Electronic Supplementary Material (ESI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2020
Updated 12/03/2020

SUPPORTING INFORMATION

Unravelling the effect of the E545K mutation on PI3Kα kinase

Ioannis Galdadas[a], Francesco Luigi Gervasio[a,b,c*], Zoe Cournia[d*]

[a] Department of Chemistry, University College London, London WC1E 6BT, United Kingdom

[b] Institute of Structural and Molecular Biology, University College London, London WC1E 6BT, United Kingdom

[c] Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland.

[d] Biomedical Research Foundation, Academy of Athens, Athens 11527, Greece

1. Protein structure preparation

The X-ray structure of the human WT PI3Kα was retrieved from the Protein Data Bank (PDB ID: 4OVU)[1]. The missing residues in the catalytic, p110α, and regulatory, p85α, domains were modelled and refined through homology and loop modelling using the Prime software of the Schrödinger suite [Prime, Schrödinger, LLC, New York, NY, 2019]. The SH3, GAD, and cSH2 domains regions in the *PIK3R1* gene (Figure 1, main text) were not modelled due to lack of reliable structural information. In the absence of crystallographic data regarding the structure of the E545K mutant, residue $Glu545_{p110\alpha}$ of the WT structure was mutated *in silico* to $Lys545_{p110\alpha}$. After the introduction of the point mutation, the position of the backbone nitrogen atom of Lys545 in the helical domain was optimal for the formation of a hydrogen bond with the backbone oxygen of the neighbouring $Leu380_{p85\alpha}$ of the nSH2 domain. It should be noted that no PIP2 or ATP was included in any of the simulated systems. The protonation state of each residue was determined using the PROPKA3.0 algorithm[2] at pH 7, which assigned the usual charge states to all residues, except for $Asp603_{p110\alpha}$ and $Asp560_{p85\alpha}$ that were predicted to be protonated. Lys545 was protonated.

2. Unbiased MD simulations set-up

The GROMACS 2016.4[3] MD engine compiled with the plumed 2.4.1[4,5] plug-in was used for all simulations. All simulations used the CHARMM36m force field[6] to describe the protein and a modified TIP3P water model referred to as CHARMM TIP3P[6], which has been fine-tuned to optimise the dispersion part of the protein-water interactions, while minimising the

perturbation of the repulsive part. CHARMM36m was selected because it has been shown to reproduce well the dynamics of both folded and intrinsically disordered proteins and can qualitatively capture the effects of pathogenic mutations[7,8]. The systems were solvated with ~11,000 CHARMM TIP3P water molecules in a dodecahedral box with periodic boundary conditions, while Na^+ and Cl^- ions were added to reach charge neutrality and the final concentration of 15 mM (the total number of atoms was ~350,000).

Prior to MD, the energy of all systems were minimised using the steepest descent integrator and were then equilibrated in the NVT ensemble for 5 ns, using a Berendsen barostat[9] at 1 bar with initial velocities sampled from the Boltzmann distribution at 300 K. The temperature was kept constant at 300 K by a velocity-rescale thermostat[10] and a time step of 2 fs. The long-range electrostatics were calculated by the particle mesh Ewald algorithm, with Fourier spacing of 0.16 nm, combined with a switching function for the direct space between 0.8 and 1.0 nm for better energy conservation. The systems were equilibrated for additional 10 ns in the canonical (NPT) ensemble applying position constraints to the protein (with a restraint spring constant of 1 $kJmol^{-1}nm^{-1}$) and keeping the pressure constant at 1 bar by coupling isotropically the system every 2 ps to a Parinello-Rahman barostat[11]. A 1 μs long and two more independent 500 ns long unconstrained production runs were carried out for each of the systems in the canonical (NPT) ensemble.

3. Enhanced sampling simulations set-up

Complex conformational changes are most often associated with transitions between states that are separated by energy barriers, and they can be adequately studied with enhanced sampling methods, which allow estimating the free energy of the separated states as a function of appropriate collective variables (CVs) that approximate the reaction coordinate.

Since 1 μs of unbiased MD was insufficient to sample any large conformational changes that could explain the effect of the E545K mutation on the mutation-induced activation of PI3Kα, we used multiple-walkers metadynamics (metaD) simulations using a total of eight walkers to increase the sampling. This approach has been proven to be an effective strategy to explore large-scale conformational changes, and allosteric effects in a number of drug targets[12–14]. The method is based on the concurrent simulation of the system by different “walkers”, while adding a history-dependent potential on the most important degrees of freedom, which are expected to drive the process of interest; in this case the detachment of the nHS2 domain from the helical domain. The history-dependent potential on these CVs not only discourages the system to revisit the same regions of the conformational space but also favours the exploration of new, stable or metastable configurations until the system explores the whole (or a large portion of) phase space. In the multiple walkers metaD, all walkers fill the same free energy

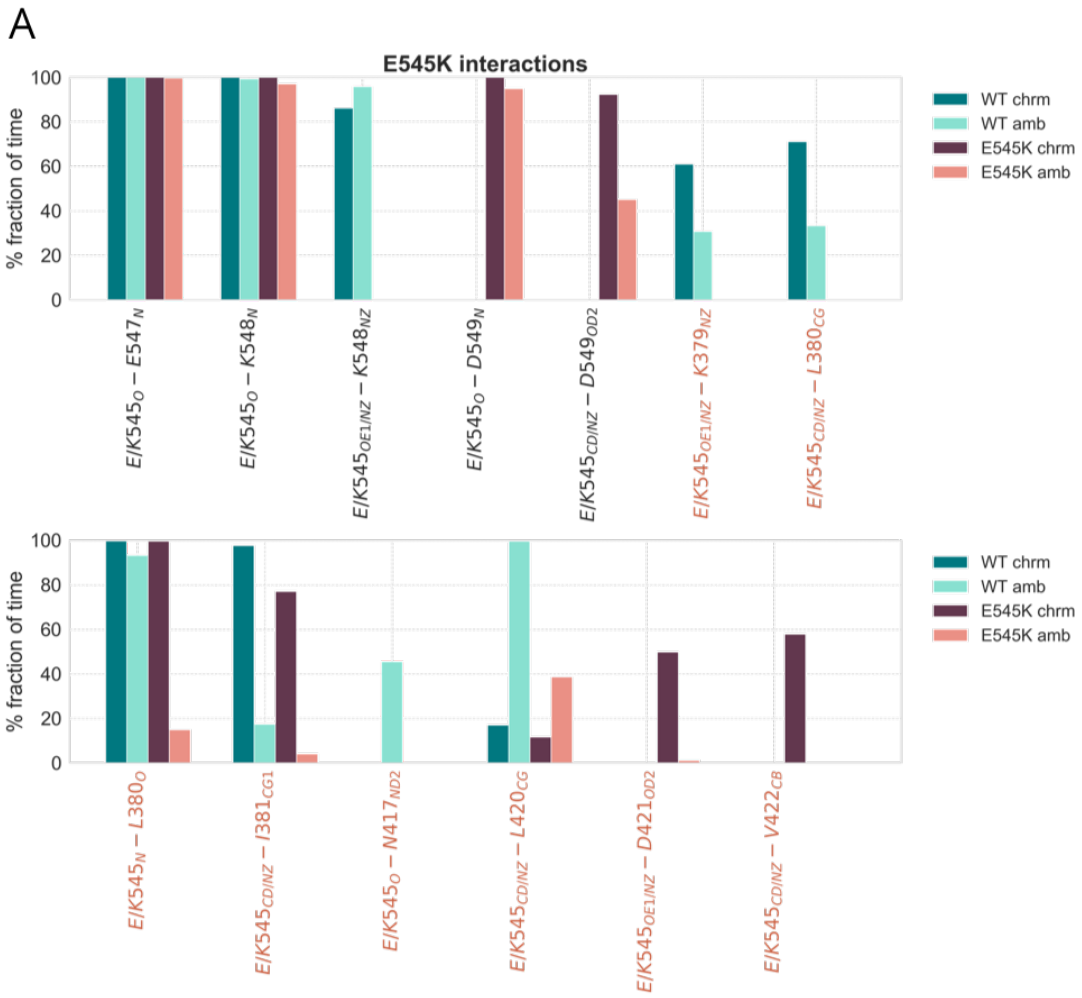
well without interacting with each other and therefore contribute simultaneously to a single combined reconstruction of the free energy surface (FES).

Given the size of the system (~1,600 amino acid long) and the significant computational resources needed to sample its conformations, we could not extend the simulations much beyond 400 ns/system. To explore the conformational changes in what, for a system of this size, is a relatively short time, the bias had to be deposited quickly and thus the resulting free energy surfaces are somewhat rough.

Each of the eight walkers was subjected to well-tempered metaD in which a Gaussian deposited on the collective variable space every 2 ps with height $W = W_o e^{-V(s,t)/(f-1)T}$, where W_0 = 5 kJ/mol is the Gaussian height, T is the temperature of each walker (310 K), f = 10 is the bias factor, and V(s,t) is the bias potential at time t and CV value s. The following two CVs were used: CV1, the distance between the centres of mass of the nSH2 and helical domains, and CV2, the distance in contact map space to the open state. The centres of mass of the nSH2 (residues Met322-Lys430, p85α) and helical (residues Arg524-Cys695, p110α) domains were defined using every second C_α carbon of the residues of each domain. The contact map distance was defined as $CV2(R) = 1/N \sum_{\gamma \in \Gamma}(D_\gamma(R) - D_\gamma(R_{open}))^2$, where $D_\gamma(R)$ is a sigmoidal function that measures the degree of formation of the contact γ in the structure R and is defined as $D_\gamma(R) = W_\gamma \frac{1-(r_\gamma/r_\gamma^0)^n}{1-(r_\gamma/r_\gamma^0)^m}$, in which r_γ is the contact distance in the structure R, r_γ^0 is the contact distance in either the reference open or closed conformation depending on which the contact γ is specific for, W_γ is the weight of the contact and is set to one for regular contacts, N is a normalization constant, n = 6, and m = 12. The set of contacts Γ defining the contact map distance for CV2 was determined using two structures that would be representative of the closed and the open states, respectively. The representative conformation for the closed state was chosen to be the resulting conformation obtained after the energy minimisation and equilibration of the WT, while the representative conformation for the open state was selected from our previous MD results in which the introduction of E545K resulted in a spontaneous detachment of the nSH2 domain from the helical domain[15]. Contacts selection was based on distance cutoff, i.e. a particular contact was selected if the separation of representative atoms fell to at or below 5 Å. For the representative atoms per residue see the definition adopted in TimeScapes Analytics Package[16]. Then, among all pairs in the closed or open conformation, only those pairs that specifically discriminate between the two structures were kept, i.e., only the pairs that appear either in one or the other discarding those pairs common to both. The final set, Γ, constituted of 65 regular contacts and eight salt-bridge contacts involving interactions between residues of the nSH2 and the helical and kinase domains. The contacts that corresponded to side chain salt-bridge interactions were weighed 3 times more than a regular contact. The widths of the Gaussians for the two CVs was set to σ_1 = 0.035 nm and σ_2 = 0.5.

Given the size of the interacting surfaces during the reattachment of nSH2 and the underlying complexity that is associated with non-specific binding when it comes to protein-protein interactions, we used a funnel-shaped potential[17] to confine the movement of the nSH2 domain after its detachment from the helical domain and aid in the convergence of the free energy. This approach has been applied successfully in the past to study the binding mechanism of peptide and non-peptide ligands to pharmacologically relevant targets[14]. The funnel-shaped potential was placed such that the detachment vector defined by the two centres of mass of the nSH2 and helical domain would align with a vector passing through the middle of the funnel (Fig. S9). Moreover, the size and shape of the funnel was such that would allow the centre of mass of the nSH2 domain to move both away from the centre of mass of the helical domain, as well as to explore different orientations around it.

4. Analysis


A
E545K interactions
% fraction of time
100
80
60
40
20
0
WT chrm
WT amb
E545K chrm
E545K amb
E/K545O – E547N
E/K545O – K548N
E/K545OE1/NZ – K548NZ
E/K545O – D549N
E/K545CD/NZ – D549OD2
E/K545OE1/NZ – K379NZ
E/K545CD/NZ – L380CG
E/K545N – L380O
E/K545CD/NZ – I381CG1
E/K545O – N417ND2
E/K545CD/NZ – L420CG
E/K545OE1/NZ – D421OD2
E/K545CD/NZ – V422CB

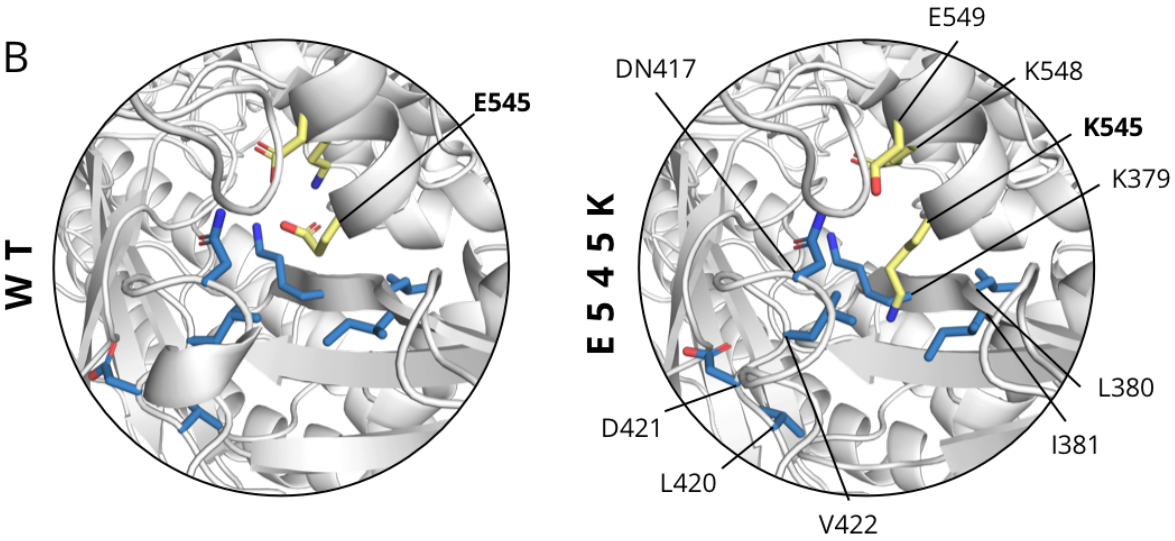

B
W T
E545
E 5 4 5 K
E549
DN417
K548
K545
K379
L380
I381
D421
L420
V422

Figure S1. (A) Interactions of E/K545 with neighbouring residues that are maintained over the course of the unbiased simulations for a fraction of time higher than 40% for at least one of the systems. The estimation of the % time fraction that the residues are considered to interact is based on distance criteria ($d < 0.7$ nm for hydrophobic interactions, $d < 0.35$ nm for hydrogen bonds, and $d < 0.4$ nm for salt-bridges). The representative side chain and backbone atoms of the interacting residues that have been considered for every interaction are denoted in subscript. Interactions between side chains of the same charge were considered not formed by default. Interactions between E/K545 and residues of the p110α domain are depicted in black, while interactions with residues of the p85α domain are depicted in orange. The estimation of the % fraction of interaction time with the Amber99SB-ILDN force field (amb label in the plot) refer to the detachment simulation described in ref. [15]. This dataset can be accessed at https://repo.vi-seem.eu/handle/21.15102/VISEEM-254 and https://repo.vi-seem.eu/handle/21.15102/VISEEM-253. (B) Graphical representation of the plotted interactions that E545 and K545 form. Residues that belong to the helical domain are depicted in yellow, while residues that belong to the nSH2 domain are depicted in blue sticks.

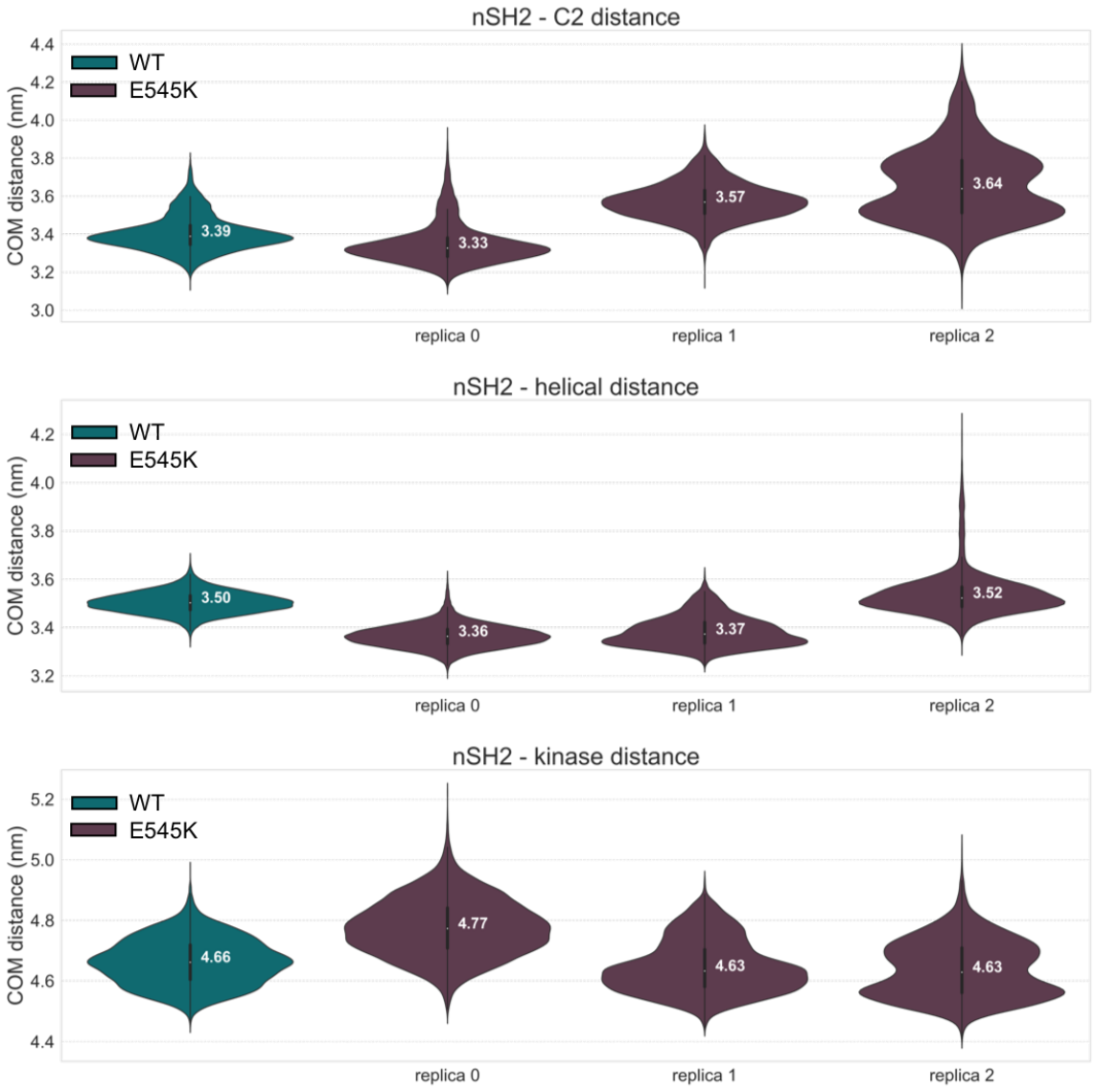


Figure S2. Distribution of the time evolution of the distance between the centres of mass of the nSH2 and C2 domains, the nSH2 and helical domains, and the nSH2 and kinase domains over the course of the unbiased simulations. The median value of each distribution is displaced next to the kernel density estimation of the each underlying distribution.

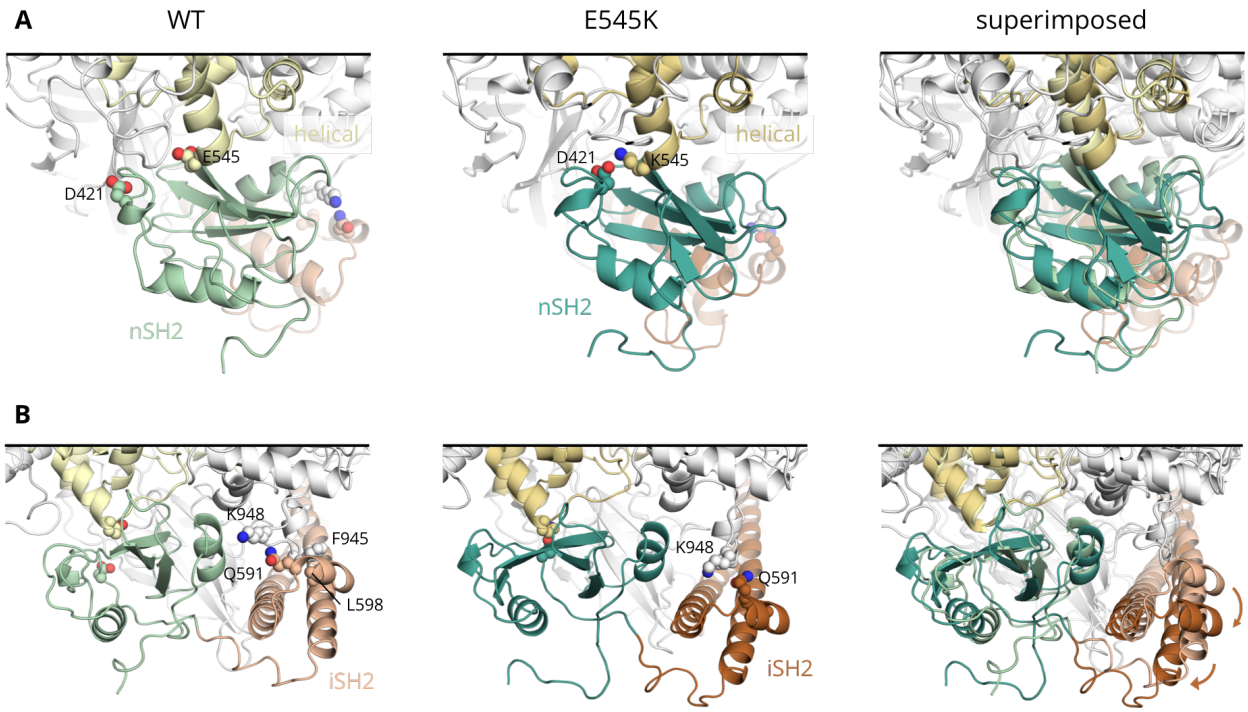
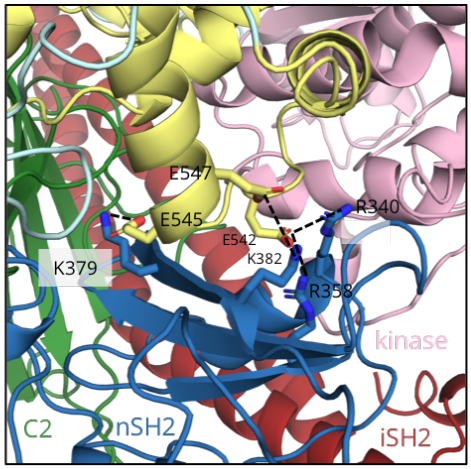


Figure S3. Representative structures of the unbiased MD simulations of the E545K with respect to the WT. (A) Close-up view around position 545, (B) Close-up view around the activation-loop and the iSH2 domain.

A

		WT	E545K
residue 1	residue 2	interaction occupancy %	
nSH2	**C2**		
Asp349(CG)	Arg357(CZ)	61.2	46.4
nSH2	**helical**		
Arg340(CZ)	Glu542(CD)	30.9	17.6
Arg358(CZ)	Glu542(CD)	65.1	0.0
Lys379(NZ)	Glu545(CD)	56.8	-
Lys382(NZ)	Glu547(CD)	43.9	77.9
Lys419(NZ)	Glu579(CD)	0.4	62.2
Asp421(CG)	Lys545(NZ)	-	48.4
nSH2	**kinase**		
Arg340(CZ)	Asp1017(CG)	0.4	31.4
Glu341(CD)	Lys942(NZ)	0.0	33.1
Glu342(CD)	Lys948(NZ)	74.1	4.0
nSH2	**iSH2**		
Glu345(CD)	Arg574(CZ)	46.3	15.3
Glu345(CD)	Lys575(NZ)	27.4	60.8

B



C

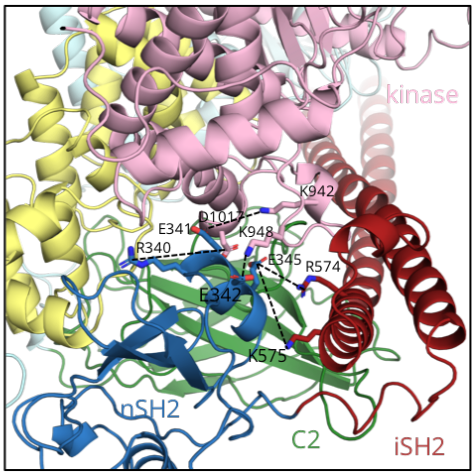


Figure S4. (A) Salt-bridge interactions on the interface of the nSH2 domain with the surrounding domains for replica 0, where the nSH2 domain slides around the helical domain. To avoid highlighting interactions that occur only sporadically in the unbiased simulations, only those salt-bridge interactions that are maintained for more than 30% of the simulation in either the WT or the E545K have been included in the table. The interacting atoms based on which distance the salt-bridge formation time was evaluated are included in the parentheses. The interactions in which the Glu545 and Lys545 are involved in have been highlighted (grey). (B) Graphical representation of the residues that are involved in the salt-bridge interactions between the nSH2 (blue) and the helical (yellow) or C2 (green) domains mapped on the structure of the WT. (C) Graphical representation of the residues that are involved in the salt-bridge interactions between the nSH2 and the kinase (pink) or iSH2 (red) domains in the WT. The crystal structure (PDB ID 4OVU) was used to depict the residues.

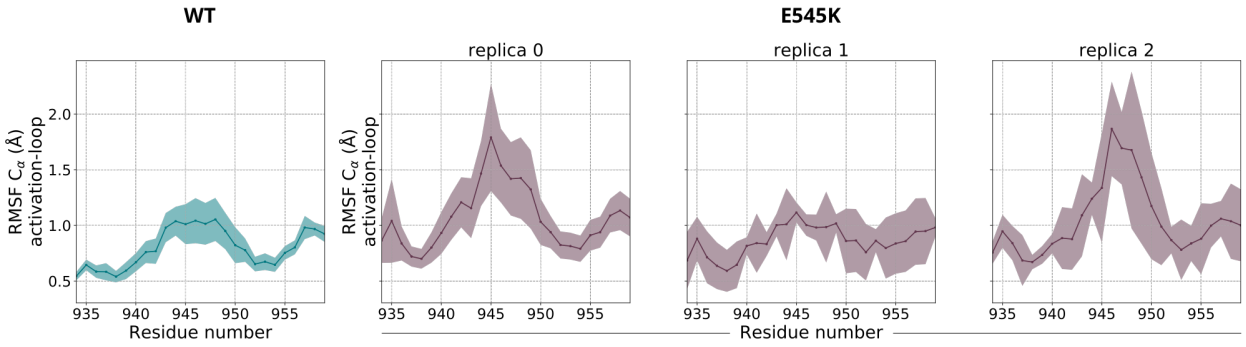


Figure S5. Root mean square fluctuation of the activation loop residues in the WT (cyan) and E545K (purple) PI3Kα of the kinase domain as seen in the unbiased simulations. The RMSF calculation was performed in overlapping time-windows and the average value is depicted with a solid line while the shading represents the deviation of the fluctuation per residue. Replicas 0 and 2 correspond to simulations, where the loop that connects helix B of nSH2 with the iSH2 domain pulls the iα3 helix of iSH2 away from the activation loop.

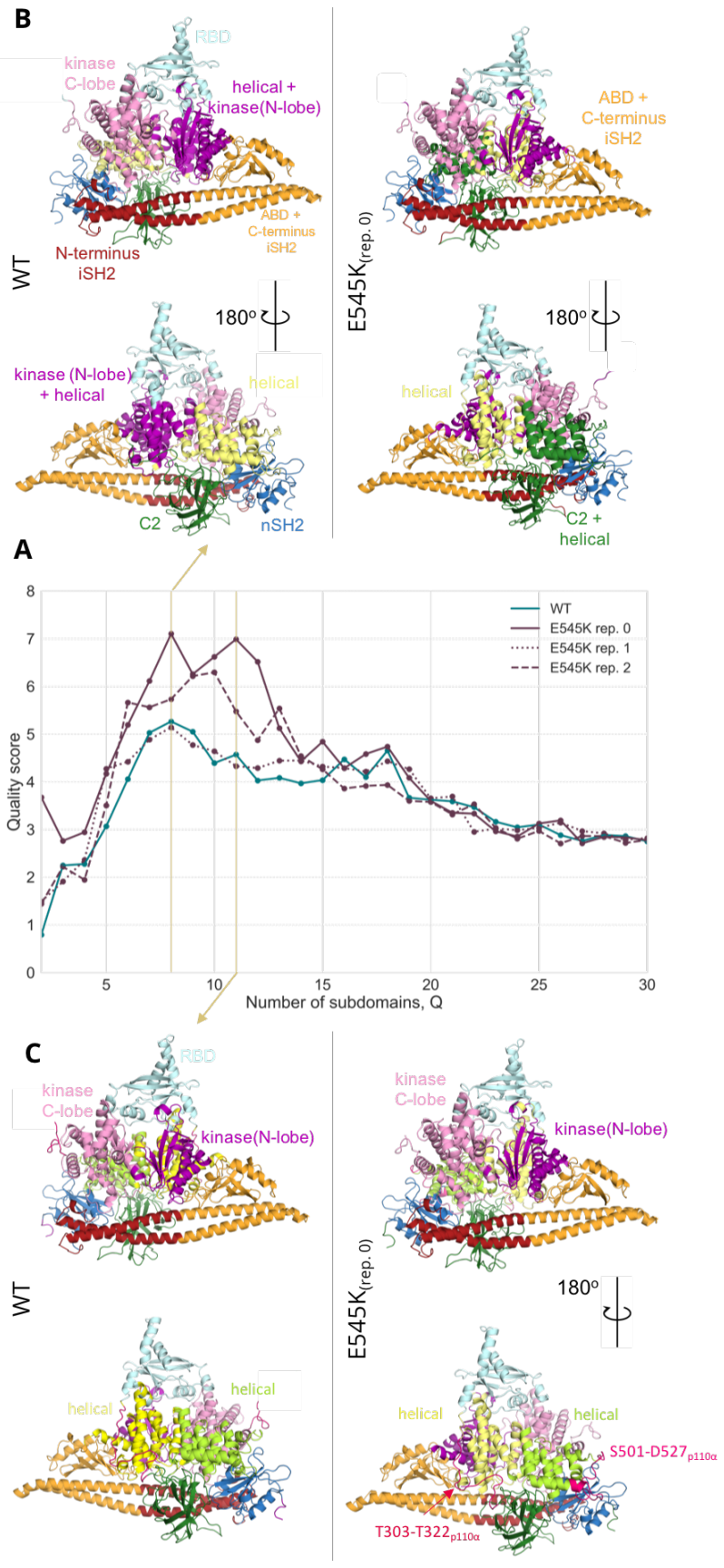

B
RBD
kinase
C-lobe
helical +
kinase(N-lobe)
ABD +
C-terminus
iSH2
N-terminus
iSH2
WT
180°
kinase (N-lobe)
+ helical
helical
C2
nSH2
E545K(rep. 0)
ABD +
C-terminus
iSH2
180°
helical
C2 +
helical
A
WT
E545K rep. 0
E545K rep. 1
E545K rep. 2
Quality score
Number of subdomains, Q
C
RBD
kinase
C-lobe
kinase(N-lobe)
WT
helical
helical
kinase
C-lobe
kinase(N-lobe)
E545K(rep. 0)
180°
helical
helical
S501-D527p110α
T303-T322p110α

Figure S6. Dynamical domain decomposition analysis based on the unbiased MD simulations. Residues that belong to the same domain as mapped by the Dynamical domain decomposition analysis are depicted in the same colour and exhibit highly correlated movement. (A) Quality score with respect to the number of domains. This score quantifies the significance of the subdomain decomposition based on the balance of intra- and inter-domain distance fluctuations. (B) Graphical representation of the subdomains that corresponds to the first maximum of the quality score (Q=8). A similar to Fig. 1 colour code of the domains has been used, where possible. (C) Graphical representation of the subdomains that correspond to the second maximum of the quality score (Q=11).

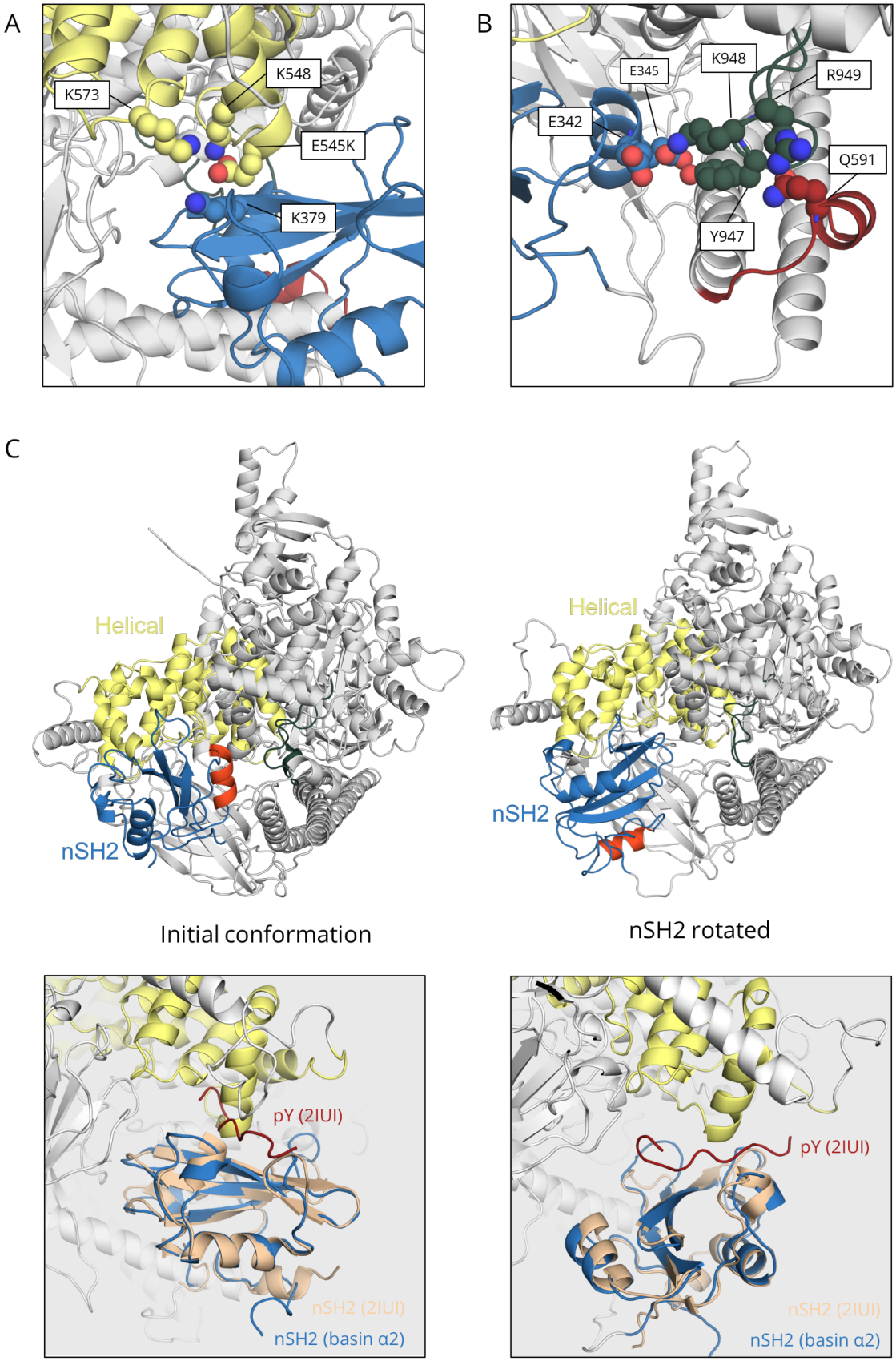

A
K573
K548
E545K
K379
B
E345
K948
R949
E342
Q591
Y947
C
Helical
nSH2
Initial conformation
Helical
nSH2
nSH2 rotated
pY (2IUI)
nSH2 (2IUI)
nSH2 (basin α2)
pY (2IUI)
nSH2 (2IUI)
nSH2 (basin α2)

Figure S7. (A) Close-up of the interactions in which Glu545 participates in basin α1 of the free energy surface of the WT. Residues that belong to the nSH2 domain are depicted in blue spheres, while residues that belong to the helical domain are depicted in yellow spheres. (B) Close-up of the interactions that stabilise the activation loop in the inactive conformation in basin α1 of the free energy surface of the WT. Residues that belong to the nSH2 domain are depicted in blue spheres, residues that belong to the activation-loop of the kinase domain are depicted in dark green spheres, while residues that belong to the iα3 helix of the iSH2 domain are depicted in red spheres. (C) Representative conformation of PI3Kα from basin α2 of the free energy surface of the WT, with respect to the initial conformation of the simulation. Helix A (residues 339–347) of the nSH2 domain (blue) is depicted in orange to showcase the rotation of the domain relative to the helical domain (yellow). Superposition of the crystal structure of the nSH2 domain (beige) bound to a PDGFR phosphopeptide (pY red, PDB ID 2IUI) with the two conformations is presented on the bottom to showcase the better accessibility of the nSH2-rotated conformation to pY binding.

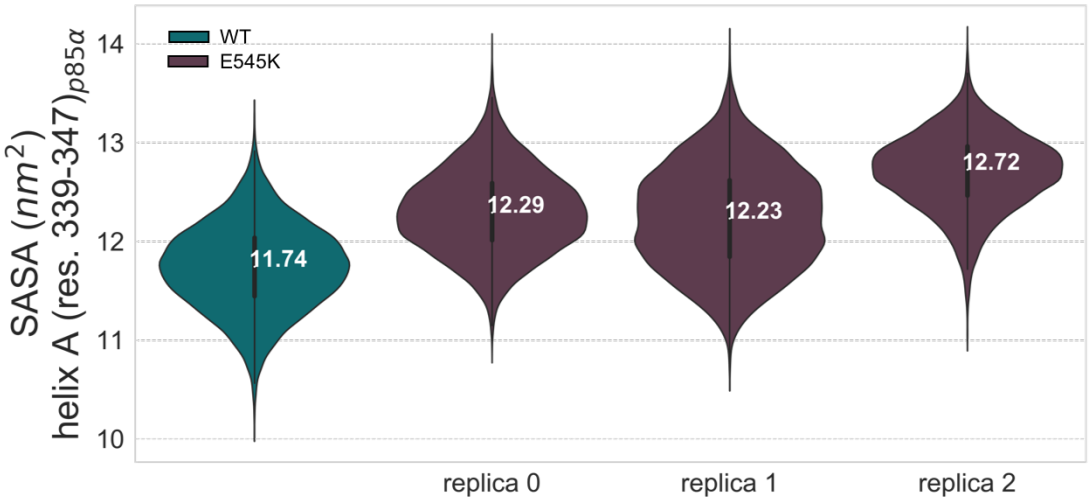


Figure S8. Distribution of the time evolution of the solvent accessible surface area of helix A of the nSH2 domain of the WT (cyan) and E545K (purple) PI3Kα over the course of the unbiased simulations.

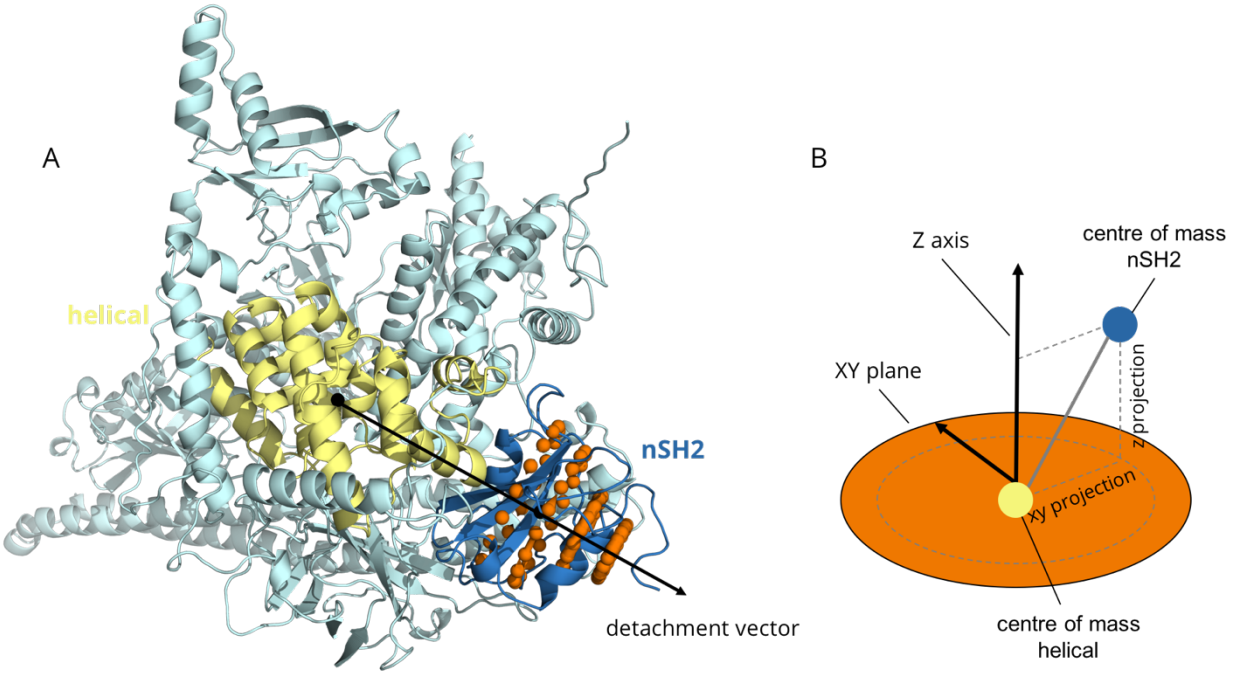


Figure S9. (A) Graphical representation of the funnel-shaped potential (orange balls) that was used to confine the movement of the centre of mass of the nSH2 domain (blue) with respect to the helical domain (yellow); the centres of mass of the two domains were used to define a detachment vector. (B) The PI3Kα complex was rotated in every step during the course of the simulation in a way that the unbinding vector would coincide with the Z-axis. The shape of the funnel was such that allowed nSH2 to sample both the closed and open conformations in which the nSH2 domain had completely detached from the helical domain.

Bibliography

1 M. S. Miller, O. Schmidt-Kittler, D. M. Bolduc, E. T. Brower, D. Chaves-Moreira, M. Allaire, K. W. Kinzler, I. G. Jennings, P. E. Thompson, P. A. Cole, L. M. Amzel, B. Vogelstein and S. B. Gabelli, Structural basis of nSH2 regulation and lipid binding in PI3Kα, *Oncotarget*, 2014, **5**, 5198–5208.

2 T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations, *Nucleic Acids Res.*, 2004, **32**, 665–667.

3 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*, 2015, **1–2**, 19–25.

4 M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia and M. Parrinello, PLUMED: A portable plugin for free-energy calculations with molecular dynamics, *Comput. Phys. Commun.*, 2009, **180**, 1961–1972.

5 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, PLUMED 2: New feathers for an old bird, *Comput. Phys. Commun.*, 2014, **185**, 604–613.

6 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot, H. Grubmüller and A. D. MacKerell, CHARMM36m: An improved force field for folded and intrinsically disordered proteins, *Nat. Methods*, 2016, **14**, 71–73.

7 A. Kuzmanic, R. B. Pritchard, D. F. Hansen and F. L. Gervasio, Importance of the Force-Field Choice in Capturing Functionally Relevant Dynamics in Von Willebrand Factor, *J. Phys. Chem. Lett.*, 2019, acs.jpclett.9b00517.

8 P. Robustelli, S. Piana and D. E. Shaw, Developing a molecular dynamics force field for both folded and disordered protein states, *Proc. Natl. Acad. Sci.*, 2018, **0260**, 201800690.

9 H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

10 G. Bussi, D. Donadio and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.*, 2007, **126**, 014101.

11 M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

12 P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti and M. Parrinello, Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics, *J. Phys. Chem. B*, 2006, **110**, 3533–3539.

13 G. Saladino, L. Gauthier, M. Bianciotto and F. L. Gervasio, Assessing the Performance of Metadynamics and Path Variables in Predicting the Binding Free Energies of p38 Inhibitors, *J. Chem. Theory Comput.*, 2012, **8**, 1165–1170.

14 N. Saleh, G. Saladino, F. L. Gervasio, E. Haensele, L. Banting, D. C. Whitley, J. Sopkova-de Oliveira Santos, R. Bureau and T. Clark, A Three-Site Mechanism for Agonist/Antagonist Selective Binding to Vasopressin Receptors, *Angew. Chemie - Int. Ed.*, 2016, **55**, 8008–8012.

15 H. Leontiadou, I. Galdadas, C. Athanasiou and Z. Cournia, Insights into the mechanism of the PIK3CA E545K activating mutation using MD simulations, *Sci. Rep.*, 2018, **8**, 15544.

16 W. Wriggers, K. A. Stafford, Y. Shan, S. Piana-Agostinetti, P. Maragakis, K. Lindorff-Larsen, P. J. Miller, M. P. Eastwood, R. O. Dror and D. E. Shaw, Automated Event Detection and Activity Monitoring in Long Time-Scale Molecular Dynamics, *Biophys. J.*, 2009, **96**, 364a-

365a.

17 V. Limongelli, M. Bonomi and M. Parrinello, Funnel metadynamics as accurate binding free-energy method, *Proc. Natl. Acad. Sci.*, 2013, **110**, 6358–6363.