PAPER

Check for updates

How a single mutation alters the protein structure: a simulation investigation on protein tyrosine phosphatase SHP2†

Yingnan Hou, [ab] Xiaoli Lu, [ab] Ziyao Xu, [c] Jiarun Qu [c] and Jing Huang [ID] *[ab]

Protein tyrosine phosphatase SHP2 is a key regulator modulating several signaling pathways. The oncogenic mutation E76K in SHP2 releases the enzyme from an autoinhibited, closed conformation into an active, open conformation. Here, we investigated the conformational dynamics of SHP2 and the effect of the E76K mutation on its conformational ensemble *via* extensive molecular dynamics (MD) and metadynamics (MetaD) simulations. Our simulations provide atomistic details on how the E76K mutated SHP2 prefers the open state and also reveal that the transition between the closed and the open states is highly collective. Several intermediate metastable states during the conformational transition between the closed and the open states were also investigated. Understanding how the single E76K mutation induces the conformational change in SHP2 could facilitate the further design of SHP2 inhibitors.

Introduction

The type 11 non-receptor tyrosine-protein phosphatase SHP2, encoded by the PTPN11 gene, is a critical regulator of signal transduction.[1] It acts on a variety of downstream receptors and cytoplasmic kinases to regulate cell survival and proliferation primarily through activation of the Ras-Raf-MEK-ERK pathway.[2,3] It also functions as a key modulator of the programmed cell death 1 (PD-1) and B- and T-lymphocyte attenuator (BTLA) immune checkpoint pathways.[4,5] SHP2 is an oncogenic tyrosine phosphatase.[6] Activating mutations of SHP2 are related to developmental disordered diseases such as Noonan syndrome and various cancer types including leukemia, neuroblastoma and lung cancer.[4,7,8] Targeting SHP2 has been demonstrated to be a successful anti-cancer strategy and small-molecule inhibitors of SHP2 such as TNO155, RMC-4630, and JAB-3068 are under different phases of clinical trials. Recently Choi *et al.* found that the SHP2-MAPK pathway regulates insulin receptor endocytosis such that SHP2 inhibition prolongs insulin action and improves insulin sensitivity, suggesting that SHP2 is also a valuable therapeutic target for diabetes.[9–11]

SHP2 contains two tandemly arranged Src-homology-2 (SH2) domains (N-SH2 and C-SH2), a catalytic protein tyrosine phosphatase (PTP) domain and a C-terminal tail that has at least two phosphatase sites.[12] The X-ray structure of the wild-type SHP2 (SHP2-WT) reveals that the N-SH2 domain tightly interacts with the PTP domain, blocking the catalytic pocket of the PTP domain being accessible to its substrates and resulting in an autoinhibited closed state (PDB id: 2SHP).[13] Normally, the autoinhibited conformation of SHP2 is activated by the binding of tyrosine-phosphorylated peptides to the tandem SH2 domains, triggering a series of structural rearrangements in the N-SH2 domain to drive its release from the PTP domain.[14] Recent investigations on the activation mechanism of SHP2 proposed that the key allosteric switch triggering SHP2 activation might not be the widely-accepted opening of the N-SH2 binding cleft but the opening of the central beta-sheet of N-SH2 based on extensive molecule dynamics (MD) simulations and free energy calculations.[15,16]

Distinct from the SHP2-WT in the autoinhibited inactivate state, SHP2 with oncogenic mutations shows enhanced basal activity without the stimulation of tyrosine-phosphorylated peptides.[17] The X-ray structure of SHP2 with the E76K mutation (SHP2-E76K), one of the most frequently observed oncogenic mutations, reveals that the N-SH2 domain locates far away from the catalytic pocket of the PTP domain hence the catalytic site is fully exposed to the solvent (PDB id: 6CRF).[18] In comparison with the closed conformation of SHP2-WT, there is a dramatic rearrangement of the three domains in SHP2-E76K, including a 120° rotation of the C-SH2 domain with respect to the PTP domain and the relocation of the N-SH2 domain to a totally different interface with the PTP domain (Fig. 1A).[18] As alternative conformational states of a protein can be

[a]Westlake AI Therapeutics Lab, Westlake Laboratory of Life Sciences and Biomedicine, 18 Shilongshan Road, Hangzhou 310024, Zhejiang, China. E-mail: huangjing@westlake.edu.cn

[b]Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou 310024, Zhejiang, China

[c]BioMap, 2 Kexueyuan South Road, Beijing 100000, China

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2ra07472a
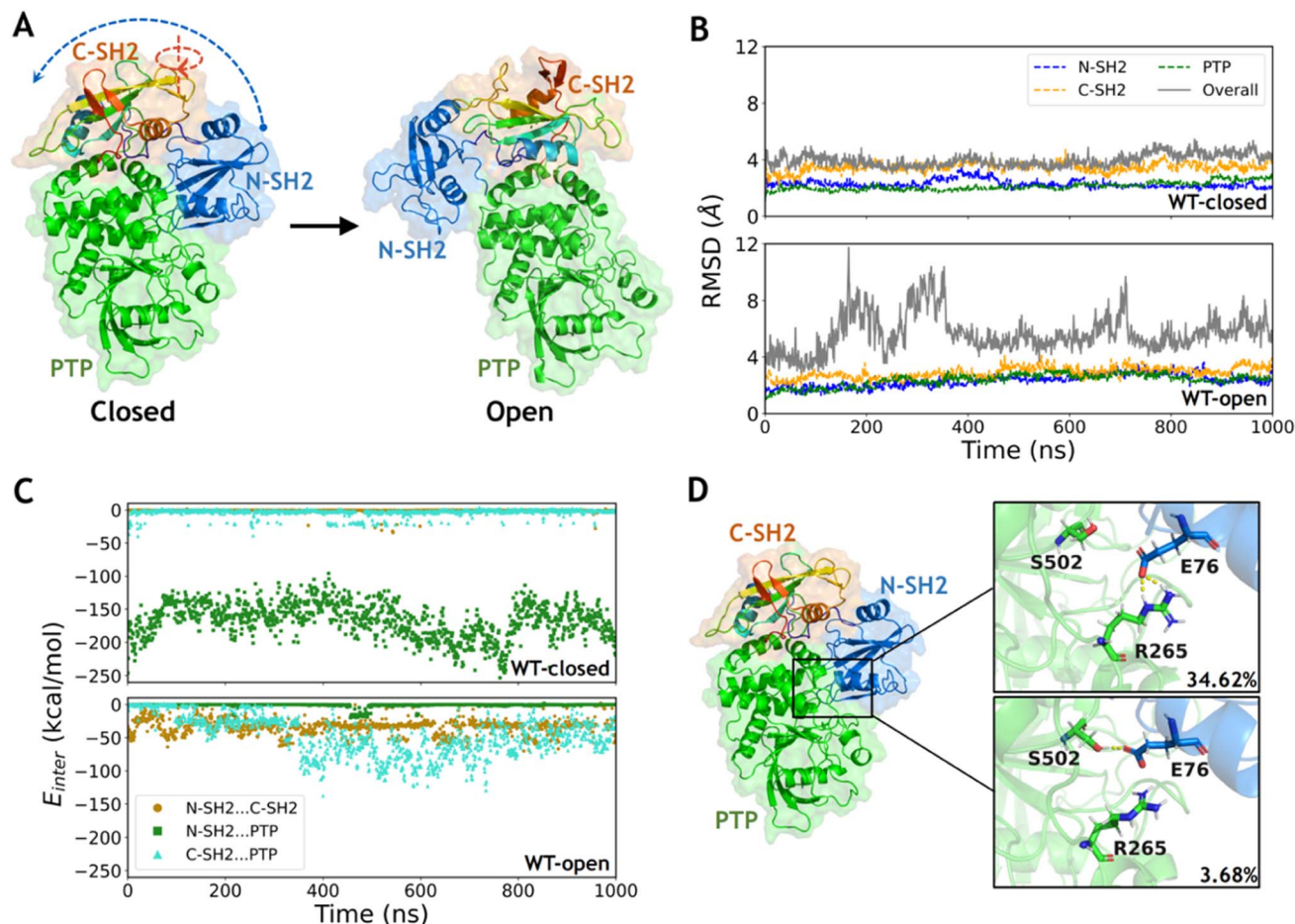
Fig. 1 The closed state of wild type SHP2 is more stable than the open state due to the stronger interaction between the N-SH2 and PTP domains. (A) The major structural differences between the closed and open states are the rotation of the C-SH2 domain and the relocation of N-SH2 to form another interface with the PTP surface. (B) The time evolution of RMSDs in each domain and overall structure in the WT-closed and WT-open systems. The RMSDs in each domain were calculated with alignment to that in the first frame (dashed lines), while the overall RMSDs were calculated with alignment to PTP in the first frame (solid lines). (C) The interaction energies between each domain pair during the simulations of the WT-closed and WT-open systems. (D) The hydrogen bond network formed by side chains between the N-SH2 and PTP domains in the WT-closed system with occupancies labeled.

beneficially targeted for structure-based rational drug design,[19] it would be helpful to unravel the conformational ensemble of SHP2.

NMR spectroscopy and X-ray crystallography studies suggested that wild-type SHP2 exchanges between closed and open conformations, while the E76K mutation shifts the equilibrium toward the open conformational state.[20] Although the dramatic conformational differences between SHP2-WT in the closed conformation and SHP2-E76K in the open conformation can be characterized through their X-ray structures, how the E76K mutation in SHP2 drives such a large conformational change as well as the key points along the transition path are still unclear. It's commonly observed but poorly understood that a single mutation in a protein sequence may alter the structural native state as well as the probability densities of structural ensembles.[21–23] SHP2 can serve as a model system to study how protein conformational transition can be induced by a single mutation.

In this work, we adopted MD simulations and the enhanced sampling metadynamics (MetaD) to study the conformational dynamics of SHP2 and the effect of the E76K mutation on the conformational transitions. With the X-ray structures of wild type SHP2 at the closed conformation (WT-closed) and E76K mutated SHP2 at the open conformation (E76K-open) determined, we generated the structures of wild type SHP2 at the open conformation (WT-open) and E76K mutated SHP2 at the closed conformation (E76K-closed). From 1 μs conventional MD (cMD) simulations, the dynamic properties of two conformational states in both the WT and E76K systems were investigated. The structural analysis and interaction energy calculations showed that the E76K mutation alters the local conformation and results in the weaker interactions between N-SH2 domain and PTP domain. To further investigate whether and how the single mutation induces the conformational transition from closed state to open state, we defined a series of collective variables (CVs) and employed MetaD to perform the

enhanced sampling. Our thermodynamic analysis reveals that E76K mutated SHP2 prefers the open state with high dynamic, while the conformational space for the transition between the two states is high dimensional. Taken together, our dynamic and thermodynamic analysis help to understand how the single mutation E76K alters the conformations of SHP2.

Results

The interaction interface between N-SH2 and PTP domains contributes to a highly stable closed state

The two representative states of SHP2 are mainly different in the relative positions among the three domains (Fig. 1A). The conformations of each separated domain are similar in the two crystal structures, in which the root mean square deviation (RMSD) of the backbone heavy atoms (the same below) are 2.1 Å, 1.5 Å and 1.4 Å in the N-SH2, C-SH2 and PTP domains, respectively. However, the RMSD between the overall structures equals 27.8 Å when aligning with the most similar PTP domain, indicating a large conformational change between the closed and the open states. The open conformation of wild type SHP2 is generated by setting the mutation back in the X-ray structure of SHP2-E76K. Conventional MD simulations of SHP2-WT starting from these two states were performed for 1 μs each (ESI, Table S1,† simulation 1–2). To investigate the dynamics of SHP2 in the simulation systems, the time evolution of RMSDs in each domain and the overall structure were calculated with alignment to those in the first frame (the same below). The time evolution of RMSDs in the three domains are 2.3 ± 0.3 Å, 3.5 ± 0.3 Å, 2.1 ± 0.3 Å in the WT-closed system and 2.3 ± 0.4 Å, 2.9 ± 0.4 Å, 2.4 ± 0.4 Å in the WT-open system, respectively (Fig. 1B), indicating the internal stability of the three domains. The overall RMSDs with alignment to the PTP domain are 4.0 ± 0.4 Å in the WT-closed system and 5.7 ± 1.4 Å in the WT-open system, suggesting that the closed state is more stable (Fig. 1B and S1A†). The higher overall RMSDs indicate the relative positions of the three domains are not stable, especially for the open state (Fig. S1B†).

To investigate the contribution of the interaction among the three domains, the inter-domain non-bonded interaction energies were computed from 1 μs MD simulations of the two systems. In the WT-closed system, the interaction energies between the N-SH2 and C-SH2 domains, as well as between C-SH2 and PTP domains are weak (on average −0.7 and −3.7 kcal mol^{-1}), while the interaction between N-SH2 and PTP domains is very strong with average interaction energy being −166.5 kcal mol^{-1} (Fig. 1C). In contrast, the interactions between N-SH2 and C-SH2, between C-SH2 and PTP, and between N-SH2 and PTP domains are all relatively weak in the WT-open system, with average interaction energies being −33.2, −46.6 and −1.1 kcal mol^{-1}, respectively (Fig. 1C). These results suggest that the interaction interface between the N-SH2 and PTP domain contributes most to stabilizing the closed state.

To further investigate the interaction interface between the N-SH2 and PTP domains, sidechain hydrogen bonding interactions were analyzed. A stable sidechain interaction network is formed in the WT-closed system, including E76 with R265 and

with S502 in 34.62% and 3.68% occupancies, respectively (Fig. 1D). In contrast, no hydrogen bond is formed by side chains between the two domains in the WT-open system. These results indicate that the residue E76 does play an important role to maintain the interaction interface between the N-SH2 and PTP domains for stabilizing the closed state for SHP2-WT. Changing the physicochemical properties of this interfacial residue by mutation most likely lead to alteration in conformations. Next, we focused on the effect of conformation change in SHP2 induced by E76K, a single mutation with opposite electrostatic properties.

E76K mutation drives the escape from the closed state by altering the interaction interface between N-SH2 and PTP domains

To investigate the conformational transition between the closed and open states induced by the typical single mutation E76K in SHP2, MD simulations of SHP2-E76K starting from the closed and the open conformations were carried out for 1 μs each (ESI, Table S1,† simulations 3–4). The time evolution of RMSDs in the three domains (all below 3.2 ± 0.3 Å) are similar with that in the WT systems, indicating that the internal conformations of the three domains are also stable in both the E76K-closed and E76K-open systems (Fig. 2A). However, the overall RMSDs with alignment to the PTP domain are 6.9 ± 0.9 Å in the E76K-closed system and 7.7 ± 3.6 Å in the E76K-open system (Fig. 2A). The observation that both closed and open states are more flexible with the E76K mutation hints an uphill step on the protein fitness landscape.[24] We note that the melting temperature (T_m) of SHP2 reduces from 51.3 ± 0.1 °C to 42.1 ± 0.1 °C with the E76K mutation.[25] Structural inspection shows that the relative movements between domains contribute most to the conformational dynamics, especially in the open state (Fig. S1†). Besides, the overall RMSDs in the E76K-open system became small again after ∼600 ns simulation (less than 4.5 Å), suggesting that the open state of SHP2-E76K might be a shallow minimum (Fig. 2A).[18] The overall RMSDs for the E76K-closed system were significantly larger than those for the WT-closed system, suggesting that E76K mutation might be unfavorable to sustain the closed state.[18,20]

The non-bonded interaction energies among the three domains were calculated from 1 μs MD simulations of the two E76K systems to investigate the contribution of the interaction interface among domains. In the E76K-closed system, the interaction energy between the N-SH2 and PTP domains shows two obvious changes: it rapidly drops from ∼−148.4 kcal mol^{-1} to ∼−129.5 kcal mol^{-1} within the first 50 ns simulation and drops again to ∼−80.5 kcal mol^{-1} after 800 ns (Fig. 2B). This is consistent with the time evolution of the RMSDs of the N-SH2 domain when aligning with the PTP domain (Fig. S1C†), indicating that the relative movement of the N-SH2 domain with respect to the PTP domain leads to the weakening of the interaction energy between them.

Further investigation on sidechain hydrogen bonds between the N-SH2 and the PTP domains finds that no hydrogen bond is formed in the E76K-closed system. The structural analysis for
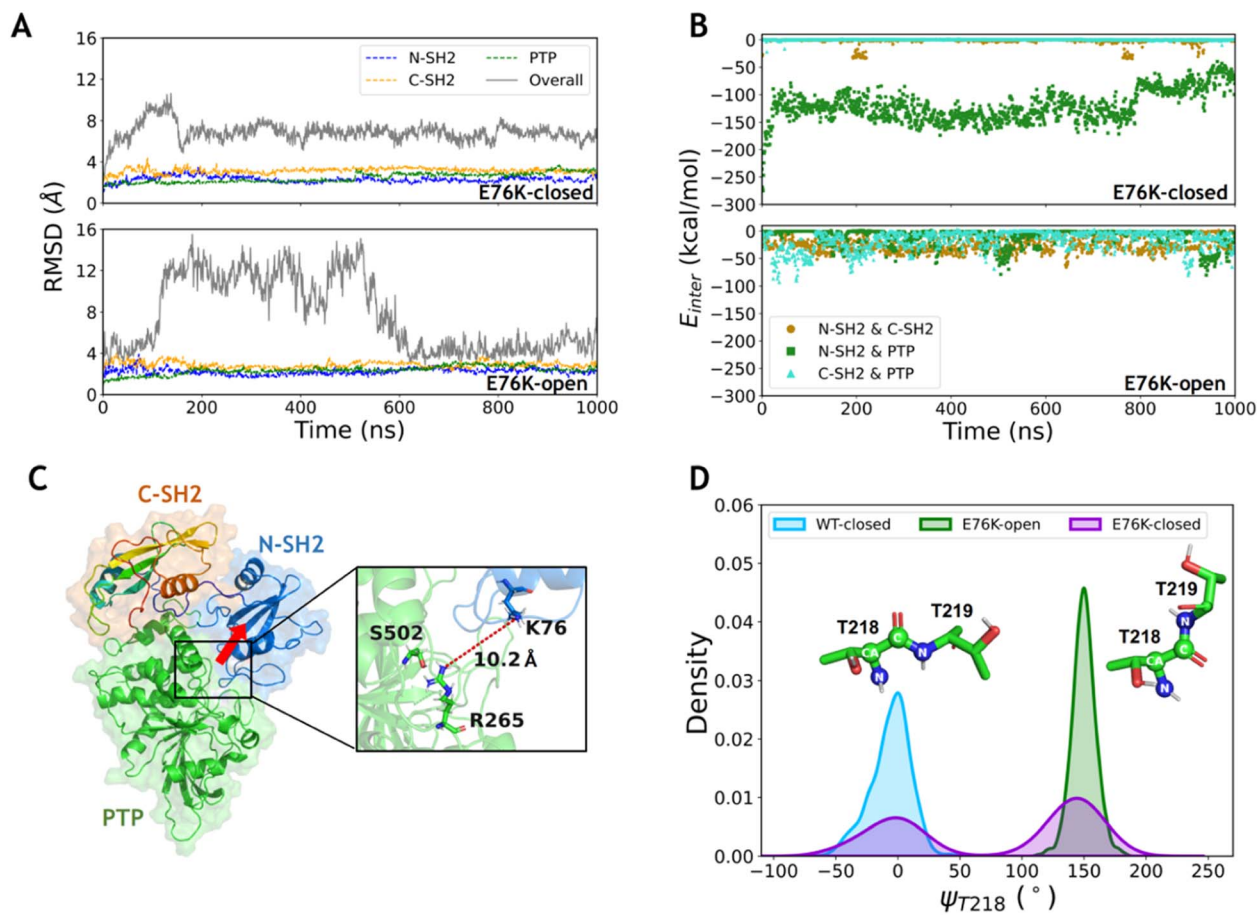
Fig. 2 The E76K mutation eliminates the hydrogen bond network between the N-SH2 and PTP domains. (A) The time evolution of RMSDs of each domain and the overall structure for the E76K-closed and E76K-open systems. The RMSDs of each domain were calculated with alignment to that in the first frame (dashed lines), while the RMSDs of the overall structure were calculated with alignment to PTP in the first frame (solid lines). (B) The interaction energies between each domain pair during the MD simulations of the E76K-closed and E76K-open systems. The interaction energy between N-SH2 and PTP domains decreases during the 1000 ns simulations in the E76K-closed systems. (C) N-SH2 detaches from the PTP domain, and no hydrogen bond formed between them. The conformation is the last MD frame for the E76K-closed system, and the red row indicates the direction of the detachment of N-SH2 from the PTP compared with that in the initial closed state. (D) The distributions of the backbone dihedral angle ψ_{T218} in the E76K-closed, E76K-open and WT-closed systems.

the last frame from 1 μs MD simulation shows that N-SH2 detaches from PTP, resulting the distance between K76 and R265 being larger than 10.2 Å (Fig. 2C). Based on MD simulation results, no new interaction can be formed to compensate the loss of E76-R265 and E76-S502 interactions. Our results indicate that E76K mutation alters the hydrogen bond network in the interaction interface between N-SH2 and PTP domains, thus drives the detachment of N-SH2 from PTP and results in SHP2 escaping from the closed state.

In addition, the backbone dihedral of residue T218 (ψ_{T218}) in the linker region between C-SH2 and PTP domains was calculated according to its possible effect on the motion of C-SH2 domain.[20] Interestingly, the distributions of sampled ψ_{T218} are different in the WT-closed, E76K-open, and E76K-closed systems. The ψ_{T218} was centrally distributed with mean and standard deviation at $-5.9 \pm 16.3°$ in the WT-closed system and $149.9 \pm 9.5°$ in the E76K-open system. For the E76K-closed system a bimodal distribution was found with two peaks that

correspond to the WT-closed and the E76K-open distributions (Fig. 2D). The transition of this key internal coordinate between the closed and open states was captured during the 1 μs MD simulation in the E76K-closed system. However, the conformational transition from the closed state to the open state is not observed during the 1 μs MD simulation, thus enhanced sampling was adopted to further investigate the conformational changes.

Conformational transition from the closed to the open state induced by the E76K mutation is highly collective

Metadynamics is an efficient enhanced sampling technique that biases MD trajectories with a properly chosen set of CVs.[26] To explore the conformational transition between the closed state and the open state of SHP2 using 2-dimensional (2D) well-tempered metadynamics (WT-MetaD) simulations, we defined several CVs characterizing the structural differences between the two states. Two distances were defined as a pair of CVs using

the centers of mass (CoMs) of three groups of Cα atoms selected from rigid regions near the two different interaction interfaces between the N-SH2 and the PTP domains in the two states ($Dist_{NP}^1$ and $Dist_{NP}^2$, Fig. 3A, Table 1). In addition, to characterize the global conformational difference, in particular the rotation of the N-SH2 domain with respect to the PTP and the C-SH2 domains, a pseudo dihedral angle was defined by carefully selecting four Cα atoms from the three domains (Dih_{NCPP}, Fig. 3A and Table 1), resulting in a difference of about 125° between the two states. Taking Dih_{NCPP} together with the key backbone dihedral ψ_{T218} (Fig. 2D and Table 1), a pair of CVs were constructed to perform MetaD simulations. In addition, the distance between the Cβ atoms of residues K76 and R265

($Dist_{K76\_R265}$, Table 1) combined with Dih_{NCPP} were also used as a pair of CVs for MetaD simulations. We note that all three combinations of paired CVs can distinguish the closed and open states of the crystal structures and conformations from cMD simulations (Fig. S2–S4†). The 2D WT-MetaD simulations for the E76K mutated and the WT systems were carried out starting from the closed state with initial conformations taken from the corresponding 100 ns cMD trajectories (simulation IDs 5–10, Table S1–S2†). Exactly the same biasing parameters such as the height and the width of the Gaussian bias potential were used for the E76K and the WT systems (Table S2†).

Free energy landscapes projected on the three pairs of CVs for the E76K and WT systems were generated after ~100 ns
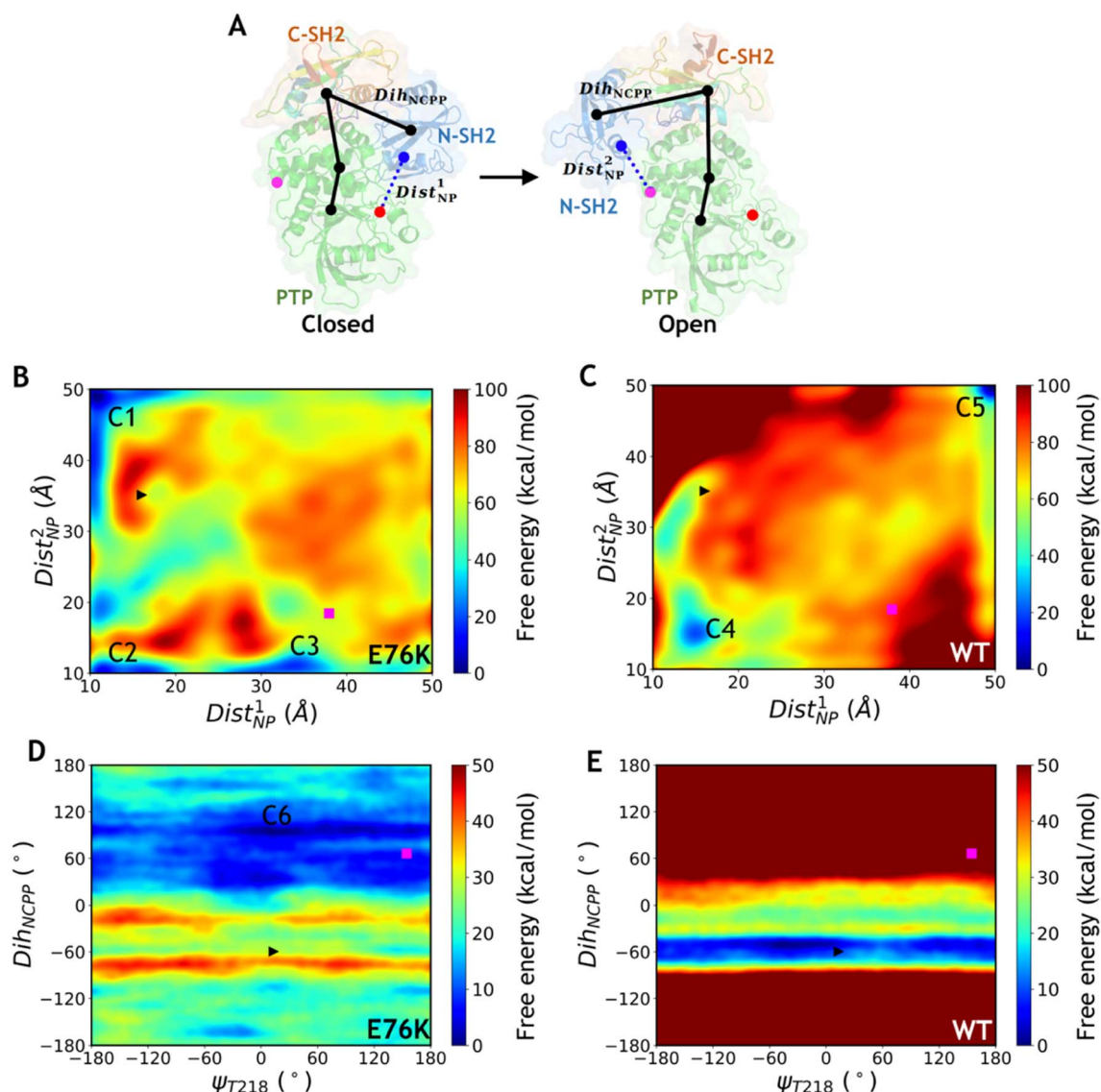
Fig. 3 Free energy landscapes obtained by 2D well-tempered MetaD simulations of the E76K and WT systems. (A) Schematic diagram for the definitions of $Dist_{NP}^1$, $Dist_{NP}^2$ and Dih_{NCPP}. The colored points represent the center of mass of Cα atoms at the interface between N-SH2 and PTP domains. The black points represent the Cα atoms of residues S44, T168, V497 and Q510 for Dih_{NCPP}. (B–C) Free energy landscapes projected on the distance CVs $Dist_{NP}^1$ and $Dist_{NP}^2$ for the E76K (B) and WT (C) systems. (D–E) Free energy landscapes projected on the dihedral CVs ψ_{T218} and Dih_{NCPP} for the E76K (D) and WT (E) systems. The initial closed state is marked with a black right triangle and the target open state is marked with a magenta square. Some intermediate metastable states are labeled as C1–C6.

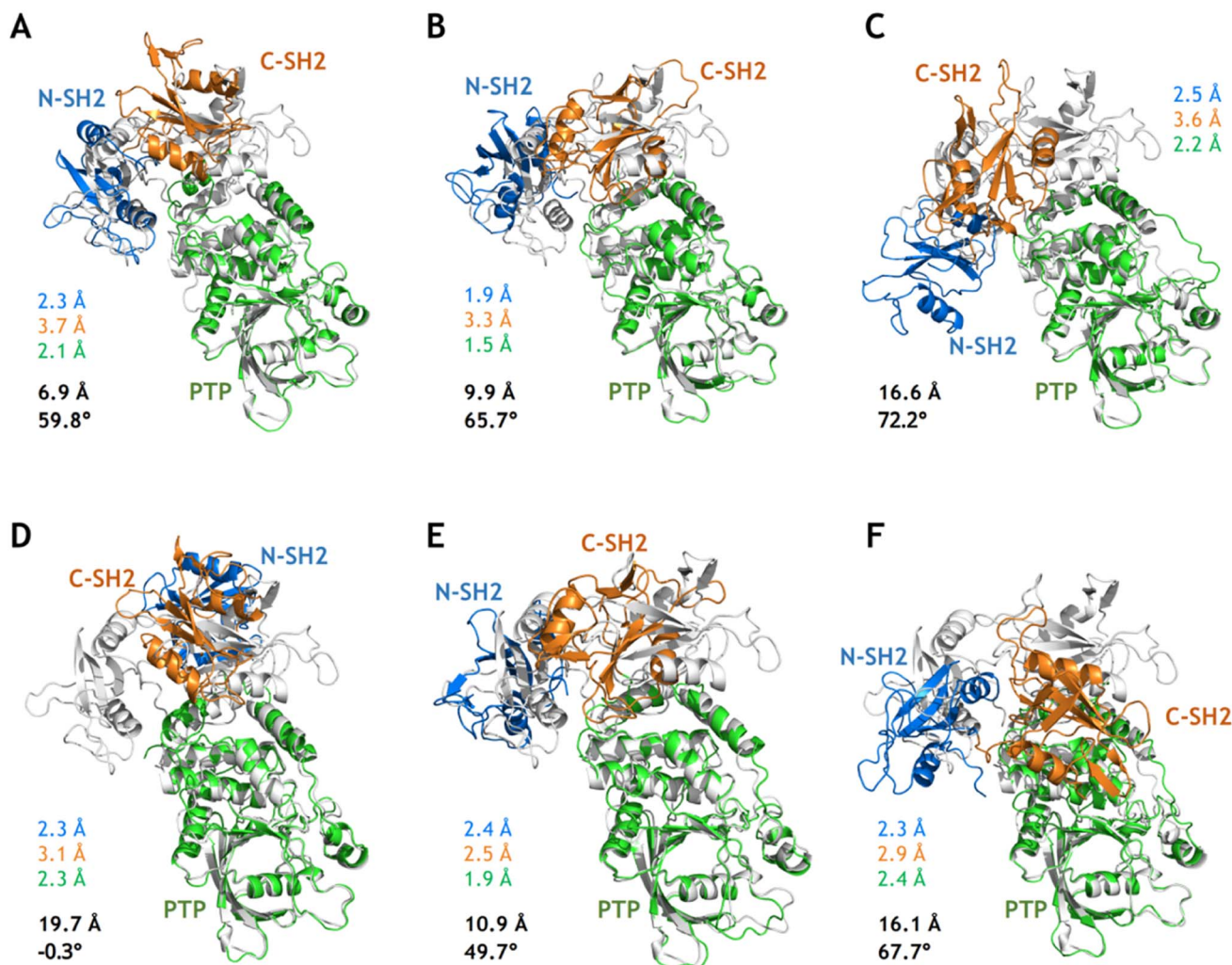
Fig. 4 Conformations closest to the target open state from MetaD simulations and after relaxation from cMD simulations. (A–C) The conformations closest to the target open state from the 100 ns MetaD simulations with the distance CVs ($Dist_{NP}^1$ and $Dist_{NP}^2$, A), the dihedral CVs (ψ_{T218} and Dih_{NCPP}, B) and the distance and dihedral CVs ($Dist_{K76\_R265}$ and Dih_{NCPP}, C), respectively. (D–F) The relaxed conformations after 100 ns cMD simulations for the corresponding conformations in (A–C). The RMSDs in the N-SH2 (blue), C-SH2 (orange), PTP (green) domains with alignment to that in the target open state (light color) are labeled nearby the corresponding conformations. The overall RMSD (black) with alignment to the PTP domain in the target open state and the Dih_{NCPP} are also labeled.

MetaD simulation with the distance CVs ($Dist_{NP}^1$ and $Dist_{NP}^2$) for the E76K system, C1 adopts a low $Dist_{NP}^1$ and a large $Dist_{NP}^2$, C2 adopts both low $Dist_{NP}^1$ and $Dist_{NP}^2$, and C3 adopts a large $Dist_{NP}^1$ and a low $Dist_{NP}^2$, respectively (Fig. 3B). The relative positions of the three domains in C1 are more similar to the closed state than the open state, with the overall RMSDs with alignment to PTP in the initial closed and open states being 18.8 Å and 25.5 Å, respectively (Fig. 5A). The three domains, especially the PTP domain, underwent certain degrees of deformation by themselves, with the RMSDs in the N-SH2, C-SH2 and PTP domains when aligning to those in the open state being 3.9 Å, 3.2 Å and 5.4 Å, respectively (Fig. 5A). Besides, the alpha helices in the long linker region between the C-SH2 and the PTP domains (residues N217 to G246) unwind to a loose state (Fig. 5A). For the C2 state, the N-SH2 domain rotated along the PTP domain with

the minimal distance of Cα atoms between the two domains being 4.2 Å, resulting in both low $Dist_{NP}^1$ and $Dist_{NP}^2$ (Fig. 5B). The C3 conformation represents a metastable state near the target open state, with the overall RMSDs with alignment to PTP in the open state being 15.0 Å and Dih_{NCPP} being 47.3° (Fig. 5C).

For the intermediate states C4–C5 from the MetaD simulation with the distance CVs for the WT system, the relative positions of the three domains are distinct from both the initial closed and the targe open states, with overall RMSD and Dih_{NCPP} are 30.3 Å and −57.7° in the C4 state (Fig. 5D), and 37.4 Å and −66.7° in the C5 state (Fig. 5E), respectively. Deformation of the PTP domain was also observed with the PTP RMSDs being 6.0 Å (C4 state) and 6.4 Å (C5 state). The difference between the C1–C3 metastable states for SHP2-E76K and the C4–C5 states for SHP2-WT was evident. The intermediate state C6 was
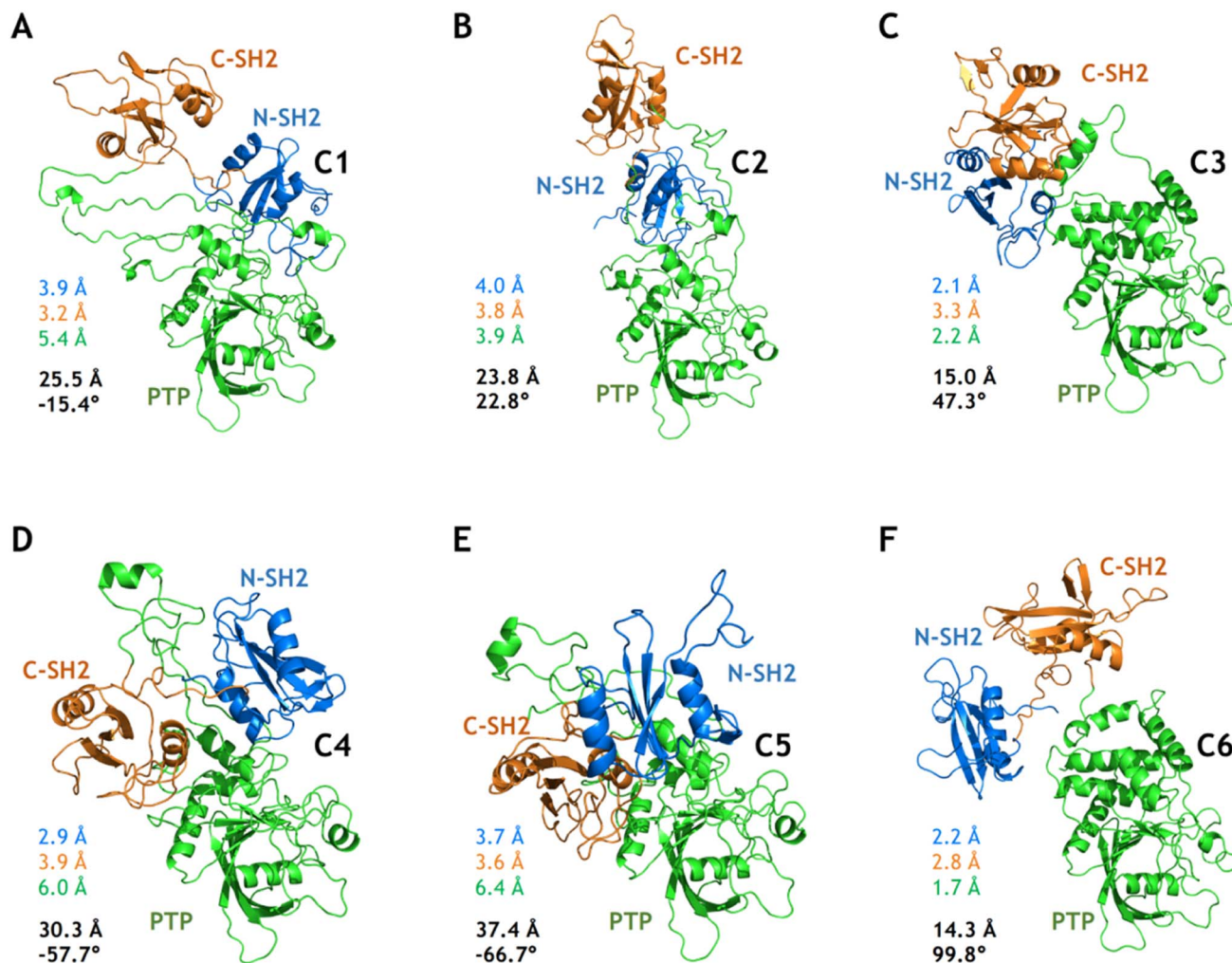
Fig. 5 Six intermediate metastable states captured from the MetaD simulations. The conformations in A–F are mapped to the C1–C6 in Fig. 3. The RMSDs in the N-SH2 (blue), C-SH2 (orange), PTP (green) domains with alignment to that in the target open state (light color) are labeled. The overall RMSD (black) with alignment to the PTP domain in the target open state together with the Dih_{NCPP} are also labeled.

sampled from the MetaD simulation with the dihedral CVs (ψ_{T218} and Dih_{NCPP}) for the E76K system, with overall RMSD being 14.3 Å and Dih_{NCPP} being 99.8° (Fig. 5F). We note that all the six intermediate metastable states sampled from the MetaD simulations were never observed in trajectories of the four 1 μs cMD simulations starting from the closed or the open states.

Discussion and conclusion

In this work, we investigated the conformational dynamics of SHP2 and the effect of the single E76K mutation on the transition of conformational states using explicit-solvent all-atom simulations starting from two previously determined experimental structures. Microsecond conventional MD simulations indicate that SHP2 in the open state is much more dynamics compared with the autoinhibited closed state in solution, although each domain is stable by itself. Furthermore, well-tempered MetaD simulations provide thermodynamic

evidence on that E76K mutated SHP2 prefers the open state and that the transition between the closed and the open states is highly collective.

In particular, our simulations illustrate that both the open and close states of SHP2 exits as basins in SHP2-WT and SHP2-E76K, while the shape of the basins is altered in the free energy landscape as the probability densities of the structural ensembles are changed by the E76K mutation. The conformational transition of multiple conformations can happen at all scales driven by the free energy of the local structures.[27] The long loop linkers between adjacent domains in SHP2 provide the structural basis for the relatively free movement between domains. In SHP2-E76K, the single mutation significantly reduces the interaction energy between the N-SH2 and the PTP domains, but increases the disorder of the three domains and results in higher conformational flexibility.

Significant advances have been achieved recently in protein structure prediction using end-to-end deep neural networks

(DNNs), for example the AlphaFold[28–30] and the RoseTTAFold.[31] Further explorations based on AlphaFold have been actively pursued from different aspects, such as sampling alternative conformational states,[32] establishing protein fold atlas,[33] and detecting topologically links in predicted structures.[34] AlphaFold was designed to predict the native structure from coevolution information of a target sequence, where the effect of single mutation was largely filtered during the establishment of coevolution from multiple sequence alignment (MSA). Consequently, AlphaFold shows limited accuracy in predicting the impact of missense mutations as reported by several studies.[35–37] It is known that one single mutation in a sequence may completely alter the native structure, with additional examples like GA98/GB98 [21] and APOE3/APOE4.[38] It is also common that a single mutation can greatly impact the stability of a protein as shown in mutation-related thermodynamic databases.[39,40] Actually, understanding the impact of missense mutations on protein structure and dynamics helps to reveal their biological effects as well as roles in related diseases.[36] Here we dissected the conformational transition induced by a single mutation for the SHP2 system from the physical point of view *via* atomistic simulations. We hope simulation studies could provide helpful insights on the development of DNN-based methods for mutation-sensitive structural prediction.

MetaD is a widely adopted method to extend the sampling capabilities of MD simulations for studying conformational transitions. With the well pre-defined CVs to describe the most relevant (typically also slowest) motion of the system, MetaD simulations have been successfully applied to understand complicated biological processes. For example, using the residual helicity and coordination number as CVs in MetaD simulations, we recently investigated the coupled binding and unwinding process in the α-helical APP TMD recognized by γ-secretase.[41] In this work, with three sets of manually crafted CVs, we were also able to observe the conformational transition from the close state to the open state in the E76K mutated SHP2 *via* MetaD simulations. However, the free energy landscapes projected on the three sets of CVs are different from each other, and the major conformational states samples by MetaD with different CVs are also significantly different. Both phenomena indicate that the conformational dynamics during the transition might involve a large number of degrees of freedom and be highly collective. The 2D CVs in this study are not sufficient to describe such high dimensional conformational dynamics.

Indeed, one of the main difficulties of bias-based enhanced sampling methods such as MetaD lies in the proper choice of CVs. To maintain the efficiency of exploring the high dimensional conformational space, only a very limited number of CVs can be used, however they need to cover all the slow modes of motion for the simulation system. It is extremely challenging to meet these two conflicting requirements to define proper CVs. In other words, selection of good CVs for MetaD requires to have sufficient prior knowledge of the complicated system, which can often only be gained after successful sampling *via* MetaD simulations. To alleviate the problem of optimal choice of CVs, some strategies have been implemented, including bias-exchange MetaD,[42] parallel bias MetaD,[43] and DNN-based CV

optimization method such as RAVE.[44] In addition, we note that CV-independent enhanced sampling methods such as temperature replica exchange or solute scaling might be beneficial for studying such complicated processes.[45,46]

SHP2 is a promising drug target due to its oncogenic role in multiple cancer-critical cellular processes in human cancers.[47] Efforts to develop small molecules that target SHP2 are ongoing, and several SHP2 allosteric inhibitors are currently in clinical trials for the treatment of solid tumors.[48] However, while the reported allosteric inhibitors are highly effective against the wild type SHP2, none of them show significant activity against the most frequent oncogenic SHP2 variants including the E76K and the G60R that drive leukemogenesis.[20,48] In addition, E76K is one of the most observed SHP2 activating mutations in diseases with activation related to the close-to-open conformational change discussed in this work. The small-molecule inhibitor SHP099 has been reported to restore the active form of E76K mutated SHP2, which interacts with the long loop linking the N-SH2 and C-SH2 domains, the loop linking C-SH2 and PTP, as well as the helix of PTP domain.[20] These loop linkers are found to be important for the flip of the N-SH2 domain and the rotation of the C-SH2 domain in the structural analysis of the multiple key intermediate states from the MetaD simulations. Therefore, our study of the conformational transitions induced by the E76K mutation in SHP2 may provide useful insights into the further drug design targeting SHP2. It would be interesting to simulate the conformational dynamics of SHP2 in the presence of SHP099, which will be pursued in future work.

Methods

System setup

The X-ray structure of the wild-type SHP2 (PDB id: 2SHP, chain A and residues K2 to L525) was used to prepare the initial atomic coordinates of the MD simulation systems starting from the closed state. Missing residues were patched based on other available SHP2 structures. CHARMM[49,50] was used to build the missing hydrogen atom coordinates and perform energy minimization. For the WT-closed system, the SHP2-WT protein was put at the center of a cubic box with a length of 100 Å, which is large enough to contain the protein and at least 16 Å of solvent on all dimensions. It was then solvated with 55 586 explicit TIP3P[51] water molecules and one sodium ion to neutralize the system, resulting in a system of 175 137 atoms in total. For the E76K-closed system, 55 587 TIP3P water molecules and one chloride ion were added to solvate the SHP2-E76K protein in the closed conformational state, resulting in a total of 175 147 atoms.

Similarly, the crystallographic structure of SHP2-E76K (PDB id: 6CRF, chain A) was used to prepare the initial atomic coordinates for the systems starting from the open state. A cubic box with a length of 100 Å was used. For the WT-open system, SHP2 was solvated with 55 586 TIP3P water molecules and one sodium ion (in total 175 137 atoms). While for the E76K-open system, SHP2-E76K was solvated with 55 587 TIP3P water molecules and one chloride ion (in total 175 147 atoms).

Proteins were modeled by the CHARMM36m force field,[52] which is both accurate and transferable for protein simulations.[53]

Molecular dynamics and metadynamics simulations

For the four systems, MD simulations were performed using OpenMM[54,55] with periodic boundary condition (PBC) applied in the isothermal–isobaric (NPT) ensemble. The long-range interactions were evaluated using particle-mesh Ewald (PME) summation with a 12 Å cutoff in real space, with Ewald error tolerance set to 0.001.[56] The Lennard-Jones interactions were truncated at 12 Å with an atom-based force switching function starting at 10 Å. Hydrogen-containing bonds were constrained and the Langevin integrator was used with a time step of 2 fs. The temperature was maintained at 298 K using the Andersen thermostat[57] and the pressure was maintained at 1 atm using a Monto Carlo barostat[58] with the box dimension being updated every 25 steps. Frames were saved every 800 ps.

Metadynamics simulations were carried out using OpenMM with a plugin for PLUMED.[59] The initial configurations for MetaD simulations were extracted from the corresponding cMD frames (simulation ID: 3) at 100 ns. Simulation details were the same as cMD simulations. The CVs employed in the MetaD simulations were defined referred to the crystal structures in the closed and open states (Table 1). The parameters for applying bias potential in MetaD can be found in the ESI Table S2.†

Analysis of simulation trajectories

To analysis the sampled structures, VMD[60] and MDAnalysis[61] were used to process the trajectories from the MD and MetaD simulations. The structural stability was characterized by RMSD calculations for backbone atoms with different ways of alignment. Hydrogen bonds were counted with a donor–acceptor distance cut-off of 3.5 Å and an angle cut-off of 20°. The non-bonded interaction energies between domains were calculated with CHARMM using the same non-bonded parameters as MD simulations. The free energy profiles were constructed with PLUMED. All structural figures were generated using PyMOL (V2.3.4, http://pymol.org/).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 M. D. Asmamaw, et al., A comprehensive review of SHP2 and its role in cancer, Cellular Oncology, 2022, 1–25.

2 J. Fiebelkow, et al., The tyrosine phosphatase SHP2 increases robustness and information transfer within IL-6-induced JAK/STAT signalling, Cell Commun. Signaling, 2021, 19(1), 1–19.

3 T. Noguchi, et al., Role of SH-PTP2, a protein-tyrosine phosphatase with Src homology 2 domains, in insulin-stimulated Ras activation, Mol. Cell. Biol., 1994, 14(10), 6674–6682.

4 J. Zhang, F. Zhang and R. Niu, Functions of Shp2 in cancer, J. Cell. Mol. Med., 2015, 19(9), 2075–2083.

5 J. Li, et al., PD-1/SHP-2 inhibits Tc1/Th1 phenotypic responses and the activation of T cells in the tumor microenvironment, Cancer Res., 2015, 75(3), 508–518.

6 G. Chan, D. Kalaitzidis and B. G. Neel, The tyrosine phosphatase Shp2 (PTPN11) in cancer, Cancer Metastasis Rev., 2008, 27(2), 179–192.

7 M. Tartaglia, et al., Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome, Nat. Genet., 2001, 29(4), 465–468.

8 M. Tartaglia, et al., Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia, Nat. Genet., 2003, 34(2), 148–150.

9 E. Choi, et al., Mitotic Checkpoint Regulators Control Insulin Signaling and Metabolic Homeostasis, Cell, 2016, 166(3), 567–581.

10 E. Choi, et al., Mitotic regulators and the SHP2-MAPK pathway promote IR endocytosis and feedback regulation of insulin signaling, Nat. Commun., 2019, 10(1), 1473.

11 E. Choi and H. Yu, Spindle Checkpoint Regulators in Insulin Signaling, Front. Cell Dev. Biol., 2018, 6, 161.

12 B. G. Neel, H. Gu and L. Pao, The 'Shp'ing news: SH2 domain-containing tyrosine phosphatases in cell signaling, Trends Biochem. Sci., 2003, 28(6), 284–293.

13 P. Hof, et al., Crystal structure of the tyrosine phosphatase SHP-2, Cell, 1998, 92(4), 441–450.

14 S. Pluskey, et al., Potent stimulation of SH-PTP2 phosphatase activity by simultaneous occupancy of both SH2 domains, J. Biol. Chem., 1995, 270(7), 2897–2900.

15 M. Anselmi and J. S. Hub, An allosteric interaction controls the activation mechanism of SHP2 tyrosine phosphatase, Sci. Rep., 2020, 10(1), 1–15.

16 M. Anselmi and J. S. Hub, The loops of the N-SH2 binding cleft do not serve as allosteric switch in SHP2 activation, Proc. Natl. Acad. Sci. U. S. A., 2021, 118(17), e2025107118.

17 J. R. LaRochelle, et al., Structural and Functional Consequences of Three Cancer-Associated Mutations of the Oncogenic Phosphatase SHP2, Biochemistry, 2016, 55(15), 2269–2277.

18 J. R. LaRochelle, et al., Structural reorganization of SHP2 by oncogenic mutations and implications for oncoprotein resistance to allosteric inhibition, Nat. Commun., 2018, 9(1), 1–10.

19 I. G. Rodriguez-Bussey, U. Doshi and D. Hamelberg, Enhanced molecular dynamics sampling of drug target conformations, *Biopolymers*, 2016, **105**(1), 35–42.

20 R. A. Pádua, *et al.*, Mechanism of activating mutations and allosteric drug inhibition of the phosphatase SHP2, *Nat. Commun.*, 2018, **9**(1), 1–14.

21 P. A. Alexander, *et al.*, A minimal sequence code for switching protein structure and function, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**(50), 21149–21154.

22 M. Lorch, *et al.*, Effects of core mutations on the folding of a β-sheet protein: implications for backbone organization in the I-state, *Biochemistry*, 1999, **38**(4), 1377–1385.

23 X. I. Ambroggio and B. Kuhlman, Design of protein conformational switches, *Curr. Opin. Struct. Biol.*, 2006, **16**(4), 525–530.

24 P. A. Romero and F. H. Arnold, Exploring protein fitness landscapes by directed evolution, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**(12), 866–876.

25 C. Romero, *et al.*, A cellular target engagement assay for the characterization of SHP2 (PTPN11) phosphatase inhibitors, *J. Biol. Chem.*, 2020, **295**(9), 2601–2613.

26 G. Bussi and A. Laio, Using metadynamics to explore complex free-energy landscapes, *Nat. Rev. Phys.*, 2020, **2**(4), 200–212.

27 K.-i. Okazaki, *et al.*, Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**(32), 11844–11849.

28 R. Evans, *et al.*, Protein complex prediction with AlphaFold-Multimer, *bioRxiv*, 2021, preprint.

29 J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold, *Nature*, 2021, **596**(7873), 583–589.

30 A. W. Senior, *et al.*, Improved protein structure prediction using potentials from deep learning, *Nature*, 2020, **577**(7792), 706–710.

31 M. Baek, *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network, *Science*, 2021, **373**(6557), 871–876.

32 D. Del Alamo, *et al.*, Sampling alternative conformational states of transporters and receptors with AlphaFold2, *Elife*, 2022, **11**, e75751.

33 T. Xie, *et al.*, Rational exploration of fold atlas for human solute carrier proteins, *Structure*, 2022, **30**(9), 1321–1330.

34 Y. Hou, *et al.*, Topological Links in Predicted Protein Complex Structures Reveal Limitations of AlphaFold, *bioRxiv*, 2022, preprint.

35 M. A. Pak, *et al.*, Using AlphaFold to predict the impact of single mutations on protein stability and function, *bioRxiv*, 2021, preprint.

36 G. R. Buel and K. J. Walters, Can AlphaFold2 predict the impact of missense mutations on structure?, *Nat. Struct. Mol. Biol.*, 2022, **29**(1), 1–2.

37 Y. Zhang, *et al.*, Applications of AlphaFold beyond Protein Structure Prediction, *bioRxiv*, 2021, preprint.

38 C. Frieden and K. Garai, Structural differences between apoE3 and apoE4 may be useful in developing therapeutic

39 R. Nikam, *et al.*, ProThermDB: thermodynamic database for proteins and mutants revisited after 15 years, *Nucleic Acids Res.*, 2020, **49**(D1), D420–D424.

40 K. A. Bava, *et al.*, ProTherm, version 4.0: thermodynamic database for proteins and mutants, *Nucleic Acids Res.*, 2004, **32**(1), D120–D121.

41 X. Lu and J. Huang, A thermodynamic investigation of amyloid precursor protein processing by human γ-secretase, *Commun. Biol.*, 2022, **5**(1), 1–10.

42 S. Piana and A. Laio, A bias-exchange approach to protein folding, *J. Phys. Chem. B*, 2007, **111**(17), 4553–4559.

43 J. Pfaendtner and M. Bonomi, Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics, *J. Chem. Theory Comput.*, 2015, **11**(11), 5062–5067.

44 J. M. L. Ribeiro, *et al.*, Reweighted autoencoded variational Bayes for enhanced sampling (RAVE), *J. Chem. Phys.*, 2018, **149**(7), 072301.

45 J. Huang and A. D. MacKerell Jr, Induction of peptide bond dipoles drives cooperative helix formation in the (AAQAA) 3 peptide, *Biophys. J.*, 2014, **107**(4), 991–997.

46 M. Yang, J. Huang and A. D. MacKerell Jr, Enhanced conformational sampling using replica exchange with concurrent solute scaling and hamiltonian biasing realized in one dimension, *J. Chem. Theory Comput.*, 2015, **11**(6), 2855–2867.

47 X. Yuan, *et al.*, Recent advances of SHP2 inhibitors in cancer therapy: Current development and clinical application, *J. Med. Chem.*, 2020, **63**(20), 11368–11396.

48 D. Raveendra-Panickar, *et al.*, Discovery of novel furanylbenzamide inhibitors that target oncogenic tyrosine phosphatase SHP2 in leukemia cells, *J. Biol. Chem.*, 2022, **298**(1), 101477.

49 B. R. Brooks, *et al.*, CHARMM: the biomolecular simulation program, *J. Comput. Chem.*, 2009, **30**(10), 1545–1614.

50 B. R. Brooks, *et al.*, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem.*, 1983, **4**(2), 187–217.

51 W. L. Jorgensen, *et al.*, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 1983, **79**(2), 926–935.

52 J. Huang, *et al.*, CHARMM36m: an improved force field for folded and intrinsically disordered proteins, *Nat. Methods*, 2017, **14**(1), 71–73.

53 Y. Xu and J. Huang, Validating the CHARMM36m protein force field with LJ-PME reveals altered hydrogen bonding dynamics under elevated pressures, *Commun. Chem.*, 2021, **4**(1), 1–11.

54 P. Eastman and V. Pande, OpenMM: A hardware-independent framework for molecular simulations, *Comput. Sci. Eng.*, 2010, **12**(4), 34–39.

55 P. Eastman, *et al.*, OpenMM 7: Rapid development of high performance algorithms for molecular dynamics, *PLoS Comput. Biol.*, 2017, **13**(7), e1005659.

agents for Alzheimer's disease, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**(23), 8913–8918.

56 U. Essmann, *et al.*, A smooth particle mesh Ewald method, *J. Chem. Phys.*, 1995, **103**(19), 8577–8593.

57 H. Andersen, Molecular dynamics at constant temperature and/or pressure, *J. Chem. Phys.*, 1980, **72**, 2384.

58 J. Åqvist, *et al.*, Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm, *Chem. Phys. Lett.*, 2004, **384**(4–6), 288–294.

59 G. A. Tribello, *et al.*, PLUMED 2: New feathers for an old bird, *Comput. Phys. Commun.*, 2014, **185**(2), 604–613.

60 W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graphics*, 1996, **14**(1), 33–38.

61 R. J. Gowers, *et al.*, MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations, in *Proceedings of the 15th python in science conference*, SciPy Austin, TX, 2016.