties in Fig. 4. The convergence of MD is clearly established in all cases, with fluctuations below 10% of the corresponding absolute values.

2 Characteristics and convergence of well-tempered metadynamics simulations

The values for the parameters of the well-tempered metadynamics (WTM) simulations[3] are listed in Table 2. The convergence of WTM simulations needs to be assessed because it is important to ensure that the number of transition events between stable states is statistically meaningful and also that sampling of all phase space of the collective variables (CV) has been achieved[4]. In the case of protein-bilayer systems with hundreds of thousands of atoms, to cover the full range of the CV space is a challenge. In the present work, two angular CV (Φ and Θ as defined in Section 3 of the manuscript) were selected.

Usually the size of the hills of the Gaussian kernels deposited along the simulation is monitorized. As the simulation progresses and the added bias grows, the Gaussian height would be progressively reduced, eventually including low-height spikes. In the present case, the height of the biased potential decreased accordingly along the simulation run, as indicated in Fig. 5. A quasi-flat profile is already seen after 450 ns, although some spikes due to large fluctuations in the values of CV are observed.

[a‡]*School of Pharmacy, Shanghai Jiaotong University, Shanghai,China. E-mail: huixia.lu@sjtu.edu.cn*
[b‡]*Department of Physics, Technical University of Catalonia-Barcelona Tech, B5-209 Northern Campus, Jordi Girona 1-3, 08034 Barcelona, Catalonia, Spain. E-mail: jordi.marti@upc.edu*
*Corresponding author
‡These authors contributed equally to this work

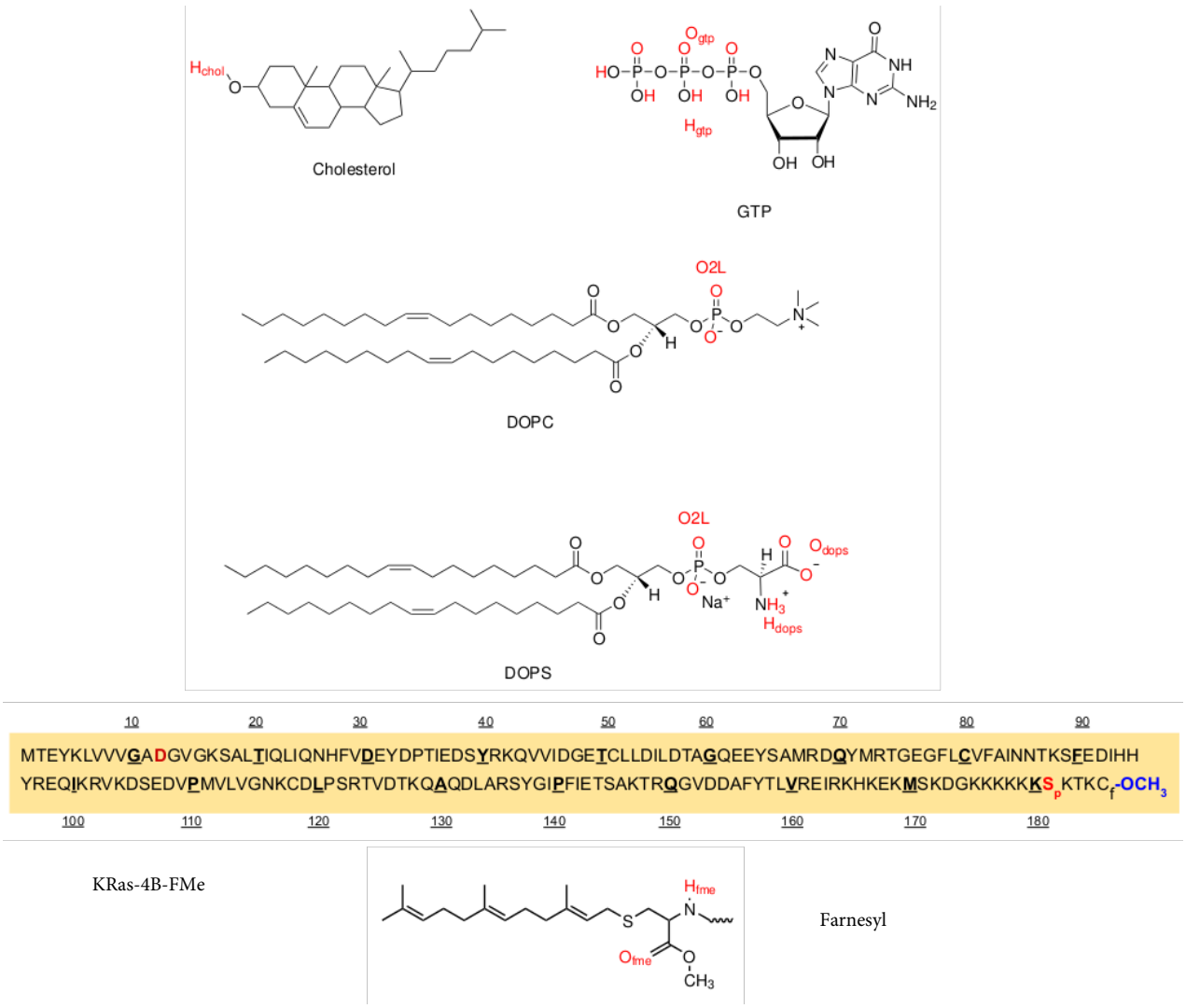

Fig. 1 Sketches of GTP, DOPC, DOPS, cholesterol, KRas-4B-FMe, with details of the farnesyl (FAR) group, labelled as C_f.

Despite the fact that the Gaussian height is decreasing to zero, it has been extensively shown that other measures of convergence of a metadynamics simulation must be explored in order to make sure the system is fully converged. To do this, we will proceed as follows:

1. The time evolution of the two CV along the full time span of our simulations has been reported in Fig. 6, we can see that the system is diffusing efficiently in the collective variable space and the bias potential produces that both CV are able to diffuse in full phase space. We can only observe a few regions where the diffusion was less efficient, indicating angular configurations not likely to reach and corresponding to the regions of highest free-energy as shown in Fig.5 of the main text.

2. To assess the convergence of the WTM simulation we have primarily calculated the integrated free-energy profiles as a function of simulation time (see Fig. 7). As we can see that the estimated free-energies do not change significantly in the final part of our simulations (starting at 550 ns) which also ensures us the convergence of the simulations.

3. Convergence can be further assessed in different ways with more precision. We have considered two additional methods: (1) reporting the *block analysis* of the average error along free-energy profiles as a function of the block length and (2) performing the *committor analysis* of the CV (see PLUMED project[5,6]). In method (1) (see Fig.8) we report the size of the average error in the free-energy profiles calculated from two different sets of well-tempered metadynamics simulations as a function of the block size. The initial 25% of all metadynamics trajectories were discarded. As expected, the errors increase with the block length until they reach a plateau in all cases. The average errors tend to stabilise around 0.42-0.43 kJ/mol for the two considered CV. For this calculation, scripts provided by the PLUMED project[5,6] have been employed. Method (2) consists in throwing a large number of short MD simulations started at the local transition state (TS) between the two basins A and B, as indicated in Figs. 5-7 of main text and check whether

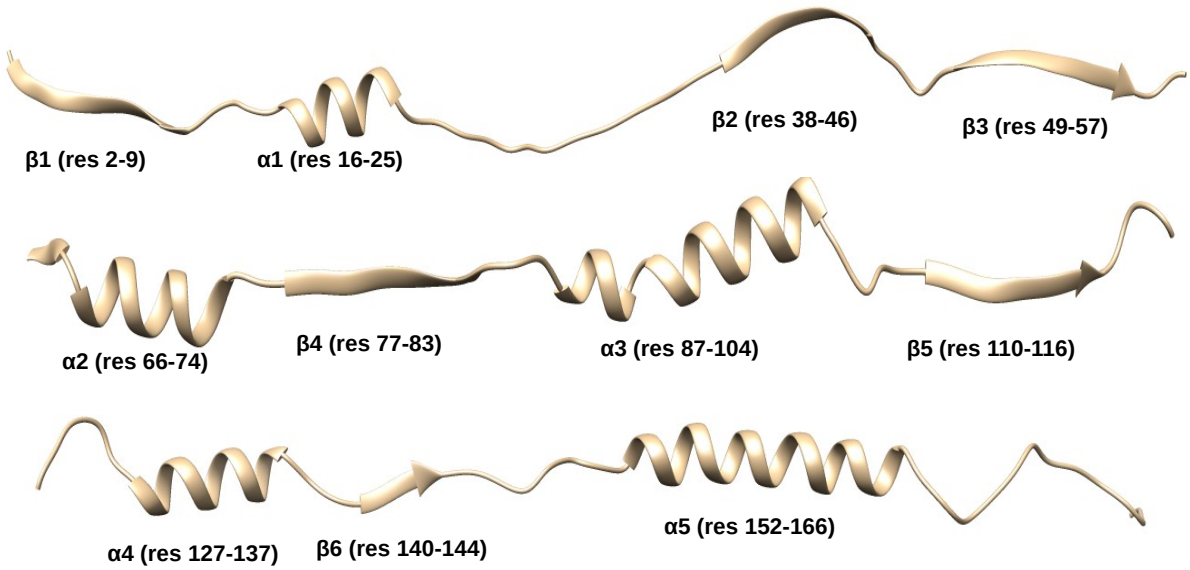

Fig. 2 Secondary structure of KRas-4B, image generated by Chimera-1.15[2]

the final values of those CV end up in one of the two stable basins of the free-energy hypersurface. After throwing around 500 MD trajectories started at the TS, we found (see Fig. 9) that 47.6 % of them ended in basin A and 52.4% ended in basin B. Given the similar probabilities we can conclude that the choice of CV was successful.

3 Calculation of the minimum free energy paths

In order to obtain and evaluate the paths connecting the two stable states located on the 2D free energy landscape we have considered a process including three steps: (a) fixing the two local minima; (b) locating a coarse path and (c) refining the path. The coordinates of the path are numerically reported in Table 3. The first coordinate of each point corresponds to CV1 (Φ) and the final coordinate to CV2 (Θ). The R-package metadynminer[7] reads HILLS files from PLUMED, calculates free energy surface by fast Bias Sum algorithm, finds minima and analyses transition paths by Nudged Elastic Band method.

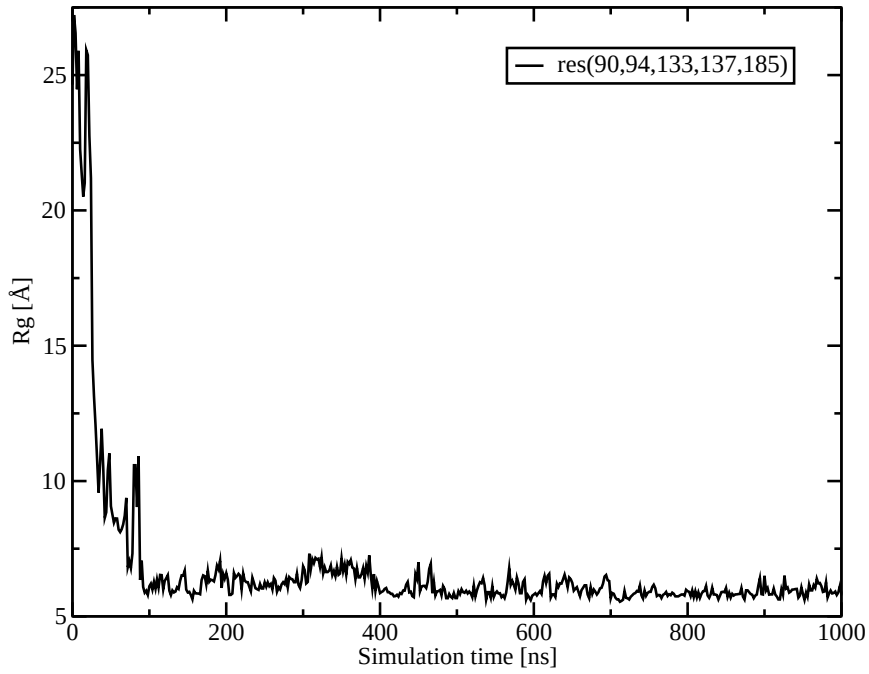

Fig. 3 Radius of gyration of the group formed by residues 90, 94, 133, 137 and 185 of KRas-4B as a function of the simulation time.

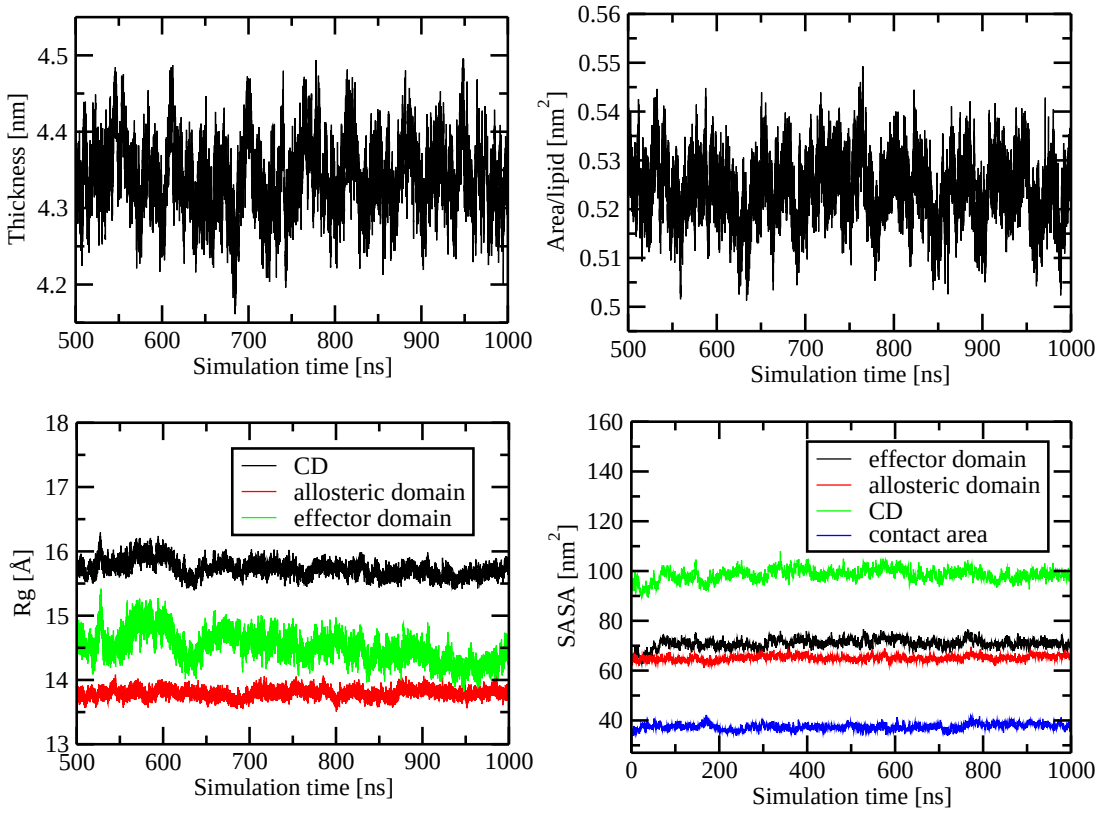

Fig. 4 Thickness, area per lipid, radius of gyration and solvent-accessible surface area as the function of simulation time.

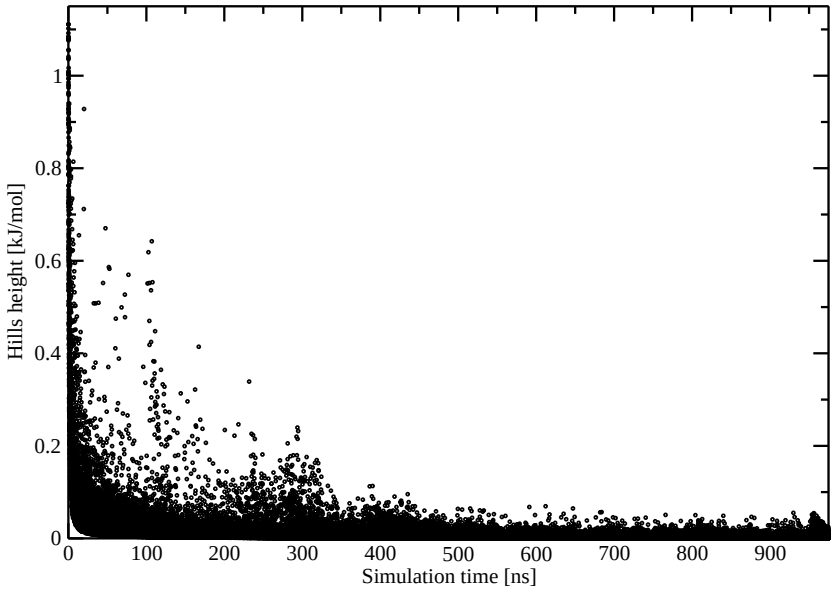

Fig. 5 Well-tempered Metadynamics hills height as a function of simulation time.

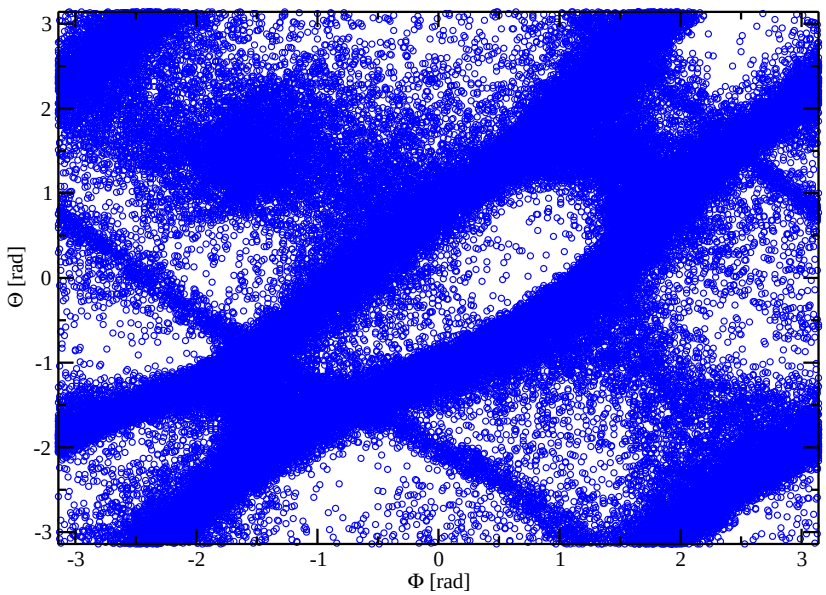

Fig. 6 Time evolution of the CV during the WTM simulations.

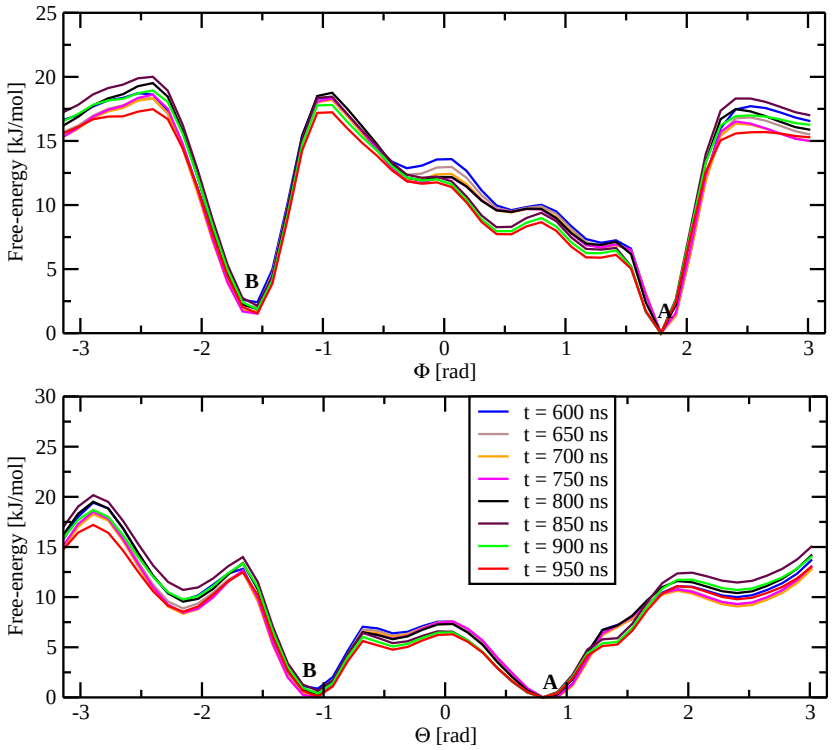

Fig. 7 Estimates of the integrated free-energy profiles as a function of the WTM CV deposited along a 975 ns-long WTM simulation time. The selected basins have been shown in Fig. 5 of the main text.

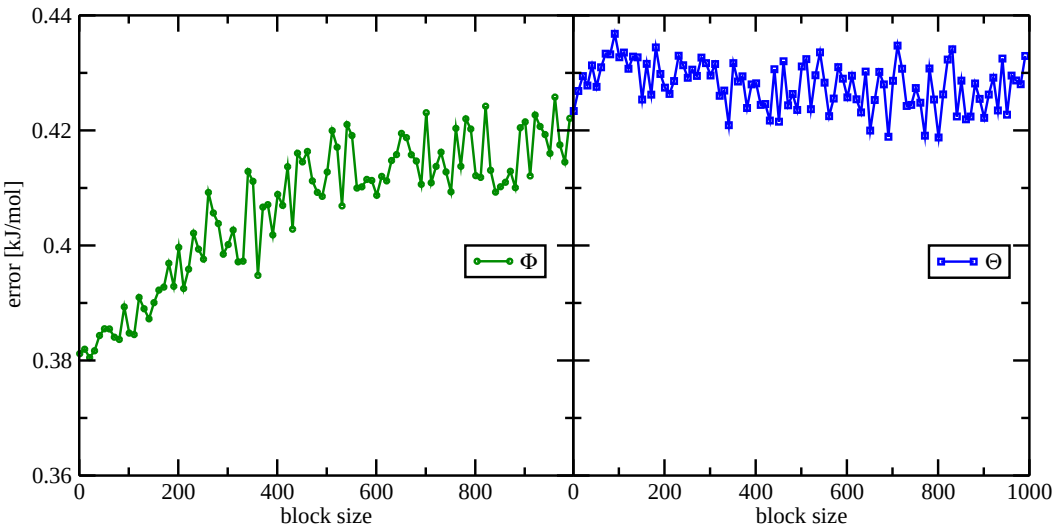

Fig. 8 Block analysis of the average errors along the free-energy profiles for Θ (left) and Φ (right) as a function of the block length. Taken from the last 400-600 ns of each metadynamics simulation.

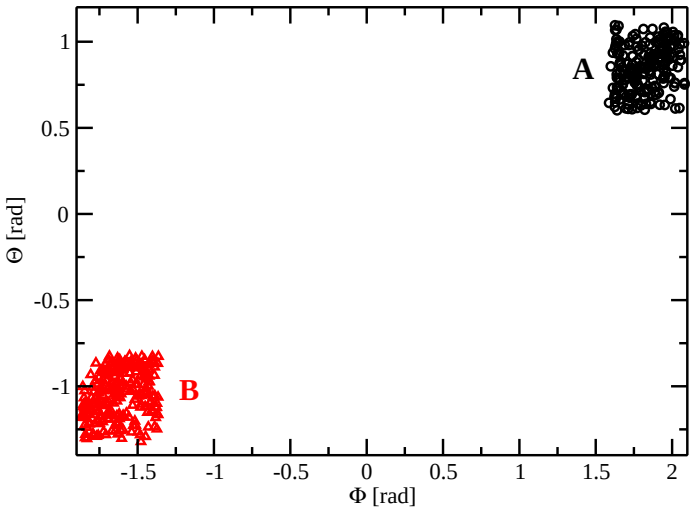

Fig. 9 Committor analysis of the stable basins A and B (see Figs.5-7 of the main text).

4 Tables

Table 1 Aminoacid components of the KRas-4B proteins.

Full name	Abbreviation
Alanine	Ala
Arginine	Arg
Asparagine	Asn
Aspartate	Asp
Cysteine	Cys
Glutamate	Glu
Glutamine	Gln
Glycine	Gly
Histidine	His
Isoleucine	Ile
Leucine	Leu
Lysine	Lys
Methionine	Met
Phenylalanine	Phe
Proline	Pro
Serine	Ser
Threonine	Thr
Tryptophan	Trp
Tyrosine	Tyr
Valine	Val

Table 2 Parameters employed in metadynamics simulations.

Gaussian width of CV1 [rad]	0.35
Gaussian width of CV2 [rad]	0.35
Starting (Gaussian) hill [kJ/mol]	1.0
Deposition stride [ps]	1
Bias factor	10
Temperature	310.0 K
Simulation time [ns]	1000

Table 3 Coordinates of segments forming selected paths for oncogenic KRas-4B-FMe. Along the path, locations of free energy spots (fespot, in kJ/mol) are displayed.

Minimum free energy path			
State	path x	path y	fespot
B	-1.58	-1.04	0.84
	-1.68	-1.23	3.48
	-1.62	-1.46	14.95
	-1.48	-1.66	19.90
	-1.31	-1.83	14.05
	-1.07	-1.77	17.47
	-0.85	-1.66	18.40
	-0.63	-1.56	16.56
	-0.39	-1.48	14.89
	-0.16	-1.37	14.71
	0.06	-1.24	15.11
	0.28	-1.09	13.94
	0.50	-0.95	11.96
	0.73	-0.80	10.02
	0.94	-0.64	7.70
	1.14	-0.45	5.48
C	1.19	-0.38	5.14
	1.31	-0.22	5.45
	1.45	0.01	6.22
	1.58	0.26	4.97
	1.69	0.52	2.18
A	1.78	0.77	0.00
	1.85	0.95	0.56
	1.85	1.14	5.28
	1.78	1.22	11.39
	1.69	1.31	21.00
	1.61	1.38	26.80
	1.51	1.46	30.90
TS	1.41	1.53	30.90
	1.33	1.63	28.16
	1.22	1.69	22.14
	1.14	1.77	15.87
	0.94	1.78	12.15
	0.87	1.71	10.20
	0.73	1.58	8.12
	0.65	1.51	7.11
D	0.48	1.39	6.34
	0.37	1.29	7.13
	0.25	1.19	8.74
	0.02	0.97	11.27
	-0.32	0.63	12.07
	-0.42	0.51	13.20
	-0.52	0.38	14.23
	-0.69	0.11	15.40
	-0.87	-0.15	17.64
	-0.97	-0.28	18.41
	-1.07	-0.41	17.69
	-1.17	-0.53	14.78
	-1.28	-0.65	9.66
	-1.38	-0.77	4.89
	-1.48	-0.90	1.81
B	-1.58	-1.04	0.84

Notes and references

1 H. J. Berendsen, D. van der Spoel and R. van Drunen, *Computer Physics Communications*, 1995, **91**, 43–56.

2 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *Journal of computational chemistry*, 2004, **25**, 1605–1612.

3 A. Barducci, G. Bussi and M. Parrinello, *Physical review letters*, 2008, **100**, 020603.

4 Y. Zhou, P. Prakash, H. Liang, K.-J. Cho, A. A. Gorfe and J. F. Hancock, *Cell*, 2017, **168**, 239–251.

5 M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci and R. A. e. a. Broglia, *Computer Physics Communications*, 2009, **180**, 1961–1972.

6 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Computer Physics Communications*, 2014, **185**, 604–613.

7 P. Hošek and V. Spiwok, *Computer Physics Communications*, 2016, **198**, 222–229.