Computational and Structural Biotechnology Journal 23 (2024) 3118–3131

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Free energy landscape of the PI3Kα C-terminal activation

Danai Maria Kotzampasi [a,b], Michail Papadourakis [a], John E. Burke [c,d], Zoe Cournia [a,*]

[a] Biomedical Research Foundation, Academy of Athens, Athens 11527, Greece
[b] Department of Biology, University of Crete, Heraklion 71500, Greece
[c] Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC V8W 2Y2, Canada
[d] Department of Biochemistry and Molecular Biology, The University of British Columbia, Vancouver, BC V6T 1Z3, Canada

ARTICLE INFO

Keywords:
PI3Kα
PIK3CA
H1047R
Allostery
HDX-MS
Cancer
Molecular Dynamics simulations
Lipid kinase

ABSTRACT

The gene *PIK3CA*, encoding the catalytic subunit p110α of PI3Kα, is the second most frequently mutated gene in cancer, with the highest frequency oncogenic mutants occurring in the C-terminus of the kinase domain. The C-terminus has a dual function in regulating the kinase, playing a putative auto-inhibitory role for kinase activity and being absolutely essential for binding to the cell membrane. However, the molecular mechanisms by which these C-terminal oncogenic mutations cause PI3Kα overactivation remain unclear. To understand how a spectrum of C-terminal mutations of PI3Kα alter kinase activity compared to the WT, we perform unbiased and biased Molecular Dynamics simulations of several C-terminal mutants and report the free energy landscapes for the C-terminal "closed-to-open" transition in the WT, H1047R, G1049R, M1043L and N1068KLKR mutants. Results are consistent with HDX-MS experimental data and provide a molecular explanation why H1047R and G1049R reorient the C-terminus with a different mechanism compared to the WT and M1043L and N1068KLKR mutants. Moreover, we show that in the H1047R mutant, the cavity, where the allosteric ligands STX-478 and RLY-2608 bind, is more accessible contrary to the WT. This study provides insights into the molecular mechanisms underlying activation of oncogenic PI3Kα by C-terminal mutations and represents a valuable resource for continued efforts in the development of mutant selective inhibitors as therapeutics.

1. Introduction

The class I phosphoinositide 3-kinase alpha isoform (PI3Kα) is a lipid kinase that mediates activation of the protein kinase B (Akt)/mTOR signaling pathway, playing fundamental roles in cell proliferation, growth, and survival. When PI3Kα binds to activated Receptor Tyrosine Kinases (RTKs) and activated RAS on cell membranes, it phosphorylates the 3′-position hydroxyl group of the inositol ring of phosphatidylinositol (4,5)-bisphosphate (PIP_2) to produce phosphatidylinositol (3,4,5)-trisphosphate (PIP_3) [1,2]. PIP_3 acts as a second messenger activating numerous membrane signaling molecules and kinases, and is thus regulating protein synthesis, cell growth, and proliferation [3].

PI3Kα in the cell predominantly exists as a heterodimer; its catalytic p110α subunit binds to one of five possible regulatory subunits (p85α, p55α, p50α, p85β, and p55γ), with the most well characterized heterodimer being p110α-p85α (Fig. 1A) [4]. Both the catalytic and regulatory subunits consist of multiple domains, with p110α containing an adaptor binding domain (ABD) that mediates binding to the iSH2 coiled-coil domain of regulatory subunits, a RAS binding domain (RBD) that mediates membrane-bound RAS recruitment, a membrane-binding C2 domain (C2), a helical scaffolding domain, which interacts with the nSH2 domain of the regulatory subunits, and a bi-lobal kinase domain, which includes catalytically important regions such as the catalytic and activation loops, membrane binding loops, and the ATP and PIP_2 binding pockets (Fig. 1B). The regulatory subunit, p85α, contains two Src homology 2 domains, nSH2 and cSH2, that bind to RTKs and their adaptor proteins with a coiled-coil domain (iSH2) between them, which mediates the high-affinity interaction to the catalytic subunit, p110α. These domains are preceded by an Src homology 3 domain (SH3), a breakpoint cluster region-homology domain (BH), and two proline-rich regions, PR1 and PR2, surrounding the BH domain (Fig. 1A). In all available crystal structures and recent cryo-EM structures of apo PI3Kα, the SH3, BH, and cSH2 domains remain elusive possibly due to their dynamic nature, although in the cryo-EM experiments of Liu et al. [5], it was shown that in PI3Kα inhibited by BYL-719, regions of low-resolution electron density are in contact with the kinase domain blocking access to the catalytic site. Intriguingly, the cryo-EM structure of an activated PI3Kα state indicated that upon activation, the ABD domain and the

* Corresponding author.
E-mail address: zcournia@bioacademy.gr (Z. Cournia).

https://doi.org/10.1016/j.csbj.2024.07.010
Received 27 February 2024; Received in revised form 5 July 2024; Accepted 5 July 2024
Available online 8 July 2024

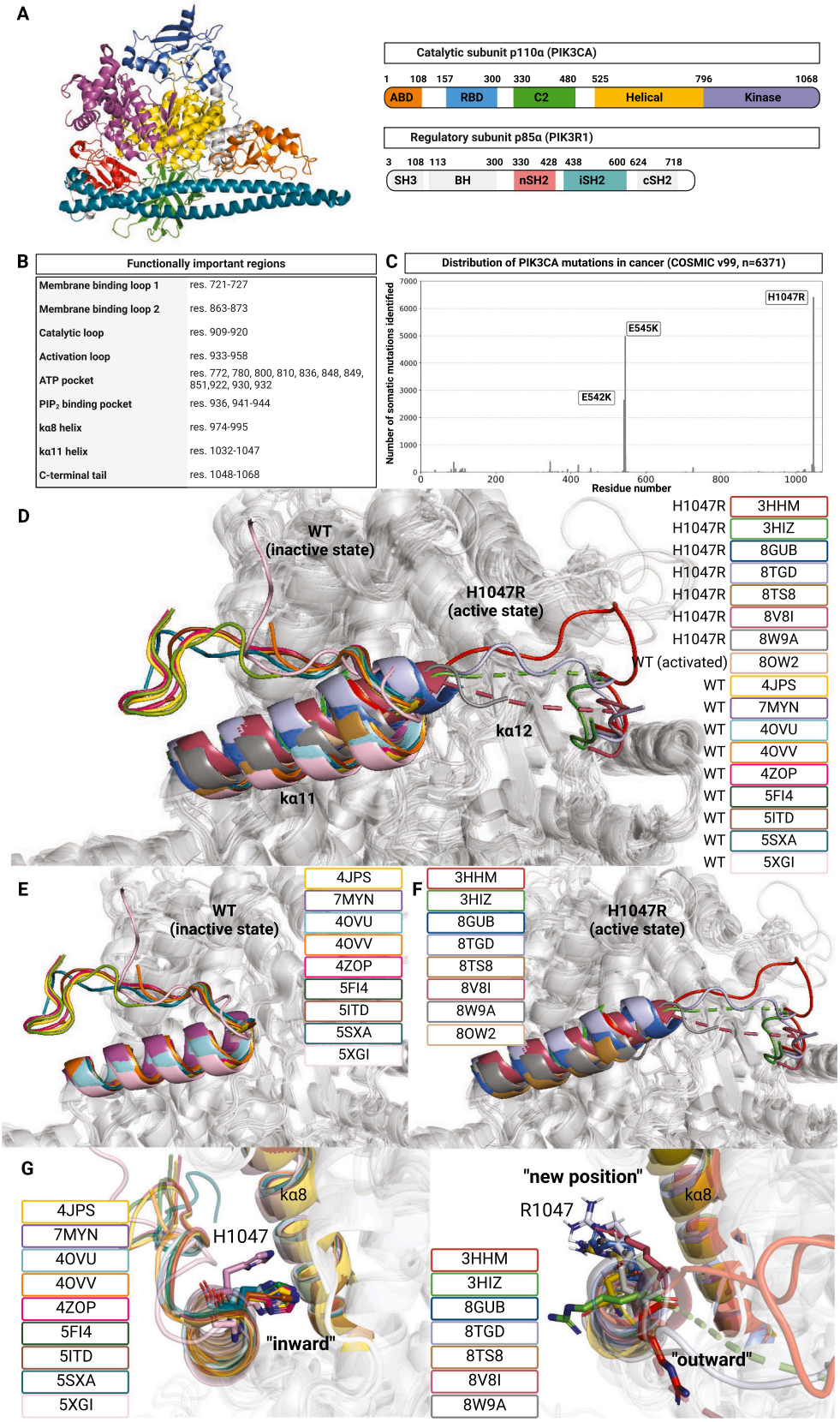

Fig. 1. Analysis of published X-ray and cryo-EM structures of PI3Kα (including both heterodimer and catalytic subunit alone). (A) The cryo-EM structure of PI3Kα heterodimer (7MYN) with its domains annotated according to the schematic [5]. (B) Functionally important residues of PI3Kα. (C) Frequency of somatic mutations in p110α versus residue number according to the COSMIC database [6]. (D) PDB structures of the C-terminal tail of PI3Kα. Two different conformations of the C-terminus have been observed, the closed (E) and the open conformation (F), with C-terminal opening being accompanied by rotation of the kα11 helix (G) exhibiting three different orientations for amino acid 1047, the "inward", "outward" (PDB IDs: 3HHM, 3HIZ), and "new position" (PDB IDs: 8GUB, 8TGD, 8TS8, 8V8I, 8W9A).

p85α regulatory subunit are mobile relative to the rest of the catalytic subunit and they cannot be observed in obtained images of the p110α catalytic core [5].

Activating oncogenic mutations that are present in many types of cancer are located throughout the sequence of *PIK3CA*, the gene encoding p110α, but concentrate in two hotspots: E542K and E545K are positioned at the interface of the p110α helical domain with the p85α nSH2 domain [7,8] and H1047R is at the C-terminus of the kinase domain, which interacts with the cell membrane when the kinase is activated (Fig. 1C) [9,10]. As the active state of PI3Kα is recruited to the cell membrane [11,12], it is not surprising that mutations leading to enhanced membrane binding of PI3Kα are associated with multiple human cancers [2,6,13]. In fact, the H1047R mutant is the most prevalent mutation in human cancers (>5,300 in the COSMIC database [6]). Other mutations may occur in the vicinity of H1047R such as M1043L (>10 in the COSMIC database, with >100 of the similar M1043I mutant [6]), G1049R (>100 mutants COSMIC database [6]), and a C-terminal frameshift mutation in which the C-terminal N1068 residue is replaced by residues KLKR [14] (>10 in the COSMIC database [6]) (Fig. 1C). While it is known that most C-terminal kinase mutations increase membrane binding [9,11,12,15,16], the exact molecular mechanism(s) by which these oncogenic mutants at the C-terminus promote the active conformation of PI3Kα are still not fully resolved.

After comparing PI3Kα structures available in the Protein Data Bank (PDB) (Fig. 1), we observe two distinct conformations of the C-terminus. In the WT structures (PDB IDs 4JPS [17], 4WAF [18], 4OVU and 4OVV [19], 4ZOP, 5XGI, 5FI4 [20], and 5SXA [21]), the C-terminus is in a closed (inactive) state making extensive contacts with helices kα8 and kα11 and orienting the membrane-binding WIF motif away from the membrane-binding interface (Fig. 1D,E). In the H1047R mutant structures (PDB IDs 8TGD [22], 8GUB [23], 3HHM and 3HIZ [16], 8TS8 [24], 8V8I [25], 8W9A [26] and the structure of the activated PI3Kα WT state, 8OW2 [27]), the C-terminus is in an entirely different orientation; an open conformation, where contacts between the C-terminus and helices kα8 and kα11 are absent and the WIF motif (residues 1057–1059) is oriented in the direction of the putative membrane-binding interface (Fig. 1D,E,F). Thus, the H1047R mutation may lead to the rearrangement of the membrane-binding C-terminus, which may alter membrane recruitment. Furthermore, structures of PI3Kα in complex with allosteric inhibitors that are in the vicinity of the H1047R mutation, such as STX-478 [22], RLY-2608 [24] and QR-7909 [26], also exhibit an open conformation of the C-terminus (Fig. 1D,F). In the cryo-EM structures of PI3Kα H1047R [23,26], the C-terminal region of H1047R (residues 1050–1062) is unresolved in contrast to the fixed orientation that they present in crystallographic structures due to crystallographic contacts (e.g. in PDB IDs 3HHM, 3HIZ, 8TGD, 8V8I). The intrinsic flexibility of the C-terminal region could be important in mediating an enhanced dynamic engagement with the cell membrane through the lipid-binding motif WIF (residues 1057-1059) [12].

The cryo-EM structure of H1047R PI3Kα inhibited by BYL-719 (PDB ID: 8GUB) was the first to reveal a notable difference compared to previous X-ray structures: the H1047R mutant rotates its sidechain to a new position in 8GUB [23] and in the more recently published H1047R structures with PDB IDs 8TGD [22], 8TS8 [24], 8V8I [25], and 8W9A [26] compared to previous structures with PDB IDs 3HHM and 3HIZ [16] (Fig. 1G). The side chain of R1047 rotates and interacts with T972, F977 and Q981 , in contrast with the WT H1047 that interact with V955 and L956. Therefore, the R1047 positioning weakens the interface between kα11 (residues 1031–1050) and the activation loop (residues 954–956). When comparing all available H1047R structures, the kα11 helix axis of the mutant is tilted 4.16° ± 0.73° with respect to the WT structure (PDB ID 4OVU used as a reference, see the SI for more information), which further disrupts the interface between the kα11 and kα8 helices (Fig. 1D). This interface disruption may be the reason why the Hydrogen/Deuterium (H/D) exchange rate of the kα8 helix (residues 962–980) is increased by 8.32 % in the H1047R mutant compared to PI3Kα WT [15]. Hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments in solution also showed a significant increase of 19.53 % in the H/D exchange rate of the PI3Kα H1047R C-terminal tail residues 1039–1055 compared to the WT [11]. Moreover, we have previously found that both H1047R and G1049R exhibit a ~75 % increase in the membrane binding and similar conformational changes, with distinct differences from M1043L and the N1068KLKR frameshift [15]. This finding suggests that different C-terminal mutants cause differential conformational changes, resulting in altered membrane binding, which may indicate different mechanisms of activation. Therefore, structural changes in the vicinity of the C-terminal "opening" may offer opportunities for isoform- and mutant-specific allosteric modulation of PI3Kα.

Although several Molecular Dynamics (MD) simulations of PI3Kα have been published over the past decade [28,7–10,24,29], only few studies have investigated the mechanism of PI3Kα C-terminal activation [9,24]. Gkeka et al. proposed a series of events that lead to the overactivation of the H1047R using SPR experiments and MD simulations of the p110α. The proposed mechanism of overactivation due to the H1047R PI3Kα mutant, includes the loss of the C-terminal autoinhibitory role, which in the WT protein controls the DRH motif to limit its access to the catalytic site [9]. The weakening of this role in the H1047R mutant through loss of intermolecular interactions is proposed as a plausible explanation of the elevated kinase activity of the enzyme [9]. Moreover, in the recently published study by Varkaris et al., where unbiased MD simulations on full-length PI3Kα and p110α alone were performed, the C-terminal tail remains stable in a closed state throughout the simulation of the full-length PI3Kα, while when only the p110α is simulated, the C-terminus begins to disengage [24].

In the present study, we investigate the mechanistic details of the "closed-to-open" C-terminal transition required for PI3Kα activity on membranes as revealed by enhanced sampling MD simulations. More specifically, we study the structural and dynamical effects of the oncogenic mutants M1043L, H1047R, G1049R, and a C-terminal frameshift, N1068KLKR on the C-terminal transition using biased and unbiased MD, compared to the WT PI3Kα. We unravel the free energy landscape for the C-terminal "closed-to-open" transition, describe conformational changes that occur across the different systems, compare with available experimental results, and quantify the interactions of the C-terminus in the closed and open state in the WT and mutant PI3Kα. Our findings enable a molecular explanation why H1047R is the most prevalent oncogenic mutation based on the fact that this mutant (and G1049R) reorients the C-terminus with a different mechanism compared to the WT and M1043L and N1068KLKR mutants. Moreover, the existence of the cryptic pocket, where the PI3Kα allosteric inhibitors, STX-478 [22] and RLY-2608 [24] bind, was investigated in the retrieved structures from the biased MD simulations. This work provides valuable insights into the molecular mechanisms underlying the C-terminal activation in PI3Kα, and reveals critical insights into conformational changes that may be exploited for design of mutant specific anti-cancer therapeutics.

2. Materials and methods

2.1. Model construction of the full-length PI3Kα and ΔABD p110α systems

For the construction of the full-length PI3Kα models, the previously published PI3Kα model by Galdadas et al. [7] was used, which is based on the crystal structure with PDB ID 4OVU. Using this model, we mutated the kinase residues M1043L, H1047R, G1049R, and N1068KLKR to generate the models of the mutant proteins (see the SI). The protonation state of each residue was determined using the PROPKA3.0 [30,31] algorithm at pH= 7, optimization of the hydrogen bond network was performed using the Protein Preparation Wizard in Schrödinger suite 2022–2 [32], and the systems were solvated with TIP3P water molecules and neutralized with HTMD [33].

While we were performing the simulations of the full-length PI3Kα, a cryo-EM study was of the phosphopeptide-bound structure of PI3Kα was published, revealing a stable core containing all p110α domains except the ABD domain oand without p85α subunit [5]. Therefore, we built a model related to the PI3Kα active state (ΔABD p110α) based on the cryo-EM structure 7MYN [5], by removing the ABD domain and the regulatory subunit. The missing residues, except C-terminal residues 1051–1068, were added using the previously published PI3Kα model by Galdadas et al. [7]. The C-terminal residues 1051–1068 were created and refined using loop modeling with the Prime software of the Schrödinger suite 2022–2 [34,35]. The next step was to mutate the kinase residues M1043L, H1047R, G1049R, and N1068KLKR to generate the models of the PI3Kα mutant proteins. The five generated all-atom models were solvated with TIP3P water molecules and neutralized with HTMD [33]. Detailed information about the model construction of the full-length PI3Kα and ΔABD p110α models can be found in the SI.

2.2. Setup of the unbiased MD simulations

The unbiased MD trajectories for the full-length PI3Kα and the ΔABD p110α systems were generated with ACEMD3 package [36], using the CHARMM36m [37] all-atom force field for protein atoms and the TIP3P potential for modeling water molecules. Three independent simulations for each protein were performed starting from different conformations (see the SI for more information). Prior to MD simulations, all systems were relaxed by 10,000 steps of energy minimization using the steepest descent algorithm of ACEMD3 [36]. Then, the systems were equilibrated using ACEMD3 first in the NVT ensemble for 5 ns with the Langevin leap-frog integrator and the Langevin thermostat with a damping constant of 1 ps^{-1} at 310 K with restraints on heavy atoms (0.1 kcal/mol/$Å^2$). A second equilibration followed in the NPT ensemble for 10 ns with the Langevin leap-frog integrator, a Monte Carlo isotropic barostat [38] to keep the pressure constant at 1 bar andthe Langevin thermostat [39] with damping constant of 0.1 ps^{-1} to keep the temperature at 310 K with restraints on protein heavy atoms (1 kcal/mol/$Å^2$ for the first 5 ns). All bonds containing hydrogen atoms were constrained using the M-SHAKE algorithm [40]. The hydrogen mass was repartitioned to 4.0 au and a timestep of 4 fs was used [41]. The cut-off distance for van der Waals and electrostatic interactions was set at 12 Å with a switching function applied beyond 10 Å. Long-range electrostatic interactions were handled using the particle-mesh Ewald summation method [42,43]. After equilibration, the production runs were performed under constant pressure, temperature, and number of particles for three independent replicas:3 μs each for the full-length PI3Kα and 1.3 μs each for the ΔABD p110α systems using the same algorithms and parameters as in the second equilibration phase described above.

2.3. Analysis of the unbiased MD trajectories

The analysis of the retrieved unbiased MD trajectories was performed using GROMACS tools v2021.6 [44,45]. The last 2.5 μs of each replica were used for the full-length PI3Kα simulations and the last 1 μs of each replica was used for the ΔABD systems. To validate our findings, we compare the results from MD simulations with existing HDX-MS experimental data [15] using the python package HDXer [46,47]. Because the HDX-MS experiments were conducted at 293 K and pD = 7.9 [15], these parameters were incorporated into the predictive model of HDXer. We used the independent replica simulations to predict H/D exchange rates for all available peptide fragments at 3, 30 and 300 s. The computational values of the H/D exchange rates of the full-length PI3Kα WT system and ΔABD p110α WT were then compared to the experimental ones [15] as described in the SI. Clustering of the trajectories was performed using the GROMACS tool gmx cluster, with the gromos method and a cutoff of 2.2 Å [48]. The convergence of our simulations was evaluated using (i) Root Mean Square Deviations (RMSD) on all protein Cα atoms and the PI3Kα active site Cα atoms and (ii) Principal Component Analysis (PCA) on the C-terminal Cα atoms with the scikit-learn python package [49] as described in the SI. Dynamical network analysis [50] was performed using VMD [51] to calculate allosteric communication metrics (see the SI for more information).

2.4. Setup and analysis of the enhanced sampling simulations

Subsequently, we performed multiple-walkers metadynamics [52] simulations with a total of 8 walkers to enhance the sampling and characterize the effect of the oncogenic mutations M1043L, H1047R, G1049R, and N1068KLKR on the "closed-to-open" C-terminal transition and calculate the corresponding free energy surfaces (FESs). These simulations were conducted withGROMACS 2021.6 MD engine [44,45] compiled with the PLUMED 2.8 plug-in [53]. Bootstrap analysis was performed on the generated FES. We sample 1,000 times from the FES and based on this sampling, we calculate mean values and standard deviations (std) for the free energy of the closed, open and transition state [54,55]. The accuracy of the metadynamics method for calculating the free energy estimates cannot be quantitatively assessed because of a combination of factors, including the influence of the force field and the adequate sampling to ensure statistical significance in the analysis. Metadynamics simulations were performed for both the full-length PI3Kα and the ΔABD p110α systems until convergence (see the SI for metadynamics convergence calculations). Bias potential is deposited by means of Gaussians acting on a few degrees of freedom named collective variables (CVs) that bias the system towards a desired direction. CV1 is the distance in contact map space from a reference state in which the C-terminal tail is the closed state with respect to the open state and CV2 is the RMSD of the Cα carbon atoms of kα12 starting from the closed state (see SI for more details on the setup). The FESs of the WT and mutants were reconstructed as a function of these two CVs. Using the R package Metadynminer, the FES was estimated as a negative imprint of the added hills. Using the same package, the free energy minima and the transition path along with the transition state were extracted for each system [56]. The FESs for each system were adjusted so that the global free energy minimum was set to zero, which was then offset by the calculated mean values of each free energy minimum (see the SI). Moreover, we calculated the relative occupancy of native contacts in the basins of the closed, the open and the transition states using the CPPTRAJ module [57] as described in the SI. The Euler angles α, β, and γ of the kα11 helix were calculated using the GROMACS tool gmx rotmat to describe the orientation of the kα11 helix of each structure with respect to the kα11 helix of the crystal structure of PI3Kα WT apo with PDB ID 4OVU [19].

All methods are presented in detail in the Supplementary Information.

3. Results

3.1. Unbiased MD simulations of the full-length PI3Kα and ΔABD p110α

3.1.1. Model validation against HDX-MS data

HDX-MS is a powerful tool for investigating conformational changes that occur in different states of PI3Kα and how oncogenic mutations affect its conformation. To validate our models, we compared the results from our MD simulations against available HDX-MS experimental data, which provide the H/D exchange rates of the full-length PI3Kα WT and the ΔABD p110α WT, and the difference between them. Here, we used the python package HDXer to predict H/D exchange rates for all the available peptide fragments at 3, 30 and 300 s from biomolecular simulations [46,58–60,47]. We used the available PI3Kα peptide fragments and compared our simulation results against experimental data for the full-length PI3Kα WT and the ΔABD p110α WT (for a full description of the methodology see the Methods section and the SI) [15].

There are several important caveats to this analysis, primarily that

the experimental HDX-MS data was all generated as relative HDX-MS data, as for a large complex like PI3Kα, it can be extremely challenging to generate absolute H/D exchange rates and the generation of a fully deuterated PI3Kα sample is not possible because PI3Kα aggregates. Therefore, the comparisons here were qualitative in nature, and were used as a rough validation of the MD simulation versus our previous experimental data. Indeed, the mean deviation between experimental and computational H/D exchange rates is 17.1 % with std 14.8 % for the full-length PI3Kα system and 16.7 % with std 13.4 % for the ΔABD p110α system. The mean deviation is computed by calculating the mean % deuteration for the different time points for each peptide fragment and the deviation between the experimental and the computational data. In general, we observe that the predictions are constantly higher than the experimental values, which is consistent with the expected 15–40 % back exchange that would be expected in the HDX-MS analysis.

Our results align approximately to our previous data [15] as shown in Fig. 2, in Table S1, and Figs. S7-S8. In particular, for the peptide fragment 120–127 of the ABD-RBD linker, the calculated percentage of H/D exchange rate increase in ΔABD p110α is up to 22.5 %, while experimental data showed an increase up to 35.5 %. These residues become more exposed upon the disengagement of the ABD domain in the ΔABD p110α system. For residues 444–475 of the C2 domain, the calculated percentages of H/D exchange rate increase in ΔABD p110α up to 5.5 %, while experimental data showed an increase up to 10.8 %.

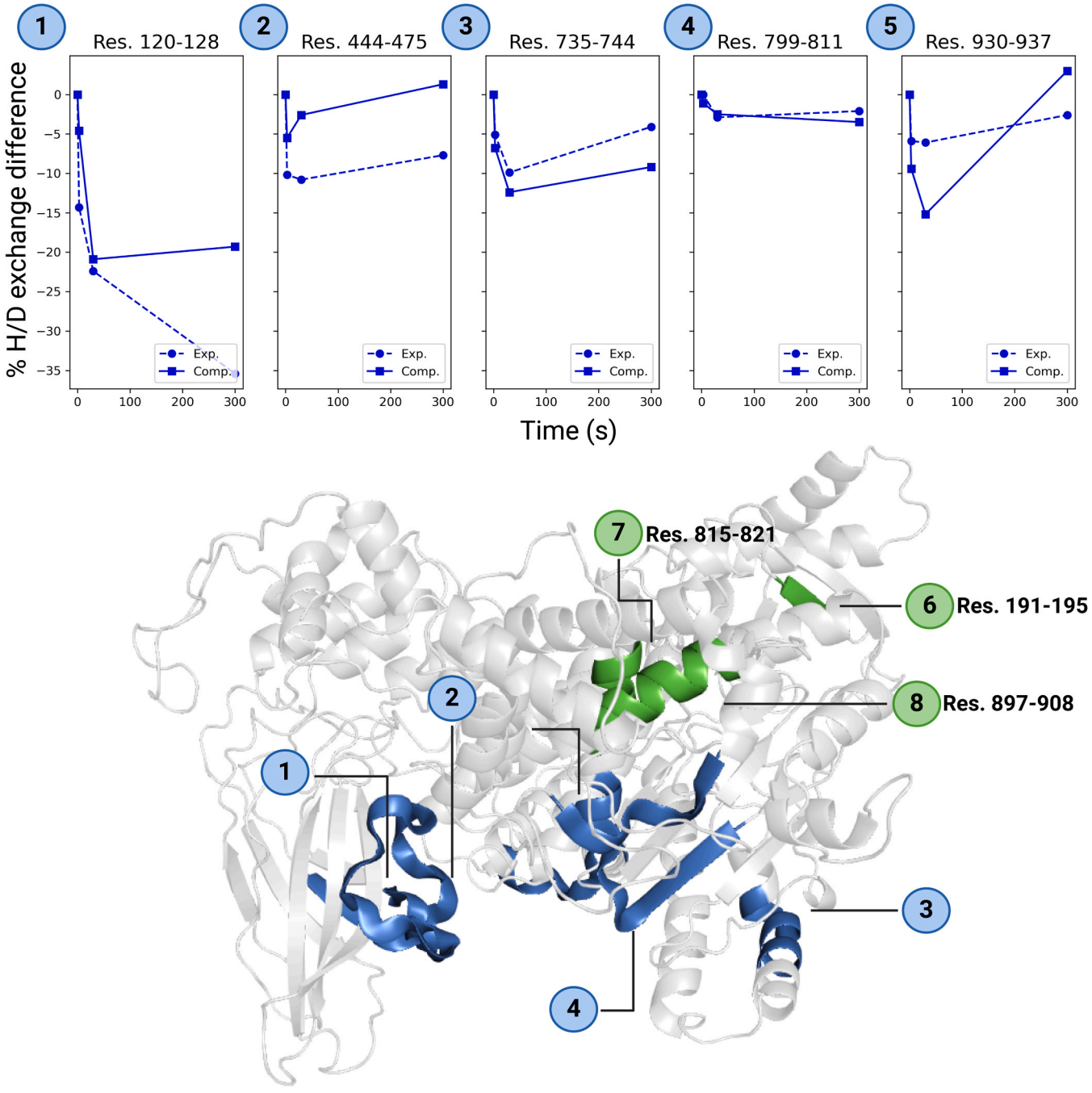

Fig. 2. H/D exchange rate differences calculated from MD simulations compared to experimental data by Jenkins et al. [15]. The H/D exchange rate differences are between the full-length PI3Kα WT and the ΔABD p110α WT. Residues colored green represent regions that showed no difference in exposure upon the removal of p85α and ABD. Residues colored blue represent the residues that show an increase in H/D exchange rate in case of the ΔABD p110α WT compared to the full-length PI3Kα WT.

These residues are in contact with the iSH2 domain. Thus, these residues become more exposed upon the disengagement of niSH2, which explains why disengagement of p85α leads to increased exchange for these residues. For peptide fragment 735–744 of the helical domain, the calculated percentage of H/D exchange rate increase in ΔABD p110α is up to 12.4 %. The experimental increase for residues 735–744 is up to 9.9 %. These residues are in contact with the ABD domain, and in the ΔABD p110α state, the removal of the p85α and of the ABD domain disrupts the contacts between the helical domain and nSH2 domain, leading to increased solvent exposure of these residues. This increase explains why disengagement of p85α leads to increased H/D exchange. For residues 930–937 of the activation loop, a 15.2 % increase in the H/D exchange rate at 30 s is calculated, while the experimentally calculated increase for this peptide fragment is 6.1 %. This increase between the active-state related ΔABD p110α state and the full length PI3Kα may be attributed to the fact that in the ΔABD p110α state, the removal of the regulatory subunit disrupts the contacts between the activation loop and the iSH2 domain and the activation loop becomes more solvent exposed. For residues 191–195, 815–821, and 897–908 no difference in H/D exchange rates was observed upon the removal of p85α and ABD domain.

3.1.2. Convergence of the WT and mutated systems

Three independent unbiased MD simulations were performed for the WT and each of the mutated full-length PI3Kα and ΔABD p110α systems starting from different conformations (see the SI for more details). RMSD was calculated between the starting protein structure Cα atoms and the trajectory frames and the Cα atoms of the active site (residues 772, 780, 776, 810, 836, 848, 849, 851, 917, 922, 930, 932, 936, and 941) and the trajectory frames. RMSD curves indicated that a plateau is reached after 500 ns for the full-length PI3Kα simulations (Figs. S1 and S2) and after 300 ns (Figs. S1 and S2) for the ΔABD p110α simulations (Figs. S3 and S4), apart from one replica in ΔABD p110α H1047R (Fig. S3C). Therefore, for the analysis of the retrieved trajectories, the last 2.5 μs of each trajectory of the full-length PI3Kα and 1 μs of each trajectory of the ΔABD p110α system were used. We performed PCA on Cα atoms of the C-terminus (residues 1047–1068) for each independent simulation for the WT and mutated full-length PI3Kα and ΔABD p110α (Figs. S5 and S6) to map the conformational space of each simulation. In the full-length PI3Kα systems, the overlap of the 2D projections of the trajectories on the first two eigenvectors indicated that the three independent simulations for the WT and mutant proteins span the same conformational phase space (Fig. S5). For the ΔABD p110α systems, the 2D projections of the WT, M1043L, G1049R, and N1068KLKR overlap and these proteins also span the same conformational space (Fig. S6). However, for the ΔABD p110α H1047R, the third replica forms a different cluster in the phase space (Fig. S6C), which could be associated with the higher RMSD values compared to the other ΔABD p110α systems (Fig. S3C).

3.1.3. C-terminal conformations

To further examine the change observed in the RMSD values and PCA analysis of ΔABD p110α H1047R and the C-terminal conformations of the different systems, we performed clustering on the Cα carbon atoms of each unbiased MD trajectory to identify representative structures. The representative structures of the full-length PI3Kα systems (Fig. S9) show that a C-terminal "closed-to-open" transition is not observed in full length PI3Kα. On the other hand, a spontaneous “opening” of the C-terminus occurs in one of the three replicas of H1047R mutant of the ΔABD p110α unbiased MD simulation within the first 50 ns (Fig. S10). This spontaneous "opening" was also observed in a recently published study by Varkaris et al., where unbiased MD simulations of full-length PI3Kα (p110α and p85α subunits) and p110α alone were performed [24]. In this study, the full-length PI3Kα WT simulations, the C-terminal tail remains stable in a closed state throughout the simulation, while in the H1047R PI3Kα the C-terminus begins to disengage from the rest of p110α, but without fully disengaging. In the unbiased MD simulations of p110α H1047R, the C-terminus becomes exposed similarly to the disengaged states observed in 3HHM, and similar to our results [16,24].

To determine which residues act as critical nodes in mediating the "closed-to-open" C-terminal transition, we calculate allosteric networks using the dynamical network analysis method [50]. We use the concept of “Residue Centrality”, which shows how important a node is to the entire protein network, as this metric was shown previously to unambiguously distinguish functional sites that are responsible for mediating allosteric interactions [61]. In this approach, the protein is represented as a graph and the network metric, “node betweenness” (which can be interchangeably referred to as "node centrality") is used. Node betweenness, which quantifies the involvement of specific residues in mediating correlated motions within the protein, with a higher value signifying that a residue is present more frequently in the shortest paths connecting residue pairs. It was previously shown that even relatively small differences in the node betweenness (larger than 0.03) may affect the global coordinating role of functional residues [61]. First, we examined whether there are differences in the networks between the ΔABD p110α H1047R closed (replica 1) and open (replica 3) C-terminal conformations. We found the networks of the open and closed conformations to be equivalent with differences in node betweenness < 0.01 (Figure S11A, B). Then, we compared replica 1 of the full-length H1047R (closed C-terminus) with replica 1 of ΔABD p110α H1047R simulation (closed C-terminus) and focus only on the kinase domain, where the C-terminal transition takes place. This comparison reveals an increrase of 0.06 ± 0.03 in node betweenness for residues around the ATP pocket (residues 797, 799, 838, 839, 846, 851, 879, and 895), the activation loop (residues 948, 953, 955, and 957) and in kα8 helix (residues (residues 961-1002) (Fig. S11A,C). Overall, the ΔABD p110α H1047R simulation shows consistently higher node betweenness compared to the full-length H1047R (0.07 ± 0.04). Accordingly, in the ΔABD p110α WT there is an increase in node betweenness compared to the full-length H1047R of 0.03 ± 0.05 (Fig. S11D). However, the ΔABD p110α H1047R shows a notable increase (> 0.11) in node betweenness compared to the ΔABD p110α WT in residues of the helical domain (611, 615, 651, and 653, which are at the interface with the RBD) and of the kinase domain (923 (ATP active site), 999 and 1002 in kα9 helix). In conclusion, the ΔABD p110α H1047R is increasing the allosteric communication within PI3Kα, which is consistent with the fact that we observed the C-terminal “opening” in the ΔABD p110α H1047R unbiased MD simulation. Therefore, some interesting questions arise: Is the full-length PI3Kα preventing the "opening" of the C-terminal tail? Is the ΔABD p110α facilitating the C-terminal “opening” and what is the free energy needed for the transition of the C-terminus from the closed to the open conformation? As classic MD simulations cannot tackle rare events, we resort to enhanced sampling simulations to answer these questions as described below.

3.2. Multiple-walkers metadynamics simulations of the full-length PI3Kα

Because 2.5 μs of unbiased MD was insufficient to sample large conformational changes in the full-length PI3Kα systems, we used multiple-walkers metadynamics simulations to enhance the sampling and examine the energy needed for the “closed-to-open” C-terminal transition in the presence and absence of oncogenic mutations. We observe that metadynamics simulations of the full-length PI3Kα systems required more than 1.5 μs to approach convergence based on monitoring the diffusion of the CVs with respect to time, as well as from the evolution of the 1D free energy profiles across the reaction coordinates with varying simulation increments (Table S2, Figs. S12 and S13). However, the CV diffusion was limited even after 1.5 μs (Fig. S13). The FESs (Fig. S14) of the WT and mutants M1043L, H1047R, G1049R, and N1068KLKR exhibited multiple free energy minima. We extracted the structures for each basin and report the free energies of the deepest minimum of the closed C-terminal conformation, of the deepest minimum of the open C-terminal conformation, and of the transition state

between them (Table S3). All transition states are within ~4 kcal/mol and thus no major differences in the free energy needed for the C-terminal transition are observed across the systems. This observation together with the fact that the CV diffusion is limited (Fig. S13) may indicate that the multiple-walkers metadynamics simulations of the full-length PI3Kα system require additional time to reach conclusive results. Thus, we did not analyze these FESs and trajectories further, and proceeded with the analysis of the multiple-walkers metadynamics simulations of the ΔABD p110α.

3.3. Multiple-walkers metadynamics simulations of the ΔABD p110α and analysis

Multiple-walkers metadynamics simulations were also performed for the ΔABD p110α systems. Metadynamics simulations on the ΔABD p110α converged only after 280 ns (Table S4, Figs. S15 and S16), i.e. five times faster than the full length PI3Kα, which may indicate that the free energy barrier for the "closed-to-open" C-terminal transition in the ΔABD p110α construct is lower, and that ΔABD p110α is related to the actual active state that associates with membranes. As in the full-length PI3Kα systems, the FESs of the WT and mutant proteins exhibit multiple free energy minima (Fig. 3). We extracted the structures for each basin (Fig. 3) and report the free energies of the deepest minima of the closed C-terminal conformation, of the deepest minima of the open C-terminal conformation, and of the free energy of the transition state (Table 1). The stds associated with the calculated free energies indicate consistent measurements within each system.

Moreover, we report the contacts between helices kα11 and kα8, between helix kα11 and the activation loop, and between helix kα11 and the catalytic loop of the basins of the closed, open and transition states using CPPTRAJ [57], and identify differences across the systems (Tables S5–S7). Additionally, the proper Euler angles α, β, and γ of the kα11 helix (see Methods) were calculated for the extracted minima structures and available crystal structures of the H1047R mutant (Table 2 and Table S8). This angle describes the rotation around the z-axis in the x-y plane (Fig. S17) and highlights the primary rotational difference in the x-y plane, which is critical for understanding the orientation shift in this context. For the definitions of the β and γ Euler angles, see the SI.

3.3.1. The global minimum is observed in the closed state of the C-terminus in the ΔABD p110α WT

The FES of the ΔABD p110α WT is shown in Fig. 3A. The global minimum across all the ΔABD p110α systems is in basin A of the WT, in which the C-terminal tail is in the closed conformation. In basin A, the WIF motif (residues 1057, 1058, and 1059), a triplet of hydrophobic residues crucial for lipid-binding, is not in the direction of the putative membrane bilayer and interacts with the kα11 helix through a dipole-ion interaction between W1057 and E1037 (Fig. 3A, closed state). The deepest minimum, where the C-terminus is in the open conformation, is located in basin G. The free energy of this structure is equal to 8.95 ± 0.64 kcal/mol. However, between basin A and basin G there are states with high free energy. The free energy of the transition state is equal to 19.46 ± 0.40 kcal/mol. In the open C-terminal conformation, the WIF motif points to the direction of the putative membrane bilayer and the dipole-ion interaction with E1037 is disrupted (Fig. 3 A, open state). Remarkably, in the ΔABD p110α WT closed and open states, there are no contacts between kα11 helix and the activation loop contrary to the full-length PI3Kα (see Introduction); these contacts are present in all ΔABD p110α mutant systems (Table S5 and S6). Because the kα11 helix axis of the H1047R mutant is tilted by ~4° with respect to the WT structure (Table S8), we calculated the Euler angles of the kα11 helix orientation. In the closed state, the WT helix kα11 shows minor reorientation with Euler angles α = -4.5°, β = -1°, and γ = -0.9°. In the open state, the angles shift to α = -0.2°, β = 6°, and γ = -5.4°. Here, the α angle remains close to zero and the positive β value shows a notable upward tilt along the y-axis. The negative γ indicates a slight counterclockwise rotation around the x-axis. In the transition state, kα11 helix exhibits significant reorientation with angles α = 10.1°, β = 10.7°, and γ = -4.6° (Table 2 and Table S8). The positive α signifies a clockwise rotation around the z-axis, the positive β indicates an upward tilt along the y-axis, and the negative γ shows a counterclockwise rotation around the x-axis. These differences between the closed and the transition states of the ΔABD p110α WT may be related to the C-terminal active state.

3.3.2. The C-terminal transition state has the lowest free energy in the H1047R and G1049R mutants compared to the WT, M1043L and N1068KLKR

The FES of the ΔABD p110α H1047R is shown in Fig. 3C and is shallower compared to the FES of the WT. The deepest minimum of the H1047R system in basin A has free energy equal to 5.67 ± 0.20 kcal/mol and the C-terminus is in the closed state. The deepest minimum of the open C-terminal conformation is located in basin C and has free energy equal to 7.55 ± 0.02 kcal/mol. The free energy of the transition state is equal to 11.67 ± 0.25 kcal/mol, which is the lowest observed among all transition states. In the open C-terminal conformation, the WIF motif is in the direction of the putative membrane allowing protein-membrane interactions.

Comparing the contacts between WT and H1047R mutant in the closed state of the C-terminus, we observe that contacts between the amino acids of kα11 and the activation loop are present in the H1047R system and absent in the WT (Table S5). These contacts form between the amino acids A1046 and 954–956 and R1047 and L956 that are present in more than 70 % of the frames of the closed state. Additionally, in the open state, contacts between A1046 and V952 and R1047 and L956 are present in more than 80 % of the frames and are absent in the WT. Regarding the catalytic loop, in the open state of H1047R mutant, contacts between Q1042 and F909 are present in 100 % of the frames, while they are absent in the WT (Table S6). Contacts between helix kα11 and the activation and catalytic loops may affect the kinase activity of the system.

By comparing contacts between helices kα8 and kα11 of H1047R with the ones in the WT, we observe that in the transition state of the H1047R system, several contacts between the two helices are disrupted. More specifically, the disrupted contacts are between L1036 and L989, F1039 and Q981, R1047 and F977, R1047 and F980, G1050 and F977, and between W1051 and F977, Q981 and Y985 (Table S7). Moreover, by comparing the C-terminal conformations with the crystal structure of the WT, 4OVU [19], in the closed state, the H1047R mutant has Euler angles α = 10.9°, β = 0.2°, and γ = 5°. The positive α, which is similar to the calculated α angle of the WT transition state, indicates a significant clockwise rotation around the z-axis. This rotation may destabilize the kα11-kα8 interface. The angle β is close to zero and suggests no tilt along the y-axis. The positive γ indicates a significant clockwise rotation around the x-axis. For the open state, the angles shift to α = 8.4°, β = 0°, and γ = 3.5°. The substantial positive α indicates a significant clockwise rotation around the z-axis, and the zero β value suggests no tilt along the y-axis, and the positive γ indicates a moderate clockwise rotation around the x-axis. In the transition state, H1047R shows angles α = 8.7°, β = 2.9°, and γ = 1.1°, reflecting a significant clockwise rotation around the z-axis (α), a minor upward tilt along the y-axis (β), and a slight clockwise rotation around the x-axis (γ) (Table 2 and Table S8). In the closed, open and transition states of H1047R mutant, the side chain of R1047 is in the same position and direction with the one observed in the recently published PDB structure 8GUB (Fig. S18) [23]. The H1047R mutation alters the interface between the kα8 and kα11 both in the closed and the transition states. This conformational change at the C-terminal tail may affect the free energy barrier and enable the “closed-to-open” C-terminal transition. A similar conformational change was observed in the recently published H1047R mutant structure with PDB ID 8W9A in complex with the allosteric inhibitor QR-7909, which binds in close proximity to the H1047 residue and disrupts the kα11-kα8 interface triggering the

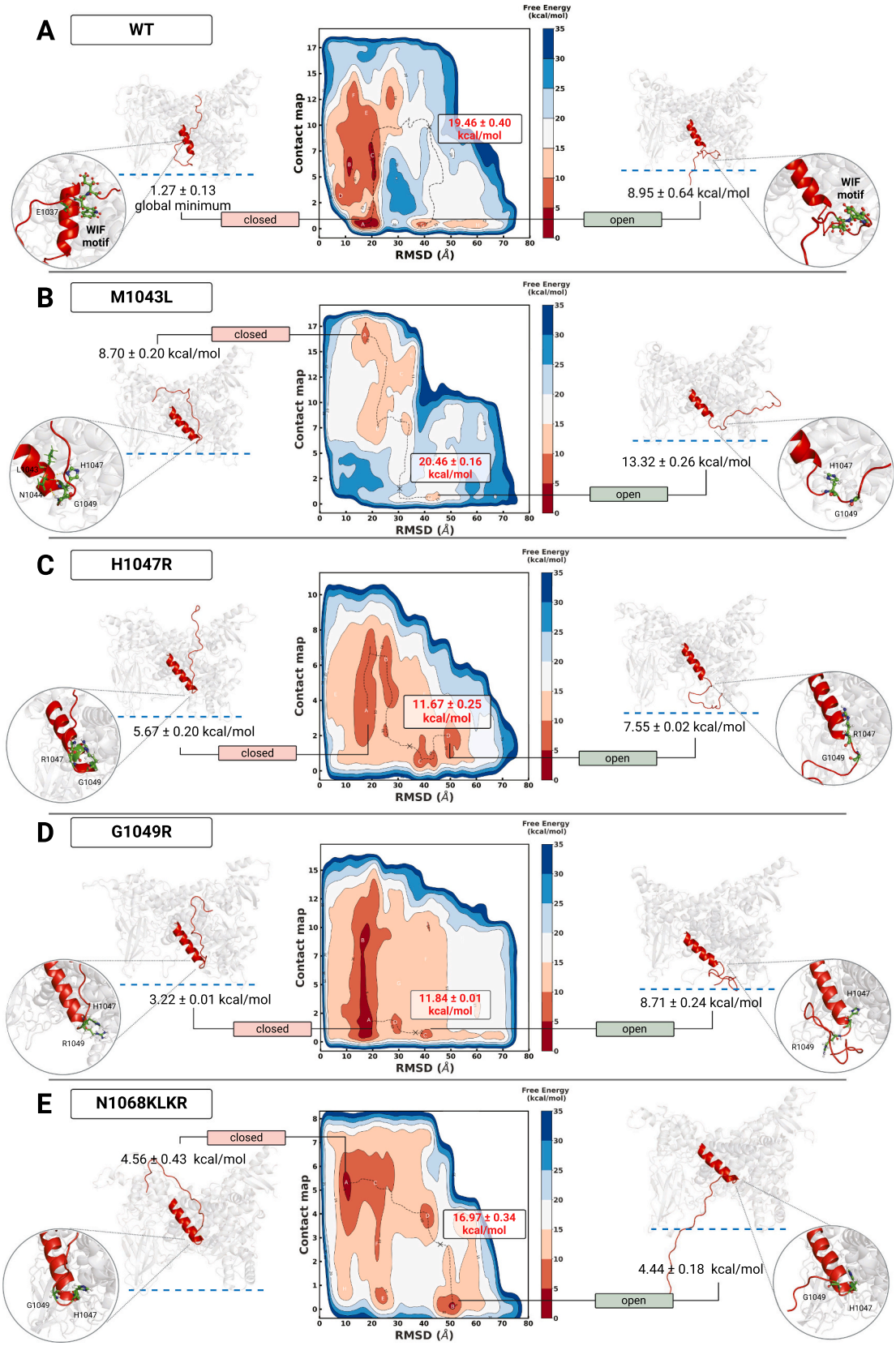

Fig. 3. Free energy surfaces of (A) the WT ΔABD p110α and (B-E) the mutant ΔABD p110α systems along the two CVs. Inside the FESs, the letters A-H indicate the structures of the free energy minima. The position of the transition state in the free energy surface is shown with the letter x and the putative membrane bilayer is shown with a blue dashed line.

Table 1
Free energy surface calculations for the ΔABD p110α systems. The free energies of the deepest minimum of the closed C-terminal conformation, of the deepest minimum of the open C-terminal conformation and the transition state are reported. The free energy of the global minimum that was initially set to zero (WT, closed state) is offset by the calculated mean values of within the grid used for bootstrapping for each system. Correspondingly, all the free energy minima were offset by their calculated mean values (see the SI).

	Free energy (kcal/mol)		
System	**Closed state**	**Open state**	**Transition state**
WT	1.27 ± 0.13	8.95 ± 0.64	19.46 ± 0.40
M1043L	8.70 ± 0.20	13.32 ± 0.26	20.46 ± 0.16
H1047R	5.67 ± 0.20	7.55 ± 0.02	**11.67 ± 0.25**
G1049R	3.22 ± 0.01	8.71 ± 0.24	**11.84 ± 0.01**
N1068KLKR	4.56 ± 0.43	4.44 ± 0.18	16.97 ± 0.34

Table 2
Euler angle α of helix kα11 in the extracted ΔABD p110α structures of the closed, open and transition state of the WT and mutants with respect to the crystal structure of PI3Kα WT, PDB ID 4OVU.

	Euler angle α (rotation)
40VU (ref)	0
3HHM	4
Closed state	
WT	-4.5
M1043L	10.5
H1047R	10.9
G1049R	6.5
N1068KLKR	3.8
Open state	
WT	-0.2
M1043L	8.6
H1047R	8.4
G1049R	16.1
N1068KLKR	14.8
Transition state	
WT	10.1
M1043L	8.5
H1047R	8.7
G1049R	3.2
N1068KLKR	3.6

reorientation of R1047 (Fig. 1G) [26]. The recently-published pyridopyrimidinones, compound 4 (PDB ID 8V8J, 8V8H), compound 7 (PDB ID 8V8V), and compound 12 (PDB ID 8V8U), also bind in the same pockets as in ref. [26]. Evidently, H1047R-selective inhibitors that bind at the interface of kα11-kα8, induce an open C-terminal state and reorient the side chain of R1047 similar to QR-7909.

The FES of the ΔABD p110α G1049R is shown in Fig. 3D and is similar with the FES of the ΔABD p110α H1047R. The deepest minimum of the G1049R system in basin A is in the closed C-terminal state and has free energy = 3.22 ± 0.01 kcal/mol, which is deeper compared to H1047R. In this structure, R1049 points outwards in the direction of the membrane. The open C-terminal conformation is located in basin C and has free energy = 8.71 ± 0.24 kcal/mol. As in the H1047R system, the free energy of the transition state is much lower compared to the WT and is equal to 11.84 ± 0.01 kcal/mol. In the open C-terminal conformation, the WIF motif is in the direction of the putative membrane allowing protein-membrane interactions. However, because the closed state of G1049R has lower free energy compared to the H1047R mutant, it is expected that the "closed-to-open" transition will occur more easily in H1047R mutant with respect to the G1049R. This finding aligns with experimental data, which report a higher membrane-binding ability of H1047R mutant compared to G1049R [15]. For the G1049R mutant in the closed state, the helix shows reorientation with angles α = 6.5°, β = 3.4°, and γ = -0.1°. The positive α indicates a clockwise rotation around the z-axis, while the positive β suggests a slight upward tilt along the y-axis. The negative γ indicates minimal rotation around the x-axis. As in the closed state of H1047R mutant, also in G1049R, a rotation of kα11 helix is observed, which may affect the kα11-kα8 interface (Table 2 and Table S8). In the open state, G1049R exhibits significant reorientation with angles α = 16.1°, β = 0.5°, and γ = 6.6°. The positive α signifies a pronounced clockwise rotation around the z-axis, the positive β indicates a minor upward tilt along the y-axis, and the positive γ shows a notable clockwise rotation around the x-axis. During the transition state, G1049R has angles α = 3.2°, β = 8.6°, and γ = -6.1°, showing a slight clockwise rotation around the z-axis (α), a considerable tilt along the y-axis (β), and a counterclockwise rotation around the x-axis (γ).

C-terminus G1049R mutant contacts reveal similarities with the H1047R mutant: contacts between A1046 of the kα11 helix and V952 of the activation loop are present in the mutated system and absent in the WT in the closed state (Table S5). Additionally, in the open state, contacts between A1046 and V952 and R1047 and L956 are present only mainly in H1047R and G1049R systems and absent in the WT (Table S6). Similarly to the H1047R system in the transition state, several contacts between the helices kα11 kα8 are disrupted. These contacts are the contacts between G1050 and F977 and between W1051 and amino acids F977, Q981 and Y985 (Table S7). Therefore, G1049R makes similar contacts with H1049R mutant, which may have impact in their mechanism of activation and imply similarities in the mechanism of C-terminal activation.

3.3.3. The M1043L mutant exhibits a different direction of the C-terminus compared to the WT and mutants being parallel to the putative membrane plane

The FES of the ΔABD p110α M1043L is shown in Fig. 3B. The deepest minimum of the M1043L system in basin A has free energy 8.70 ± 0.20 kcal/mol and corresponds to the closed C-terminal conformation. The open C-terminal conformation is located in basin G and has free energy 13.32 ± 0.26 kcal/mol. Similarly with the WT system, between basin A and basin G there are states with high free energy. The free energy of the transition state is equal to 20.46 ± 0.16 kcal/mol, which is comparable with the WT transition state free energy. In the open C-terminal conformation, the C-terminus changes direction compared to the open C-terminal states of the other systems and is parallel to the putative lipid bilayer.

Examining the contacts of kα11, M1043L mutant forms more extensive contacts between kα11 and kα8 compared to the other systems, which may trigger the different direction of the C-terminus observed in the M1043L mutant. Similarly to the other mutants in the closed state, M1043L mutation triggers the reorientation of kα11 helix, with α = 10.5°, β = -1.1°, and γ = 5.4°. The positive α value indicates a substantial clockwise rotation around the z-axis, while the negative β suggests a slight downward tilt along the y-axis. The positive γ indicates a clockwise rotation around the x-axis. In the open state, the angles are α = 8.6°, β = 1.6°, and γ = 2.2°, reflecting a significant clockwise rotation around the z-axis (α) and a minor upward tilt along the y-axis (β), with a slight clockwise rotation around the x-axis (γ). During the transition state, M1043L displays angles α = 8.5°, β = 5.4°, and γ = -0.8°, showing a noticeable clockwise rotation around the z-axis (α), an upward tilt along the y-axis (β), and a negligible counterclockwise rotation around the x-axis (γ) (Table 2 and Table S8). In the closed state, amino acid H1047 is rotated to the same position as the R1047 in PDB ID 8GUB (Fig. S18) [23]. Additionally, contacts between F1039 and Q981 and between the amino acids H1047, W1051, T1052 and T1053 of kα11 and F977 and Q981 of kα8 are present in more than 70 % of the frames and absent in the closed state of the WT. As in the other mutants, several amino acids of kα11, including the amino acids 1045–1047, make contacts with amino acids 954–956 of the activation loop in the closed state (Table S5). These contacts are absent from the WT. Moreover, in the open state, contacts between M1055 and T974, L1043 and V952 and D1045 and V952 are present in more than 40 % of the frames and absent in the WT and in the other mutants (Table S6). These contacts may

stabilize the different C-terminal conformation. These differences in contacts between the M1043L and the mutants H1047R and G1049R suggest a different mechanism of C-terminal activation.

3.3.4. *The N1068KLKR mutant may not easily reorient the C-terminus due to high energy needed for the C-terminal transition*

Finally, the FES of the ΔABD p110α N1068KLKR is shown in Fig. 3E. The deepest minimum of the N1068KLKR system in basin A has free energy 4.56 ± 0.43 kcal/mol and corresponds to the closed C-terminus. The open C-terminal conformation, which is located in basin B and has free energy 4.44 ± 0.18 kcal/mol. In the open C-terminal conformation, the kα12 along with the WIF motif and 1068KLKR are in the direction of the membrane enabling the protein-membrane interactions. Between basin A and basin B, the free energy of the transition state is equal to 16.97 ± 0.34 kcal/mol, comparable with the ones of the WT and M1043L mutant. This indicates that N1068KLKR mutant may not easily reorient the C-terminus and it needs to be further examined whether the membrane acts as a catalyst in the case of N1068KLKR.

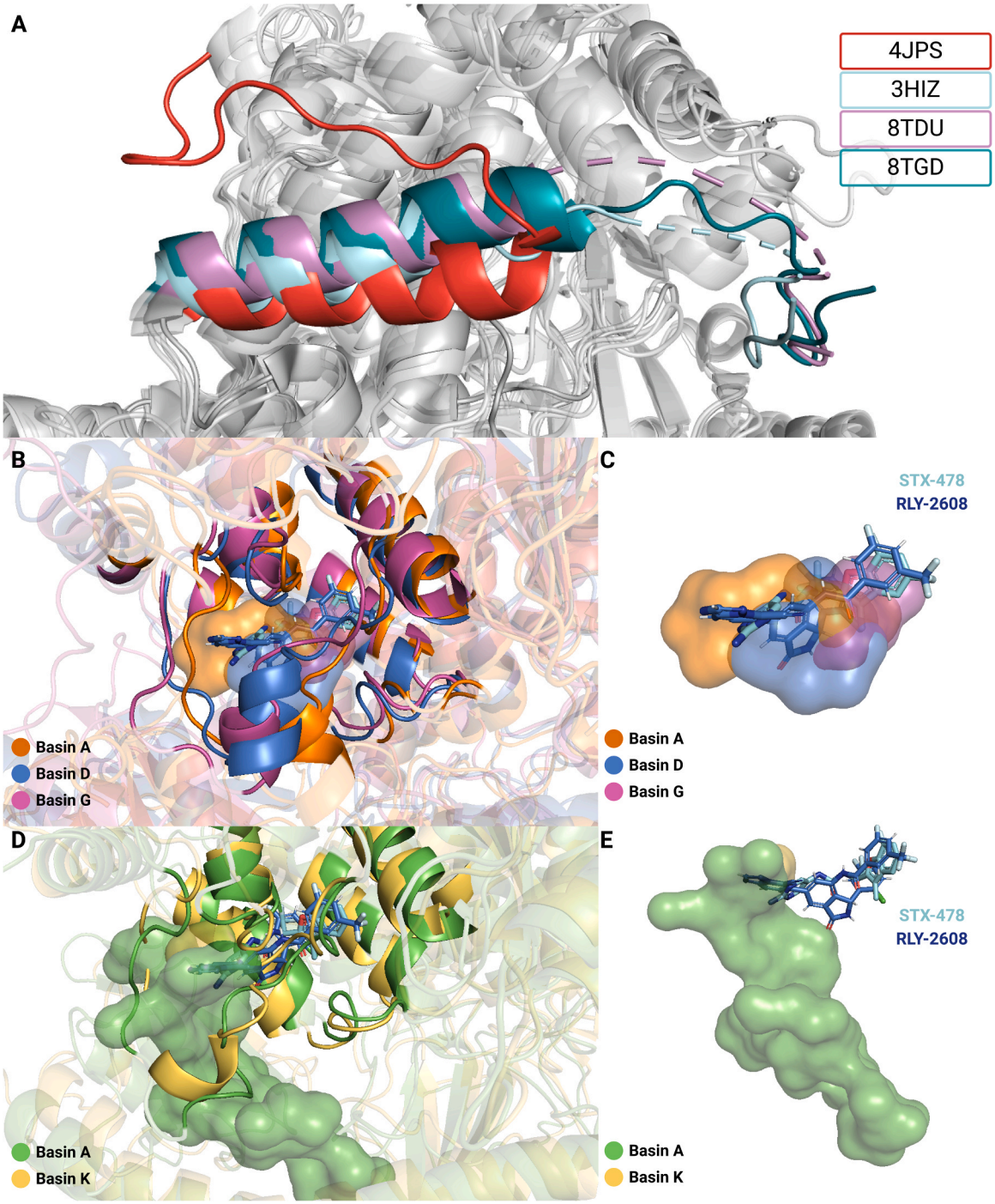

Fig. 4. Cavities in the STX-478 binding site. (A) The C-terminal tails of the STX-478 bound PDB structures, 8TDU (WT) and 8TGD (H1047R) with respect to the crystal structures with PDB IDs 4JPS (WT) and 3HIZ (H1047R). (B) Free energy minima structures extracted from basins A, D and G of the full-length WT along with (C) their cavities in the binding site of STX-478 are superimposed. (D) Minima structures extracted from basins A and K of the full-length H1047R mutant along with (E) their cavities in the binding site of STX-478. All minima structures along with the PDB structure 8TSD [24] were aligned to the crystal structure with PDB ID 8TGD [22] and the STX-478 and RLY-2608 molecules were superimposed.

Comparing the contacts of the N1068KLKR mutant with the ones of the WT and M1043L, H1047R, G1049R mutants, we report several similarities. As in the other mutants, amino acids 1045–1047 form contacts with the activation loop in the closed state and these contacts are absent in the WT (Table S5). Furthermore, contacts between Q1042 and F909 are present in all the mutated systems in the open state and absent in the WT, which may have an effect on the kinase activity of the mutants (Table S6). In contrast to the other mutants, in the closed state, the N1068KLKR mutant exhibits angles $\alpha = 3.8°$, $\beta = -2.2°$, and $\gamma = 3.5°$, which suggest a minor reorientation of kα11 helix (Table 2 and Table S8). In the open state, the angles shift to $\alpha = 14.8°$, $\beta = 2.9°$, and $\gamma = 4.1°$. The substantial positive α shows a significant clockwise rotation in the z axis, the positive β indicates a slight upward tilt along the y-axis, and the positive γ shows a clockwise rotation around the x-axis. During the transition state, N1068KLKR displays angles $\alpha = 3.6°$, $\beta = 4.8°$, and $\gamma = -2.6°$, reflecting a slight clockwise rotation in the x-y plane (α), a moderate tilt along the y-axis (β), and a minor counterclockwise rotation around the x-axis (γ). These kα11 helix reorientations are similar among all mutants, which may indicate that they play a role in kinase activation.

3.4. The STX-478 binding pocket starts to form only on the full-length minima of the WT and H1047R and not in the ΔABD p110α

The pocket of the allosteric PI3Kα inhibitor that targets prevalent PI3Kα helical- (E542K, E545K) and kinase-domain (H1047R) mutant tumors, STX-478 [22], was investigated in the extracted structures from the minima generated by the metadynamics simulations both for the WT and the H1047R mutant using fpocket [62]. We find that the pocket starts to form only in the full-length minima of the two systems and is absent in the ΔABD p110α systems, which is in accordance with the STX-478 bound experimental structures with PDB IDs 8TDU and 8TGD, which are both full-length structures for the WT and H1047R respectively. The C-terminal tails of both STX-bound PDB structures, 8TDU (WT) and 8TGD (H1047R), are in the same direction as in PDB ID 3HIZ (H1047R), in the open state (Fig. 4A). In the extracted WT structures from our metadynamics simulations, cavities on the binding site of STX-478 are observed in basin A, in which the C-terminus is in the closed state and in basins D and G in which the C-terminus is in the open state (Fig. 4B,C). In H1047R mutant, cavities on the binding site of STX-478 were observed in basins A and K, in which the C-terminus is in the open state (Fig. 4D,E). In the H1047R mutant, although the cavities cover a small part of STX-478 and RLY-2608, a large pocket is forming, which connects the active site and the allosteric pocket (Fig. 4E).

4. Discussion - Conclusions

Results presented herein describe the free energy landscape for the C-terminal “closed-to-open” transition and conformational changes that occur across different oncogenic mutations and provide valuable insights into the mechanism of the C-terminal PIK3CA oncogenic mutants M1043L, H1047R, G1049R and N1068KLKR. Unbiased MD simulations were performed for the full-length PI3Kα and ΔABD p110α systems. To validate our models, we compared the results from our MD simulations with available HDX-MS experimental data, which provide the H/D exchange rates for the full-length PI3Kα WT and the ΔABD p110α WT, as well as the differences between them. Our predicted H/D exchange rates align to previous experimental data [15]. In unbiased MD simulations of the full-length PI3Kα, the transition of the C-terminus from the closed to the open state was not observed in any of the systems, however, the C-terminal transition was observed in one of the replicas of the ΔABD p110α H1047R system. To determine which residues are important in mediating the "closed-to-open" C-terminal transition, we used the concept of “node betweenness" and compared the ΔABD p110α H1047R closed (replica 1) and open (replica 3) C-terminal simulations, which showed that the allosteric networks in the two C-terminal conformations are equivalent. Then, we compared replica 1 of the full-length H1047R (closed C-terminus) with replica 1 of ΔABD p110α H1047R simulation (closed C-terminus), focusing on the kinase domain, and observed an increase of 0.06 ± 0.03 in node betweenness for residues around the ATP pocket, the activation loop and kα8 helix, indicating an increase in the allosteric communication of the ΔABD p110α H1047R system in these regions. Overall, the ΔABD p110α H1047R is increasing the allosteric communication within PI3Kα, which is consistent with the fact that we observed the C-terminal “opening” in the ΔABD p110α H1047R unbiased MD simulation. Because we observed the C-terminal “opening” only in the ΔABD H1047R unbiased simulation, which is a rare event, we decided to perform multiple-walkers metadynamics simulations for both the full-length PI3Kα and ΔABD p110α WT and mutant systems to calculate the free energy required for the C-terminal "closed-to-open" transition. The CVs that were chosen to study this transition were, CV1, the distance in contact map space from a reference state in which the C-terminal tail is the closed state with respect to the open state and CV2, the RMSD of the protein kα12 Cα atoms starting from the closed state.

Metadynamics simulations on the full-length PI3Kα required more than 1.5 μs to approach convergence. In the full-length PI3Kα systems, all transition states are within ~4 kcal/mol, indicating no significant differences in the free energy required for the C-terminal transition across the systems. This finding, coupled with the limited CV diffusion, suggests that the multiple-walkers metadynamics simulations of the full-length PI3Kα system may need additional simulation time to achieve conclusive results. Metadynamics simulations of the ΔABD p110α converged only after 280 ns, which may suggest that the PI3Kα C-terminal "opening" may require disengagement of the ABD and p85α regulatory subunit. In the metadynamics simulations on the ΔABD p110α, the global minimum is observed in the closed state of the C-terminus in the WT, where the kα11 helix closely resembles the conformation in the crystal structure with PDB ID 4OVU. The WT transition state has a transition state free energy of ~20 kcal/mol and exhibits a reorientation of kα11, which may be related to the C-terminal active state. In contrast, the H1047R and G1049R mutants lower transition state free energy by ~8 kcal/mol and have the lowest transition state free energies among all systems ($\Delta G = \sim 12$ kcal/mol). Moreover, H1047R and G1049R mutants show similar open structure conformations, suggesting a similar mechanism of C-terminal activation compared to the WT and M1043L and N1068KLKR mutants. Both H1047R and G1049R oncogenic mutations change the interactions between helices kα8 and kα11 by triggering the rotation of the kα11 helix in the closed state. In the transition states of both H1047R and G1049R mutants, similar contacts between helices kα11 and kα8 are disrupted, which may imply similarities in their mechanism of C-terminal activation. However, as the open state of H1047R has lower free energy compared to the open state of G1049R, it is expected that the Gibbs free energy of activation for the “closed-to-open” transition will occur more easily in H1047R mutant compared to the G1049R, potentially explaining why H1047R is the most prevalent oncogenic mutation. The lowest transition state free energy of the C-terminal "opening" of the H1047R mutant aligns with experimental data [15], according to which a higher membrane-binding ability of the H1047R compared to G1049R mutant is reported. In the closed, open state, and transition states of the H1047R mutant, the side chain of R1047 is in the same position and direction as observed in the recently published PDB structure 8GUB.

Similarly to H1047R and G1049R mutations, the M1043L mutation, triggers the rotation of kα11 helix in the close state. This reorientation of the kα11 helix in the closed state leads to more extensive contacts between kα11 and kα8 that are absent in the WT and the other mutants and to the reorientation of amino acid H1047 to the same position as the R1047 in PDB ID 8GUB. Interestingly, in the open state of the M1043L mutant, a different direction of the C-terminus is observed compared to the WT and the other mutants. This direction is parallel to the putative membrane plane and is stabilized in the open state by contacts between helices kα8 and kα11 that are absent in the other systems. The different

directionality of the C-terminus may not enable the C-terminal tail to bind to the membrane and could be the reason why this mutant has lower membrane-binding ability compared to the other mutants according to the existing HDX-MS data [15]. Moreover, in the M1043L mutant, the free energy of the transition state is the highest among all systems and comparable to the WT transition state free energy (~20 kcal/mol). For N1068KLKR mutant, the free energy needed for the C-terminal "opening" is comparable to the one of M1043L and WT and we calculate that this mutation, in contrast to the other systems, does not trigger any major reorientation of the kα11 helix in the closed state. This finding suggests that the activation of this activation of this frameshift mutant is not driven by a reorientation of the C-terminus, but instead it may be driven by enhanced affinity caused by the addition of positive and hydrophobic residues (KLKR) and not with the promotion of the "closed-to-open" C-terminal transition. In all mutated systems (M1043L, H1047R, G1049R, and N1068KLKR), amino acids 1045–1047 of kα11 form contacts with the activation loop in the closed state; these contacts are absent in the WT simulation. Also, contacts between Q1042 of kα11 and F909 of the catalytic loop are present in the open state of all the mutated systems, but absent in the WT. The presence of these contacts between kα11 and the catalytic and activation loops may affect the kinase activity of the mutants. The differences in the free energy needed for C-terminal transition and the observed conformational changes across the systems enable us to provide a molecular explanation why H1047R and G1049R reorient the C-terminus with a different mechanism compared to the WT and M1043L and N1068KLKR mutants and why the H1047R mutant, with the lowest expected Gibbs free energy of activation for the "closed-to-open" transition, is the most prevalent mutant in human cancer, providing valuable insights into the molecular mechanisms underlying the C-terminal activation in PI3Kα.

We also investigated in our free energy minima structures the existence of a recently described PI3Kα allosteric pocket, where the mutant-selective PI3Kα inhibitors, RLY-2608 and STX-478, bind [22,24]. In the free energy minima structures of the WT (basins A, D and G) and H1047R mutant (basins A and K) we identified the formation of the cryptic STX-478 pocket in the full-length PI3Kα systems, but not in the ΔABD p110α, consistent with experimental PDB structures that show binding of RLY-2608 and STX-478 in the full length PI3Kα. In the H1047R minima structures, we show several cavities opening (Fig. 4) and forming a channel connecting the active site with the allosteric site; these cavities are not as extensive in the WT, which may mean that the mutant structure opening the cryptic pocket is more populated accessible to bind STX-478. Moreover, while the RMSD between the WT and mutant PDB structures bound to STX-478 are almost identical (PDB IDs: 8TDU and 8TDG, RMSD of the Cα atoms = 0.54 Å), biochemical assays and preclinical data suggest that STX-478 is selectively inhibiting the mutant enzyme. We propose that while the crystal structures have the same coordinates in the mutant and the WT after binding the compound, their pre-bound conformational states are different. The mutant appears to have a higher probability of being found in an accessible state to the inhibitor, and STX-478 and RLY-2608 may bind selectively to the mutant by conformational selection of that state. According to Buckbinder et al. [22], STX-478 binds with 15-fold higher binding affinity and has a ~6-fold faster association rate for the H1047R mutant compared to the WT. Varkaris et al. further support this finding by showing that an analogue of RLY-2608 binds with ~10-fold faster association rate to the mutant than to the WT [24]. These findings, together with our observation for the opening of the STX-478 pocket, suggest that the reason behind the selectivity of these compounds for the H1047R mutant may lie in a conformational rearrangement that occurs more favorably in the mutant enzyme compared to the WT [22,24]. In our future work we aim to examine the effect of these allosteric modulators on the dynamics of the protein, their effect on the PI3Kα C-terminal conformations and binding to the membrane.

Understanding the molecular mechanism by which oncogenic mutations in the kinase domain of PIK3CA overactivate PI3Kα has the potential to reveal novel molecular insights that may be exploited for the development of novel cancer therapeutic strategies. These results will contribute to the development of oncogene-specific PI3Kα inhibitors targeting kinase domain mutants.

Our methodology holds promise for broader applications in the study of other protein complexes. By incorporating Molecular Dynamics simulations and combining them with experimental data, we can gain atomistic insights into the structural changes induced by mutations or post translational modifications (PTMs), shedding light on their functional consequences. As with other proteins, PI3Kα contains PTMs that regulate its catalytic activity such as the direct phosphorylation of S361 in nSH2 and S652 in cSH2 by Protein Kinase C, which prevent PI3Kα from binding to the activated receptor. Additionally, different lysine residues located at the iSH2 domain are suggested to be putative SUMOylation sites, which modulate the p85 function by reducing the levels of tyrosine-phosphorylated-p85α [63]. Understanding the intricacies of such PTMs is crucial for deciphering different states of its catalytic cycle and the regulatory mechanisms governing its function. Our methodology could be extended to explore the dynamic interplay between PTMs and protein function of other systems [64,65]. Furthermore, this methodology can contribute to gaining insights into the mechanism of action of large protein complexes that could be exploited as drug targets [66,67]. By applying similar strategies to diverse biological systems, including those involving intricate regulatory networks or complex macromolecules, our methodology could be used to contribute significantly to advancing our understanding of fundamental biological processes and disease mechanisms.

Funding

DK is funded by the 4th Call for H.F.R.I. Scholarships to PhD Candidates. ZC acknowledges funding from Grant No. 1780 of the Hellenic Foundation for Research and Innovation (H.F.R.I). JEB is supported by the Cancer Research Society (operating grant 1052949), and a scholar award from Michael Smith Health Research (17686).

CRediT authorship contribution statement

Danai Maria Kotzampasi: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Michail Papadourakis:** Investigation, Methodology, Software, Validation. **John Burke:** Methodology, Validation, Writing – review & editing. **Zoe Cournia:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

JEB reports personal fees from Scorpion Therapeutics, and Reactive therapeutics, and research contracts from Calico Life Sciences.

Data availability

The input files for the unbiased and biased simulations and the scripts used for data analysis are available in the Zenodo repository: https://doi.org/10.5281/zenodo.13273642

Acknowledgments

Computational resources from PRACE project Pra23_0075 on Marconi100 hosted by CINECA in Italy and computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC facility - ARIS under project IDs pr008033_gpu/Mem-Surf and pr008033_thin/Mem-Surf are acknowledged. Figures were created with in-house python scripts, BioRender.com and with PyMol.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.07.010.

References

[1] Vadas O, Burke JE, Zhang X, Berndt A, Williams RL. Structural basis for activation and inhibition of class I phosphoinositide 3-kinases. Sci Signal 2011;4. https://doi.org/10.1126/scisignal.2002165.

[2] Kotzampasi DM, Premeti K, Papafotika A, Syropoulou V, Christoforidis S, Cournia Z, et al. The orchestrated signaling by PI3Kα and PTEN at the membrane interface. Comput Struct Biotechnol J 2022;20:5607–21. https://doi.org/10.1016/j.csbj.2022.10.007.

[3] Toker A, Cantley LC. Signalling through the lipid products of phosphoinositide-3-OH kinase. Nature 1997;387:673–6. https://doi.org/10.1038/42648.

[4] Rathinaswamy MK, Burke JE. Class I phosphoinositide 3-kinase (PI3K) regulatory subunits and their roles in signaling and disease. Adv Biol Regul 2020;75:100657. https://doi.org/10.1016/j.jbior.2019.100657.

[5] Liu X, Yang S, Hart JR, Xu Y, Zou X, Zhang H, et al. Cryo-EM structures of PI3Kα reveal conformational changes during inhibition and activation. Proc Natl Acad Sci USA 2021;118:e2109327118. https://doi.org/10.1073/pnas.2109327118.

[6] Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res 2019;47:D941–7. https://doi.org/10.1093/nar/gky1015.

[7] Galdadas I, Gervasio FL, Cournia Z. Unravelling the effect of the E545K mutation on PI3Kα kinase. Chem Sci 2020;11:3511–5. https://doi.org/10.1039/c9sc05903b.

[8] Leontiadou H, Galdadas I, Athanasiou C, Cournia Z. Insights into the mechanism of the PIK3CA E545K activating mutation using MD simulations. Sci Rep 2018;8:15544. https://doi.org/10.1038/s41598-018-27044-6.

[9] Gkeka P, Evangelidis T, Pavlaki M, Lazani V, Christoforidis S, Agianian B, et al. Investigating the structure and dynamics of the PIK3CA wild-type and H1047R oncogenic mutant. PLOS Comput Biol 2014;10:e1003895. https://doi.org/10.1371/journal.pcbi.1003895.

[10] Gkeka P, Papafotika A, Christoforidis S, Cournia Z. Exploring a non-ATP pocket for potential allosteric modulation of PI3Kα. J Phys Chem B 2015;119:1002–16. https://doi.org/10.1021/jp506423e.

[11] Burke JE, Perisic O, Masson GR, Vadas O, Williams RL. Oncogenic mutations mimic and enhance dynamic events in the natural activation of phosphoinositide 3-kinase p110α (PIK3CA). Proc Natl Acad Sci USA 2012;109:15259–64. https://doi.org/10.1073/pnas.1205508109.

[12] Hon W-C, Berndt A, Williams RL. Regulation of lipid binding underlies the activation mechanism of class IA PI3-kinases. Oncogene 2012;31:3655–66. https://doi.org/10.1038/onc.2011.532.

[13] Zhang M, Jang H, Nussinov R. The mechanism of PI3Kα activation at the atomic level. Chem Sci 2019;10. https://doi.org/10.1039/C8SC04498H.

[14] Spangle JM, Von T, Pavlick DC, Khotimsky A, Zhao JJ, Roberts TM. PIK3CA C-terminal frameshift mutations are novel oncogenic events that sensitize tumors to PI3K-α inhibition. Proc Natl Acad Sci USA 2020;117:24427–33. https://doi.org/10.1073/pnas.2000060117.

[15] Jenkins ML, Ranga-Prasad H, Parson MAH, Harris NJ, Rathinaswamy MK, Burke JE. Oncogenic mutations of PIK3CA lead to increased membrane recruitment driven by reorientation of the ABD, p85 and C-terminus. Nat Commun 2023;14:181. https://doi.org/10.1038/s41467-023-35789-6.

[16] Mandelker D, Gabelli SB, Schmidt-Kittler O, Zhu J, Cheong I, Huang C-H, et al. A frequent kinase domain mutation that changes the interaction between PI3Kalpha and the membrane. Proc Natl Acad Sci USA 2009;106:16996–7001. https://doi.org/10.1073/pnas.0908444106.

[17] Furet P, Guagnano V, Fairhurst RA, Imbach-Weese P, Bruce I, Knapp M, et al. Discovery of NVP-BYL719 a potent and selective phosphatidylinositol-3 kinase alpha inhibitor selected for clinical evaluation. Bioorg Med Chem Lett 2013;23:3741–8. https://doi.org/10.1016/j.bmcl.2013.05.007.

[18] Barsanti PA, Aversa RJ, Jin X, Pan Y, Lu Y, Elling R, et al. Structure-based drug design of novel potent and selective tetrahydropyrazolo[1,5-a]pyrazines as ATR inhibitors. ACS Med Chem Lett 2015;6:37–41. https://doi.org/10.1021/ml500353p.

[19] Miller MS, Schmidt-Kittler O, Bolduc DM, Brower ET, Chaves-Moreira D, Allaire M, et al. Structural basis of nSH2 regulation and lipid binding in PI3Kα. Oncotarget 2014;5:5198–208. https://doi.org/10.18632/oncotarget.2263.

[20] Han W, Menezes DL, Xu Y, Knapp MS, Elling R, Burger MT, et al. Discovery of imidazo[1,2-*a*]-pyridine inhibitors of pan-PI3 kinases that are efficacious in a mouse xenograft model. Bioorg Med Chem Lett 2016;26:742–6. https://doi.org/10.1016/j.bmcl.2016.01.003.

[21] Miller MS, Maheshwari S, McRobb FM, Kinzler KW, Amzel LM, Vogelstein B, et al. Identification of allosteric binding sites for PI3Kα oncogenic mutant specific inhibitor design. Bioorg Med Chem 2017;25:1481–6. https://doi.org/10.1016/j.bmc.2017.01.012.

[22] Buckbinder L, St. Jean DJJr, Tieu T, Ladd B, Hilbert B, Wang W, et al. STX-478, a mutant-selective, allosteric PI3Kα inhibitor spares metabolic dysfunction and improves therapeutic response in PI3Kα-mutant xenografts. Cancer Discov 2023;13:2432–47. https://doi.org/10.1158/2159-8290.CD-23-0396.

[23] Liu X, Zhou Q, Hart JR, Xu Y, Yang S, Yang D, et al. Cryo-EM structures of cancer-specific helical and kinase domain mutations of PI3Kα. Proc Natl Acad Sci USA 2022;119:e2215621119. https://doi.org/10.1073/pnas.2215621119.

[24] Varkaris A, Pazolli E, Gunaydin H, Wang Q, Pierce L, Boezio AA, et al. Discovery and clinical proof-of-concept of RLY-2608, a first-in-class mutant-selective allosteric PI3Kα inhibitor that decouples antitumor activity from hyperinsulinemia. Cancer Discov 2024;14:240–57. https://doi.org/10.1158/2159-8290.CD-23-0944.

[25] Ketcham JM, Harwood SJ, Aranda R, Aloiau AN, Bobek BM, Briere DM, et al. Discovery of pyridopyrimidinones that selectively inhibit the H1047R PI3Kα mutant protein. J Med Chem 2024;67:4936–49. https://doi.org/10.1021/acs.jmedchem.4c00078.

[26] Huang X, Wang K, Han J, Chen X, Wang Z, Wu T, et al. Cryo-EM structures reveal two allosteric inhibition modes of PI3KαH1047R involving a re-shaping of the activation loop. Structure 2024;32:907–917.e7. https://doi.org/10.1016/j.str.2024.03.007.

[27] Gong GQ, Bilanges B, Allsop B, Masson GR, Roberton V, Askwith T, et al. A small-molecule PI3Kα activator for cardioprotection and neuroregeneration. Nature 2023;618:159–68. https://doi.org/10.1038/s41586-023-05972-2.

[28] Zhang M, Jang H, Nussinov R. The structural basis for Ras activation of PI3Kα lipid kinase. Phys Chem Chem Phys 2019;21:12021–8. https://doi.org/10.1039/C9CP00101H.

[29] Chakrabarti M, Gabelli SB, Amzel LM. Allosteric activation of PI3Kα results in dynamic access to catalytically competent conformations. Structure 2020;28:465–474.e5. https://doi.org/10.1016/j.str.2020.01.010.

[30] Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput 2011;7:525–37. https://doi.org/10.1021/ct100578z.

[31] Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. J Chem Theory Comput 2011;7:2284–95. https://doi.org/10.1021/ct200133y.

[32] **Schrödinger Release 2022-2**: Maestro, Schrödinger, LLC, New York, NY, 2022.

[33] Doerr S, Harvey MJ, Noé F, De Fabritiis G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. J Chem Theory Comput 2016;12:1845–52. https://doi.org/10.1021/acs.jctc.6b00049.

[34] Jacobson MP, Pincus DL, Rapp CS, Day TJF, Honig B, Shaw DE, et al. A hierarchical approach to all-atom protein loop prediction. Proteins 2004;55:351–67. https://doi.org/10.1002/prot.10613.

[35] Jacobson MP, Friesner RA, Xiang Z, Honig B. On the role of the crystal environment in determining protein side-chain conformations. J Mol Biol 2002;320:597–608. https://doi.org/10.1016/s0022-2836(02)00470-9.

[36] Harvey MJ, Giupponi G, Fabritiis GD. ACEMD: accelerating biomolecular dynamics in the microsecond time scale. J Chem Theory Comput 2009;5:1632–9. https://doi.org/10.1021/ct9000685.

[37] Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 2017;14:71–3. https://doi.org/10.1038/nmeth.4067.

[38] Åqvist J, Wennerström P, Nervall M, Bjelic S, Brandsdal BO. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. Chem Phys Lett 2004;384:288–94. https://doi.org/10.1016/j.cplett.2003.12.039.

[39] Sivak DA, Chodera JD, Crooks GE. Time step rescaling recovers continuous-time dynamical properties for discrete-time Langevin integration of nonequilibrium systems. J Phys Chem B 2014;118:6466–74. https://doi.org/10.1021/jp411770f.

[40] Kräutler V, van Gunsteren WF, Hünenberger PH. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. J Comput Chem 2001;22:501–8. https://doi.org/10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V.

[41] Hopkins CW, Le Grand S, Walker RC, Roitberg AE. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. J Chem Theory Comput 2015;11:1864–74. https://doi.org/10.1021/ct5010406.

[42] Harvey MJ, De Fabritiis G. An implementation of the smooth particle Mesh Ewald method on GPU hardware. J Chem Theory Comput 2009;5:2371–7. https://doi.org/10.1021/ct900275y.

[43] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys 1995;103:8577–93. https://doi.org/10.1063/1.470117.

[44] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001.

[45] Páll S, Abraham MJ, Kutzner C, Hess B, Lindahl E. Tackling EXascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In: Markidis S, Laure E, editors. Solving Software Challenges for Exascale. Cham: Springer International Publishing; 2015. p. 3–27. https://doi.org/10.1007/978-3-319-15976-8_1.

[46] Lee PS, Bradshaw RT, Marinelli F, Kihn K, Smith A, Wintrode PL, et al. Interpreting hydrogen-deuterium exchange experiments with molecular simulations: tutorials and applications of the HDXer ensemble reweighting software [Article v1.0. Living J Comput Mol Sci 2021;3: 1521. https://doi.org/10.33011/livecoms.3.1.1521.

[47] Bradshaw RT, Marinelli F, Faraldo-Gómez JD, Forrest LR. Interpretation of HDX Data by Maximum-Entropy Reweighting of Simulated Structural Ensembles. Biophys J 2020;118:1649–64. https://doi.org/10.1016/j.bpj.2020.02.005.

[48] Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide Folding: When Simulation Meets Experiment. Angewandte Chemie International Edition 1999;38:236–40. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M.

[49] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12:2825–30.

[50] Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA: protein complexes. Proc Natl Acad Sci 2009;106:6620–5. https://doi.org/10.1073/pnas.0810961106.
[51] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14;33–8:27–8. https://doi.org/10.1016/0263-7855(96)00018-5.
[52] Raiteri P, Laio A, Gervasio FL, Micheletti C, Parrinello M. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. J Phys Chem B 2006;110:3533–9. https://doi.org/10.1021/jp054359r.
[53] Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: new feathers for an old bird. Comput Phys Commun 2014;185:604–13. https://doi.org/10.1016/j.cpc.2013.09.018.
[54] Efron B., Tibshirani R.J. An Introduction to the Bootstrap. New York: Chapman and Hall/CRC; 1994. https://doi.org/10.1201/9780429246593.
[55] Raniolo S, Limongelli V. Ligand binding free-energy calculations with funnel metadynamics. Nat Protoc 2020;15:2837–66. https://doi.org/10.1038/s41596-020-0342-4.
[56] Trapl D, Spiwok V. The R journal: analysis of the results of metadynamics simulations by metadynminer and metadynminer3d. R J 2022;14:46–58. https://doi.org/10.32614/RJ-2022-057.
[57] Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput 2013;9:3084–95. https://doi.org/10.1021/ct400341p.
[58] Best RB, Vendruscolo M. Structural interpretation of hydrogen exchange protection factors in proteins: characterization of the native state fluctuations of CI2. Structure 2006;14:97–106. https://doi.org/10.1016/j.str.2005.09.012.
[59] Bai Y, Milne JS, Mayne L, Englander SW. Primary structure effects on peptide group hydrogen exchange. Proteins 1993;17:75–86. https://doi.org/10.1002/prot.340170110.
[60] Nguyen D, Mayne L, Phillips MC, Englander SW. Reference parameters for protein hydrogen exchange rates. J Am Soc Mass Spectrom 2018;29:1936–9. https://doi.org/10.1007/s13361-018-2021-z.
[61] Tse A, Verkhivker GM. Molecular dynamics simulations and structural network analysis of c-Abl and c-Src kinase core proteins: capturing allosteric mechanisms and communication pathways from residue centrality. J Chem Inf Model 2015;55:1645–62. https://doi.org/10.1021/acs.jcim.5b00240.
[62] Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. BMC Bioinform 2009;10:168. https://doi.org/10.1186/1471-2105-10-168.
[63] de la Cruz-Herrera CF, Baz-Martínez M, Lang V, El Motiam A, Barbazán J, Couceiro R, et al. Conjugation of SUMO to p85 leads to a novel mechanism of PI3K regulation. Oncogene 2016;35:2873–80. https://doi.org/10.1038/onc.2015.356.
[64] Castelli M, Yan P, Rodina A, Digwal CS, Panchal P, Chiosis G, et al. How aberrant N-glycosylation can alter protein functionality and ligand binding: an atomistic view. Structure 2023;31:987–1004.e8. https://doi.org/10.1016/j.str.2023.05.017.
[65] Joshi H, Prakash MK. Using atomistic simulations to explore the role of methylation and ATP in chemotaxis signal transduction. ACS Omega 2022;7:27886–95. https://doi.org/10.1021/acsomega.2c00792.
[66] Castelli M, Magni A, Bonollo G, Pavoni S, Frigerio F, Oliveira ASF, et al. Molecular mechanisms of chaperone-directed protein folding: Insights from atomistic simulations. Protein Sci 2024;33:e4880. https://doi.org/10.1002/pro.4880.
[67] Galdadas I, Bonis V, Vgenopoulou P, Papadourakis M, Kakoulidis P, Stergiou G, et al. The effect of S427F mutation on RXRα activity depends on its dimeric partner. Chem Sci 2021;12:14700–10. https://doi.org/10.1039/D1SC04465F.