



An allosteric regulation mechanism of Arabidopsis Serine/Threonine kinase 1 (SIK1) through phosphorylation



Junxi Mu¹, Jiali Zhou¹, Qingqiu Gong*, Qin Xu*

State Key Laboratory of Microbial Metabolism & Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, China

ARTICLE INFO

Article history:

Received 21 October 2021
Received in revised form 23 December 2021
Accepted 23 December 2021
Available online 28 December 2021

Keywords:

Hippo
SIK1
Phosphorylation
MD simulation
Allosteric pathway

ABSTRACT

The Arabidopsis Serine/Threonine Kinase 1 (SIK1) is a Sterile 20 (STE20)/Hippo orthologue that is also categorized as a Mitogen-Activated Protein Kinase Kinase Kinase (MAP4K). Like its animal and fungi orthologues, SIK1 is required for cell cycle exit, cell expansion, polarity establishment, as well as pathogenic response. The catalytic activity of SIK1, like other MAPKs, is presumably regulated by its phosphorylation states. Since no crystal structure for SIK1 has been reported yet, we built structural models for SIK1 kinase domain in different phosphorylation states with different pocket conformation to see how this kinase may be regulated. Using computational structural biology methods, we outlined a conduction path in which a phosphorylation site on the A-loop regulates the catalytic activity of SIK1 by controlling the closing or opening of the catalytic pocket at the G-loop. Furthermore, with analyses on the dynamic motions and *in vitro* kinase assay, we confirmed that three key residues in this conduction path, Lys278, Glu295, and Arg370, are indeed important for the kinase activity of SIK1. Since these residues are conserved in all STE20 kinases examined, the regulatory mechanism that we discovered may be common in STE20 kinases.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Living organisms rely on dynamic signaling networks that often involve protein kinases to transduce signals and to respond to environmental and developmental cues. The mitogen-activated protein kinase (MAPK) cascades are among the most thoroughly studied pathways, with reported functions in cell proliferation, polarity establishment, differentiation, and cell death. A typical MAPK cascade consists of at least three classes of kinases, including Mitogen-Activated Protein Kinase (MAPK), MAPK Kinase (MAP2K), and MAPK Kinase Kinase (MAP3K), forming a kinase cascade [1,2]. In many cases, the MAP3K can be phosphorylated by the Mitogen-Activated Protein Kinase Kinase Kinase (MAP4K), which can also phosphorylate many other targets outside the MAPK pathway [3–5]. Together with their diverse sub-cellular distributions, the MAP4Ks are therefore often a hotspot and hub in MAPK and other signaling pathways.

One example of the versatile MAP4Ks is Hippo/MST1/2, the STE20 family kinase that serves as the core component of the

Salvador-Warts-Hippo (SWH) pathway, or the Hippo pathway. Hippo was initially identified in *Drosophila* genetic screens for tissue overgrowth (“ugly as a Hippo”) [6,7], and MST1/2 was initially characterized as a Mammalian Sterile Twenty-like (MST) kinase that responds to severe stresses [8] and characterized later as a Hippo counterpart in human. The SWH pathway is essentially a kinase cascade that involves a pair of kinases, Hippo/Mammalian STE20-like 1 and 2 (MST1/2), and Warts (Wts)/Large Tumor Suppressor kinases 1/2 (LATS1/2), and a pair of scaffold proteins, Salvador/SAV1, and Mps One Binder 1 (MOB1)/Mats. Downstream to the kinase cascade are two oncogenic transcriptional co-activators Yorkie (Yki)/Yes-Associated Protein (YAP)/Transcriptional Activator with PDZ-binding motif (TAZ). Upon activation, Hpo/MST1/2 kinases form a complex with Salvador (Sav)/SAV1 and phosphorylate and activate the Wts/LATS1/2-MOB1/Mats complexes. Wts/LATS then phosphorylate Yki/YAP/TAZ, resulting in their cytoplasmic retention and degradation. When the Hippo pathway is deactivated, Yki/YAP/TAZ shuttle to the nucleus and bind to the TEF family transcription factors Scalloped (SD)/Transcriptional Enhanced Associated Domain (TEAD) to initiate transcription that leads to growth, proliferation, and migration. A paramount of studies have since established the tumor suppressor roles for Hippo/MST1/2, as well as setting the Hippo pathway at

* Corresponding authors.

E-mail addresses: gongqingqiu@sjtu.edu.cn (Q. Gong), xuqin523@sjtu.edu.cn (Q. Xu).

¹ These authors contributed equally to this work.

the center stage of organ size control and regeneration [9]. Other MAP4Ks, such as the Ste20 kinase Happyhour (Hppy) and Mishapen (Msn) of *Drosophila* and their mammalian counterparts MAP4K1/2/3/5 and MAP4K4/6/7, have been characterized as alternative kinases that can phosphorylate Wts/Lats as Hpo/Lats do [10–13].

In 2015, we firstly reported that the model plant *Arabidopsis thaliana* also has a Hippo orthologue [14]. We showed that the *Arabidopsis* Serine/Threonine Kinase 1 (SIK1) was homologous to Hippo and MST1/2 in its kinase domain, and that SIK1 could fully complement the bud site selection and cell cycle exit defects in *ste20Δ* yeast cells. Through a yeast two-hybrid screen, we discovered MOB1A/B, homologues of MOB1/Mats, as SIK1-interacting partners. Both *sik1* and *mob1* null mutants were small in size and exhibited extended longevity. The *sik1* mutant was slower in exit from cell proliferation and entry into cell expansion coupled with endoreduplication, hence it has smaller, rather than larger leaves and petals with reduced cell numbers and smaller cells. Recently we reported that, similar to its metazoan and fungi counterparts, SIK1 also has roles in the polarity establishment of *Arabidopsis*, such as phyllotaxis and root gravitropism [15]. Other studies revealed a positive role for SIK1 in extracellular ROS burst and immunity towards pathogenic bacteria [16], and suggested its involvement in jasmonate-induced gene expression [17]. Although the studies above have rekindled an old interest on MAP4K functions in plants [18], the lack of any clearly determined structure of SIK1 limited further investigations on its molecular functions.

SIK1 belongs to the GCK-II subfamily of the STE20-like kinase (SLK) family [19]. It is homologous to STE20 in the budding yeast, Hpo in the fruit fly, and MST1/2 in humans, all belong to MAP4K [20]. Compared with other MAP4Ks, the N-terminal and C-terminal domains of SIK1 are unique [14]. The N-terminal domain of SIK1 was cytotoxic to *E. coli*, hence the full-length kinase could not be expressed and purified [14]. The C-terminal domain of SIK1 also lacks homology to other GCKs with known structures. Luckily, the kinase domain of SIK1 shares high homology with other Ste20 kinases and MAP4Ks [21,22]. The catalytic activity of the human MAP4K is known to be controlled by its phosphorylation states at a site on the “activation” loop (A-loop). According to available crystal structures, the phosphorylation sites on A-loops are Tpo183 for MST1 [23], Tpo180 for MST2 [21], Tpo178 for MST3 [24], Sep602 for PAK5, and Sep560 for PAK6 [25], where Tpo and Sep are abbreviations for phosphorylated Thr and Ser, respectively. On the other side of the kinase domain, a Gly-rich loop (G-loop) is relatively conserved among MSTs, which may change between relatively ordered and disordered conformations, just like a “gate” in front of the catalytic pocket [26]. However, the molecular mechanism of the phosphorylation on the A-loop to control the conformation at the G-loop is still unexplained.

Here we proposed an allosteric pathway between the A-loop and the G-loop based on molecular dynamics simulations on several homology models of the kinase domain of SIK1 at different phosphorylation states. The importance of three key residues on the path, Lys278, Glu295, and Arg370, were further validated by *in vitro* kinase assay. Considering that SIK1 and other STE20-like kinases are highly conserved in the kinase domain, this allosteric mechanism might be shared by other Ste20 kinases, thus providing clues in the activation and regulation mechanisms of these kinases.

2. Methods and materials

2.1. Homology analysis and molecular modeling

The disordered tendency of full-length SIK1 was predicted by IUPred [27], which was then confirmed by structure prediction

using RaptorX [28] and AlphaFold2 [29,30]. Since the N-terminal and C-terminal domains were predicted to be disordered, only the kinase domain of SIK1 was used as a target for homology modeling by SWISS-MODEL [31–35]. To search for proper templates, we used the sequence of aa249–503 that was predicted to be the kinase domain of SIK1 [14] to do the sequence alignment, with basic local alignment search for proteins (blastp) targeting on the protein database (PDB). We used the Clustal Omega [36] for multiple sequence alignment and the result was visualized and analyzed by Jalview [37]. Five proteins were selected as candidate templates: MST1 (PDB id: 3COM) and MST3 (PDB id: 3A7J) from the GCK sub-family of the STE20 kinases, PAK5 (PDB id: 2F57) and PAK6 (PDB id: 2C30) from the PAK sub-family of the STE20 kinases, and STK25 (also a GCK) without phosphorylated Thr or Ser (PDB id: 4NZW).

Based on multi sequence alignment and structure comparison results, two different templates were chosen to build the structures in different active states and different phosphorylation states. We chose MST1/2 (PDB id: 3COM.A) [21] as the template of the active SIK1 with Tpo405. 3COM.A has the lowest E-value among all blastp results and is the only template with a phosphorylated Tpo183. Then we chose STK25 (PDB id: 4NZW.B) [38] as the template of the inactive SIK1 with non-phosphorylated threonine (Thr405). 4NZW.B has the fourth-lowest E-value among all blastp results and has a “closed” catalytic site with ATP blocked by the G-loop. Thus we considered 3COM.A and 4NZW.B appropriate templates for modeling the activated and inactivated structures of SIK1, respectively. The sequence identity between SIK1 and the two templates are 42.9% and 47.8%, respectively, in the kinase domain.

The ATP-binding model was built by superposition of another Ste20 like kinase, Hematopoietic progenitor kinase-1 (HPK1, PDB id: 6CQD.A) [39] that contains ANP-Mg²⁺-Mg²⁺ in its catalytic pocket, with the active model of SIK1 kinase. And the N atom of ANP was modified to O atom, so as to transfer ANP into ATP. Because of the high similarity between the catalytic pockets of these two structures, the binding pose and shape of ATP-Mg²⁺-Mg²⁺ in this artificial model is more reasonable than directly docking ATP into the pocket.

2.2. Molecular dynamic simulation and data analysis

Four initial structures were built from the homology models above and modified at the residue 405 with different phosphorylation states if necessary (Table 1). All trajectories were calculated by Amber 14 package [40] with the ff03CMAP force field [41], a balanced force field for folded/intrinsically-disordered proteins and the TIP4P-Ew model for water molecules [42]. In those protein systems containing Tpo, we added parameters from Forcefield_PTM [43,44]. Counterions (Na⁺/Cl⁻) were added to neutralize the systems, which were then solvated in a truncated octahedron box of TIP4P-Ew water molecules with a buffer of 12 Å. All systems were relaxed for 6000 steps with the steepest descent minimization and 3000 steps with the conjugate gradient minimization, and were then heated for 20 ps and equilibrated for 20 ps in the constant-pressure and constant-temperature (NPT) ensemble at a temperature of 298.15 K under standard atmospheric pressure. The covalent bonds with hydrogen atoms were constrained using the SHAKE algorithm [45], and the electrostatic interactions were calculated using the particle mesh Ewald (PME) method [46] with a cutoff of 8 Å. Lennard-Jones interactions were truncated at 8 Å. The temperature coupling was controlled using a Berendsen thermostat, and pressure coupling was controlled using a Berendsen barostat [47]. To accelerate the molecular dynamic (MD) simulations, the CUDA version of PMEMD was used [48]. Each initial structure was simulated three times with different random seeds,

Table 1
The models of MD simulations.

Model	Template	Phosphorylation states	Residue at aa405	Initial Pocket	Equilibrated pocket
I	3COM.A	Phosphorylated	Tpo	Open	Open
II	3COM.A	Non-phosphorylated	Thr	Open	Close
III	4NZW.B	Phosphorylated	Tpo	Close	Open
IV	4NZW.B	Non-phosphorylated	Thr	Close	Close

The model based on 3COM.A with Tpo405 (model **I**) and the model based on 4NZW.B with Thr405 (model **IV**) are kinases in the active state or the inactive state, respectively. Then Tpo405 of model **I** was modified into Thr (model **II**) to observe the conformational changes when the dephosphorylation transfers SIK1 into the inactive state. *Vice versa*, Thr405 of model **IV** was phosphorylated in model **III**, so as to observe possible activation by this phosphorylation.

with each trajectory simulated for 500 ns. CPPTRAJ [49] was used to process all MD trajectories and calculate root mean square deviation (RMSD), two-dimensional RMSD, root mean square fluctuation (RMSF), distances, and dihedral parameters. Conformer clustering was performed with Kclust [50], using C α RMSD as reference with the cut-off as 2.5 Å.

2.3. Dynamics network and allosteric conduct pathway analysis

Dynamical fluctuation correlation network is constructed from the atomic fluctuation covariance matrix, as Eq. 1 [51–54]:

$$C_{xy} = \frac{\Delta \vec{r}_x(t) \cdot \Delta \vec{r}_y(t)}{\sqrt{\Delta \vec{r}_x(t)^2 \cdot \Delta \vec{r}_y(t)^2}} \quad (1)$$

where displacement of node x is computed respect to its time average as Eq. 2:

$$\Delta \vec{r}_x(t) = \vec{r}_x(t) - \langle \vec{r}_x(t) \rangle \quad (2)$$

where $\langle \cdot \rangle$ is the time average of node x. Here every amino acid is defined as one node. An edge between two nodes is defined if they are not covalently bonded but are with heavy atoms closer than 4.5 Å over 75% of sampling time. The strength of the edge is defined as the absolute value of the covariance matrix element (C_{xy}) as computed in Eq (1). The number of connected edges at each node is defined as the degree of the node. The relative correlation-weighted degree of the node, which indicates the importance of the node, is defined as the ratio of the total strength of all edges connected to the node over the total strengths of all nodes. It was used to calculate the dynamic cross-correlation matrix (DCCM) and was used for the community analysis. The DCCM analysis unveiled the relationships between residues and distinct regions by quantifying their relative motion. Positive values indicate that two residues are moving in the same direction, whereas negative values indicate the opposite.

The Floyd-Warshall algorithm [55] was used to calculate the shortest pathways between each pair of nodes in the networks. By calculating all shortest paths between each pair of residues, the times that a single edge was passed was defined as the edge betweenness. It can be used to represent the importance of each edge in information flow. Based on the edge betweenness, the networks were clustered into communities with the Girvan-Newman algorithm [56]. All scripts used for calculation are from earlier works [57,58].

2.4. Recombinant protein purification and in vitro kinase assay

Since the N-terminal region and thus the full-length SIK1 is toxic to *E. coli*, we used the N-terminal truncated SIK1, SIK1 Δ N with amino acids 221–836 for *in vitro* kinase assay. SIK1 Δ N^{K278A}, SIK1 Δ N^{E295A}, SIK1 Δ N^{H369A}, SIK1 Δ N^{R370A}, and SIK1 Δ N^{K373A} variants were generated by site-directed mutagenesis. Genes encoding SIK1 Δ N and the 5 variants were sub-cloned into PET30 vector with N-terminal His-MBP tag. All primers used are listed in Table S2.

The plasmids were transformed into *E. coli* for protein expression. Protein expression were induced by addition of 0.25 mM IPTG to a 100 mL culture at OD600 = 1.2, followed by incubation at 16 °C for 13 h. In general, 16 °C is preferred over 37 °C for proper folding of recombinant protein in *E. coli* [59,60]. Bacteria were then harvested and resuspended in binding buffer (50 mM Tris, 0.5 M NaCl, pH8.0). Subsequently, cell lysates were centrifuged at 15,000 rpm and proteins in supernatant were incubated with Ni-NTA Beads at 4 °C for 4 h. Proteins were eluted with 1 mL elution buffer (50 mM Tris, 0.5 M NaCl, 0.5 mM Imidazole, pH8.0). 200 ng of purified protein were added to 30 μ L reaction buffer containing 2 mM HEPES, 30 mM NaCl, 2 mM MgCl₂, and 0.5 mM ATP γ S. The reaction was carried out at 30 °C for 1 h. Then 1.5 μ L of 50 mM PNBM was added and alkylation was proceeded at 30 °C for 1 h. When the reaction is completed, each sample was mixed with sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer and analyzed by western blotting.

For western blotting, Mouse anti-MBP (1:5000 dilution; Uti-Body, China), Rabbit anti-Thiophosphate ester (1:5000; Abcam, USA), and the appropriate IgG-HRP conjugated secondary antibody (1:5000; ZSGB-Bio, China) were used. The signal was developed using High-sensitivity ECL chemiluminescence detection kit (Vazyme, China) and chemiluminescence was detected using a chemiluminescent Western Blot scanner (ChemiScope 6100T, Clinx, China). Quantifications were performed with ImageJ (<https://imagej.nih.gov/>) and normalized to the anti-MBP signal.

3. Results and discussions

3.1. Homology analysis and molecular modeling

For the full-length SIK1, the disordered tendency predicted by IUpred and the three-dimensional structures predicted by RaptorX and AlphaFold2 all indicate that only the residues numbered from 249 to 503 tend to form an ordered structure, which may be exactly the kinase domain, while most of the N-terminal and C-terminal domains are highly disordered (Fig. S1). Therefore, the MD simulations and structural analyses here mainly focus on the kinase domain of SIK1.

We selected five STE20 kinase family members with reported crystal structures for a comparison with SIK1 of their phosphorylation sites and phosphorylation states (Fig. 1A). The overall structures are well conserved among all 5 STE20 kinases, suggesting a similar catalytic regulatory mechanism (Fig. 1B-C). Although the phosphorylation sites corresponding to Thr405 in SIK1 are not identical, they are all serines or threonines. As far as we know, it is always threonine in all MSTs. Another difference on the phosphorylation site is that, in all the crystal structures compared, only Thr178 of 4NZW is in the non-phosphorylated state and deviates from those phosphorylated sites of the other four structures. The detailed multi-sequence alignment is shown in Fig. S2. The alignment showed that only the kinase domains are relatively conserved among the MAP4Ks, particularly around the catalytic sites Asp371 and Asp389, the ATP binding site Lys278, the phosphoryla-

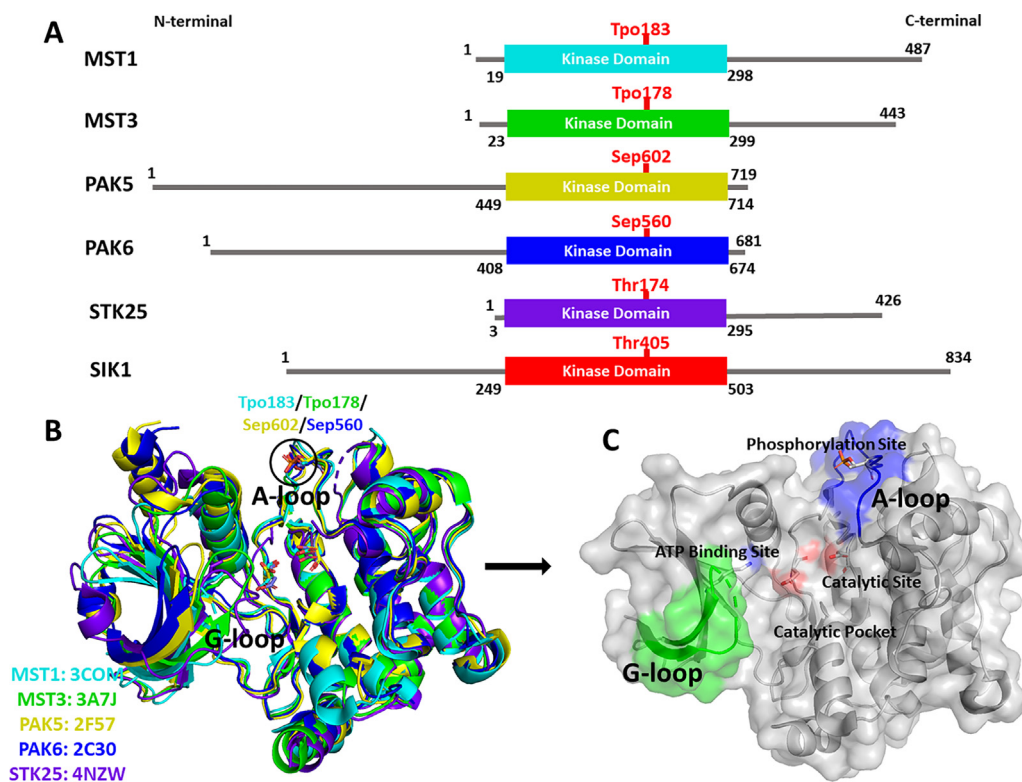


Fig. 1. Comparison between STE20 like kinases. (A) Only the kinase domains are conserved among MST1, MST3, PAK5, PAK6, STK25 and SIK1. According to multi-sequence alignment, STE20 like kinases have conserved phosphorylation sites (some are phosphothreonines while some are phosphoserine). It is reasonable to suppose that Thr405 is a phosphorylation site of SIK1. The detailed multi-sequence alignment result is shown in Fig. S2. (B) The “activation” loops (A-loop, aa398–408) and Gly-rich loops (G-loop, aa253–264) are conserved among these five homology proteins[26]. (C) A surface model of MST1 (3COM.A), with the A-loop in blue, the G-loop in green, the catalytic site in red. The residues of the phosphorylation site, the catalytic pocket and the ATP binding site shown in sticks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tion site Thr405 on the A-loop [21], and the Gly-rich loop region (G-loop) including Gly256, Gly258 and Gly261 of SIK1. By comparing a phosphorylated structure 3COM and the non-phosphorylated structure 4NZW (Fig. S3), we noticed that the structure of the G-loop may be a critical factor in determining the catalytic activation of STE20 kinases. In the non-phosphorylated structure (4NZW.B), the G-loop is unfolded at the entry of the catalytic pocket, blocking ATP from entering. In contrast, in a phosphorylated kinase (3COM.A), the G-loop forms a β -hairpin structure, leaving the catalytic pocket open, so that ATP can quickly enter the catalytic site.

Using 3COM.A and 4NZW.B as templates, we built two three-dimensional models for the kinase domain of SIK1. When comparing the two models, we noticed that the orientation of the residue Lys257 could best reflect the conformational change in the G-loop and may play a role in the regulation of the catalysis of SIK1. The model based on 3COM (in green) has the phosphorylated residue Tpo405 (Fig. 2). At the same time, its G-loop forms a β -hairpin structure that is somewhat stabilized by the salt bridge between the side chains of Lys257 and Glu254. In this model, the entry of the catalytic site is fully open for ligand binding, thus it is called “the active model” hereafter. On the opposite, “the inactive model”, shown in blue, has no phosphate group on Thr405 and the entry of its catalytic pocket is blocked by the unfolded G-loop, where the sidechain of Lys257 sticking out toward Asp334 to form a salt bridge. This model based on 4NZW may represent a non-phosphorylated and inactive state of SIK1. Similar to the phosphorylation site, the site of Lys257 of SIK1 is not fully conserved among MAP4Ks, but they are all charged residues with long side chains (Lys in SIK1, STK25, and MST3; Glu in MST1, PAK5, and PAK6) (Fig. S2). The flexibility of the A-loop and the G-loop in the 3D space might be the reason to allow the different amino acids at

the phosphorylation site and the regulation site to achieve similar function, respectively. For example, no matter the phosphorylated residue is serine or threonine, its phosphoryl group always forms strong double hydrogen bond with a nearby arginine, which corresponds to Tpo405 and Arg370 in SIK1 (to be discussed below). At the same time, the salt bridge between Lys257 and Glu254 stabilizing the G-loop in the active SIK1 is also observed in the structure of PAK6, where Glu415 and Lys412 forms a salt bridge in the inversed orientation (see Fig. S4). At the same time, the residue Asp334 of SIK1 on the opposite side of the entry is fully conserved in all these kinases. It may provide electrostatic interactions with other positively charged residue on the G-loop once it is relatively unfolded, such as the partially conserved lysines three sites ahead of Lys257. In summary, although not strictly conserved as the catalytic residues in the active site, Thr405 and Lys257 of SIK1 on the flexible A-loop and G-loop may still exhibit some regulation function common in the STE20-like kinases. Therefore, we chose Thr/Tpo405 and Lys257 as the terminal residues for further allosteric pathway analysis, as discussed below.

From these two models, we modified the phosphorylation states at the residue 405 and obtained four initial structures in total for further MD simulations, as summarized in Table 1, so as to compare and observe how the phosphorylation states can affect the dynamic conformations of the SIK1 kinase domain and finally affect its activity.

3.2. MD simulations

We ran 500 ns simulations for each model, which is relatively enough for conformational equilibration in model I and model IV, and enough to observe the conformational changes in model II

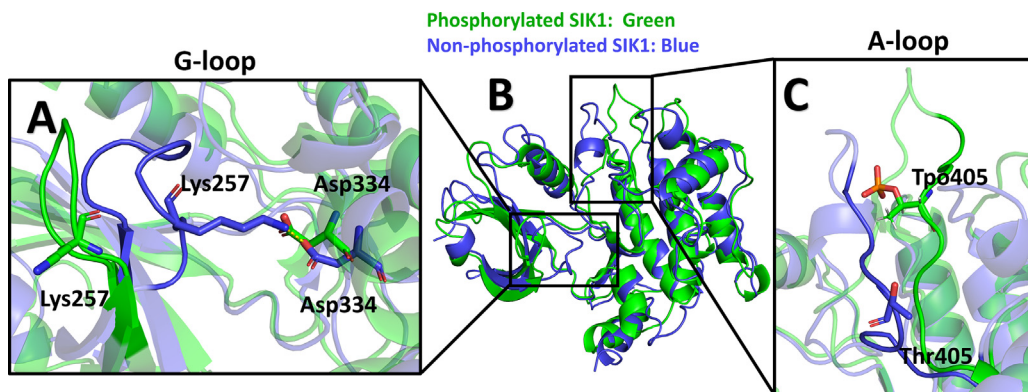


Fig. 2. The structure comparison between two SIK1 kinase homology models with different phosphorylation states. The blue one is based on 4NZW while the green one is based on 3COM. The carbon atoms of the side chains are colored same as the backbone while other heavy atoms colored by elements. (A) The difference in the orientation of the G-loop and the long side-chain of residue Lys257. (B) The global comparison between the two models. (C) The difference in the A-loop and the phosphorylation site Thr405/Tpo405. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and model **III**. As suggested by the RMSD results (Fig. S5), all four systems reached equilibrium. Consistent with our hypothesis, the model **I** and **IV** were stable and keep their pockets open or closed, respectively. In contrast, with the phosphorylation states inverted, the model **II** and **III** experienced important conformational changes, especially in the G-loop region, making the catalytic pocket closed or open, respectively.

Whether the pocket is open or closed is reflected by the different distributions of two selected distances. The first distance is the distance between the side-chain N atom of Lys257 on the G-loop and the gamma carbon atom of Asp371, one of the catalytic residues on the other side of the bottom of the catalytic pocket. The other distance we chose is the distance between the side-chain N atom of Lys257 and the gamma carbon atom of Asp334, a residue on the edge of the other side of the catalytic pocket at the entry. The two distances and their distributions are shown in Fig. 3. Although the results of the trajectories from different initial homology models (**I** & **II** versus **III** & **IV**) may not be compared directly, comparison between the distributions in the models based on the same homology model (**I** versus **II**, and **III** versus **IV**) could still suggest a shift of opening probability when the phosphorylation state is changed on the A-loop, which may also mean different chance for ATP to enter the catalytic pocket and thus change the catalytic activity.

As Fig. 3 suggested, the active model from 3COM.A with phosphorylation at Tpo405 (model **I**) tend to keep the catalytic pocket open while the one with non-phosphorylated Thr405 (model **II**) may close its catalytic pocket and transfer into the inactive state during the 500 ns simulations. The different conformations are compared by representative structures from the clustering results of MD simulations (Fig. 4). Oppositely, the inactive model from 4NZW.B tends to have the catalytic pocket open or keep it closed when the residue 405 is phosphorylated (model **III**) or dephosphorylated (model **IV**), respectively (Fig. 5). These two figures together show that the phosphorylation of the A-loop is the critical factor that determines the conformation of the G-loop and the orientation of the side chain of the key residue Lys257, which determines the opening probability of the catalytic pocket and thus the activity of SIK1. The RMSD and RMSF analysis of the trajectories are also shown in Fig. S5 and Fig. S6, which are consistent with the conclusions from the representative structures above.

3.3. Community analysis

From the MD simulations, we found enormous dynamic motions in the domain of the G-loop to open or close the catalytic

pocket upon phosphorylation or de-phosphorylation on the A-loop, respectively. To search for important domains coupled with these conformational changes, dynamic cross-correlation matrix (DCCM) were constructed from the trajectories stable at the active state (model **I**) and the inactive state (model **IV**) (Fig. S7), from which the residual DCCM was calculated for the active SIK1 kinase domain versus the inactive SIK1 kinase domain (Fig. 6). According to the residual DCCM map, the correlations between the A-loop and α G, the A-loop and the catalytic sites, as well as the G-loop and α C increased in the phosphorylated SIK1 kinase domain, where the latter two changes may be related to the allosteric regulation of the activity of SIK1 as discussed in section 3.2 above.

From these matrices, we generated the community networks of the stable models **I** and **IV**, i.e. the active SIK1 (phosphorylated) and inactive SIK1 (non-phosphorylated) for comparison (Table S1). Residues within the cut-off distance of 4.5 Å for at least 75% of the MD trajectories were categorized into the same community, which could be regarded as a synergistic functional unit within the protein structure. In Fig. 7, residues within the same community were illustrated with identical colors, and further simplified to colored circles connected by sticks, whose thicknesses are proportional to the maximum edge betweenness of the two connected communities. In both networks of SIK1 kinase domain, the community A, B, and C mainly consist of the β -sheet domain together with the G-loop, the bottom of the catalytic pocket, and part of the A-loop, respectively. The community I mainly cover the small helix α C that consists one side of the catalytic pocket together with the communities A and G, while the community B and D consist of the other side of the pocket and also part of the α -bundle domain. But the core of the α bundle domain is mainly supported by the helix α G, which is part of the community J in both networks.

Obviously, the community networks are quite different in the active and the inactive states. There are ten communities in the former but only nine in the latter, mainly because the more flexible α -bundle domain in the active state is divided into smaller communities, and the connections with the core community J are all weakened. Instead of associated with J, the community C of phosphorylated SIK1 is directly connected with community B, consistent with the stronger interaction between the A-loop and the catalytic pocket as suggested in the residual DCCM. Another obvious difference is around the β -sheet domain. The stronger connections between the communities A, B, G and I all suggest more compact packing around the β -sheet domain, including the formation of a β -hairpin at the G-loop. At the same time, the thickness of the connection between the communities A and D is lowered, possibly because the pocket entry is more open and the folded G-loop loses

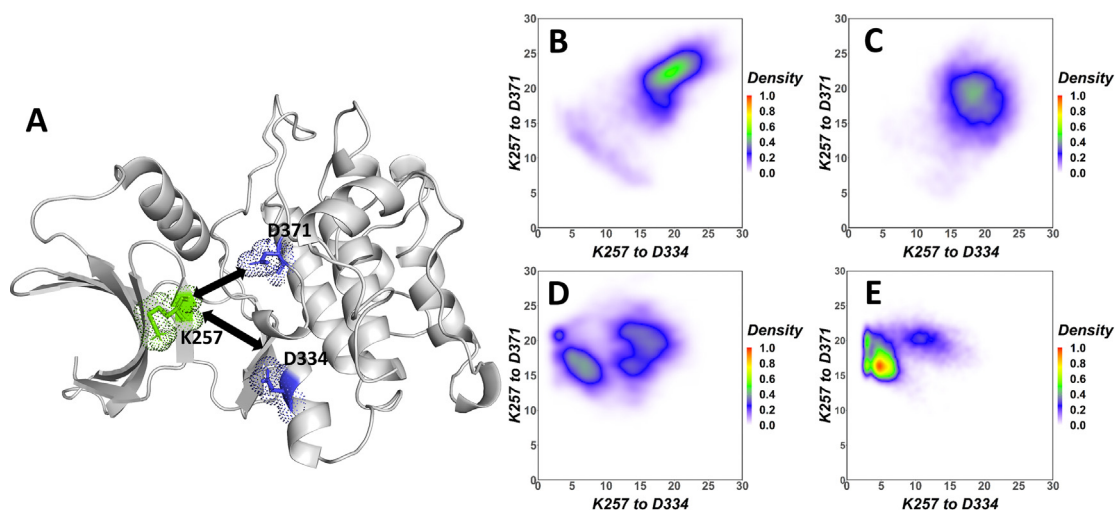


Fig. 3. The opening of the catalytic pocket in the MD trajectories of the four models. The two distances monitored are illustrated in (A), whose distributions are summarized in (B-E) with the units in angstrom. (B) The initially active and phosphorylated model keeps open (model I). The two distances are relatively concentrated in the larger region of 25 Å, indicating more space for ATP to enter. (C) The initially active but non-phosphorylated model turns to be relatively dispersed around 20 Å, suggesting the pocket turn to be more closed (model II). (D) The initially closed but phosphorylated model turns more open (model III). Although part of the frames still have the distance K257 to D334 around 5 Å, there are a number of samples with much longer distances, compared with the samples in (E) the initially closed and non-phosphorylated model. In the trajectory of the latter, the distance from K257 to D334 is highly concentrated at 5 Å, indicating that the pocket keeps closed (model IV).

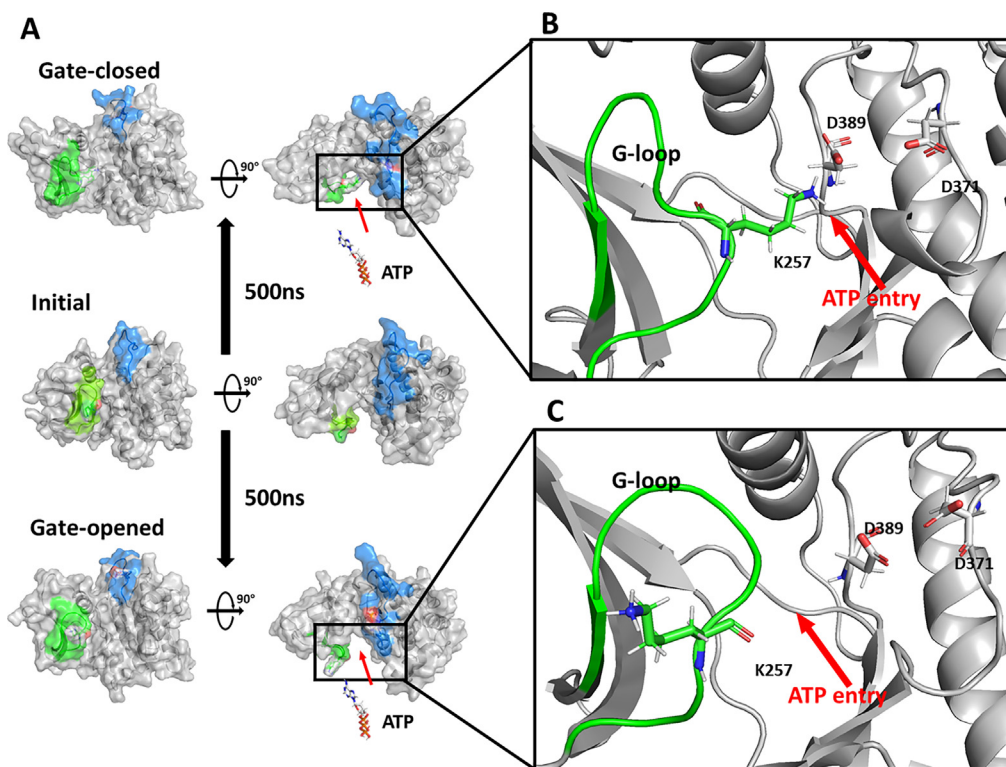


Fig. 4. The MD simulations and clustering results of the initially active model. (A) The initial structure with open catalytic pocket (middle) turns closed in the non-phosphorylated model (model II, top) and keeps open in the phosphorylated model (model I, bottom). The representative structures were obtained by k-means clustering from the last 300 ns of the 500 ns trajectories, with an RMSD cut-off of 2.5 Å. (B) The structure details of the non-phosphorylated model. The pocket is partially occupied by the G-loop in this structure, representing its inactivity. The Lys257 lies horizontally at the entry of the catalytic pocket, leading to the steric effects preventing ATP from entering the active site. This cluster accounts for 71.9% of the selected part of trajectory. (C) The structure details of the phosphorylated model. The pocket is obviously open in this structure, representing its activity. The Lys257 lies vertically in front of the catalytic pocket, allowing the ATP to enter. Due to the relatively big cut-off of the clustering (2.5 Å), all the frames were clustered into one cluster, which still indicates that the structure of the G-loop is more stable in phosphorylated SIK1 at the same clustering cut-off.

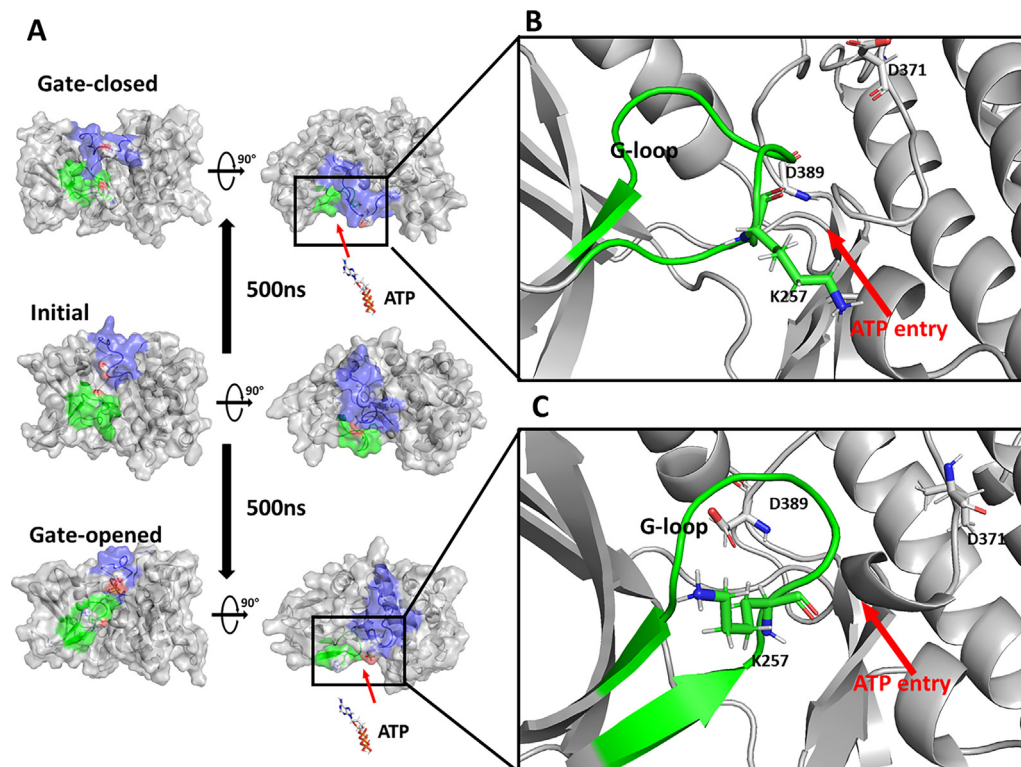


Fig. 5. The MD simulations and clustering results of the initially inactive model. (A) The initial structure with closed catalytic pocket (middle) keeps closed in the non-phosphorylated model (model IV, top) and turns open in the phosphorylated model (model III, bottom). The representative structures are the most populated clusters obtained by k-means clustering from the last 300 ns of the 500 ns trajectory, with an RMSD cut-off of 2.5 Å. (B) The structure details of the top cluster that accounts for 23.9% of the last 300 ns trajectories of the non-phosphorylated model. The pocket is blocked by the G-loop in this structure, suggesting its inactivity. The Lys257 lies horizontally at the entry of the catalytic pocket, leading to the steric effects preventing the ATP from entering. (C) The structure details of the top cluster that accounts for 61.9% of the last 300 ns trajectories of the non-phosphorylated model. The pocket is open in this structure. The Lys257 lies vertically at the entry of the catalytic pocket, allowing the ATP to enter.

its possible interaction with Asp334 in the non-phosphorylated SIK1.

3.4. Allosteric pathways

The shortest path algorithm, Floyd-Warshall's Algorithm was used to calculate the shortest paths (SP) communicating the A-loop and the G-loop, choosing Thr405/Tpo405 (the phosphorylation site on the A-loop) as the beginning residue and Lys257 (at the entry of the catalytic pocket) as the ending residue. The shortest correlation paths of phosphorylated kinase and non-phosphorylated kinase are compared in Fig. 8.

The first difference between the two shortest paths is at the upstream between the phosphorylation site and Arg370. The path of the phosphorylated kinase lacks Ile407 and Tyr409 between Tpo405 and Arg370, because the phosphate on Tpo405 can directly interact with the guanidine group of Arg370. Oppositely, Thr405 in the non-phosphorylated kinase has almost no interaction with Arg370, as observed in the MD simulations and the DCCM map. The second divergence between the two paths is between Arg370 and Lys278. Lys278 is the ATP binding site of SIK1, which corresponds to Lys59 in MST1 [21]. In the phosphorylated kinase, the communication is through the interaction between the guanidine of Arg370 and the backbone carbonyl of Val390, the backbone amino group of Phe390 and the carboxyl of Glu295, as well as the salt bridge between the side chains of Glu295 and Lys278. On the other side, the non-phosphorylated kinase replaces the latter two interactions with only the salt bridge between Asp389 and Lys278. One of the beneficial effects of the bypass through Glu295 is to release the side chain of Asp389, which is one of the critical

catalytic residues. At the downstream of both two paths, on the β -sheet domain Lys278 interacts with Val263 through backbone H-bonds and then influence Leu255 and Lys257 on the G-loop (Fig. 8).

In summary, the shortest path in the active SIK1 kinase suggested the side chains of Lys278, Glu295 and Arg370 play important roles in the communication from the phosphorylation site Tpo405 on the A-loop to the G-loop at the entry of the catalytic pocket. Interestingly, it is noticeable that all these three residues are totally conserved, even more conserved than the phosphorylation site 405 and the G-loop in the multi-sequence alignment among all compared MAP4Ks (Fig. S2). Among the three residues, Lys278 has been long recognized as critical to ATP binding [21], whereas the roles of Glu295 and Arg370 are often neglected before. Although it was mentioned that the salt bridge between Lys278 and Glu295 can only be found in the crystal structures of phosphorylated MAP4Ks [26] but not in the non-phosphorylated ones, its role to the catalysis had not been explained. Here, the scenario of the two allosteric pathways changed from the inactive to the active SIK1 kinase may perfectly explain the critical function of the fully conserved Glu295 and Arg370 in activation of STE20-like MAP4Ks.

3.5. In vitro kinase assay

To gain more insights of the distinct functions of the key residues in the allosteric path, especially the binding of ATP in the active state, we built a ATP-binding model of the kinase domain of SIK1 with ATP superposed into the catalytic pocket, and the key residues around the catalytic site Asp371 and Asp389 were examined (Fig. 9A). In this model, both Lys278 and Glu295 are irre-

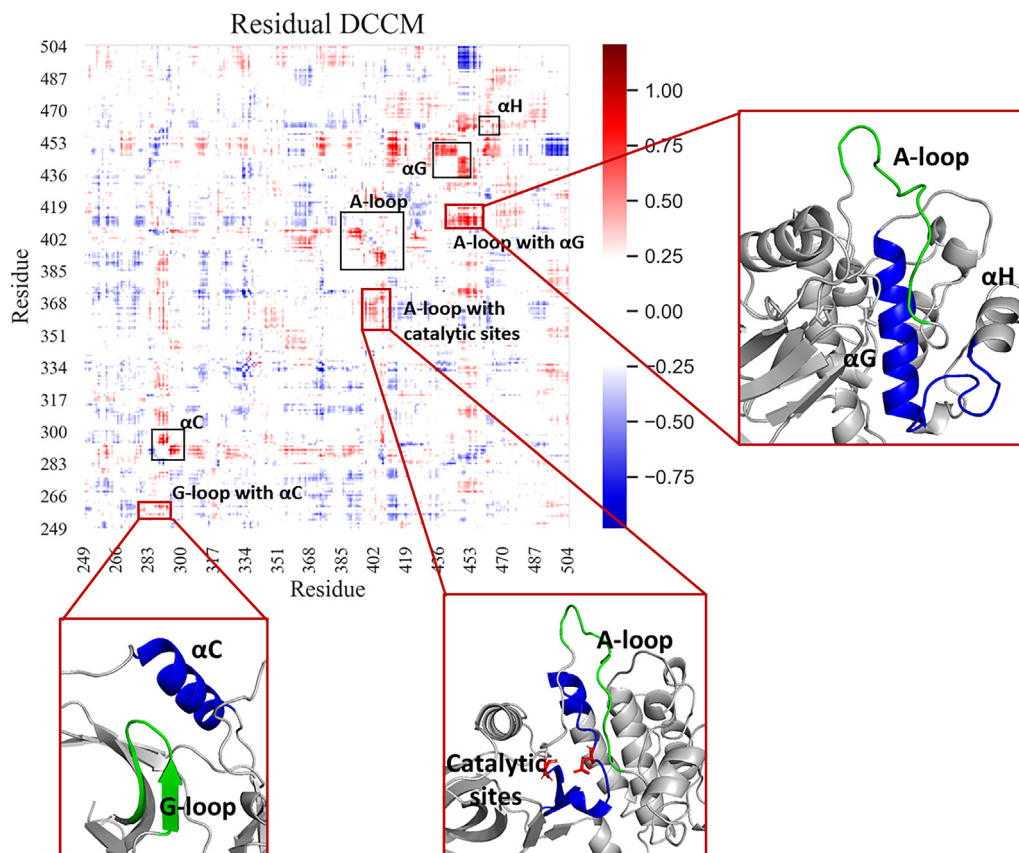


Fig. 6. The residual DCCM matrix of the active SIK1 kinase domain versus the inactive SIK1 kinase domain. The collaborations between the G-loop and αC and between the A-loop and the catalytic sites are enhanced, which may be directly related to the regulation of the catalytic activity. Meanwhile, the correlation between the A-loop and αG increased as well.

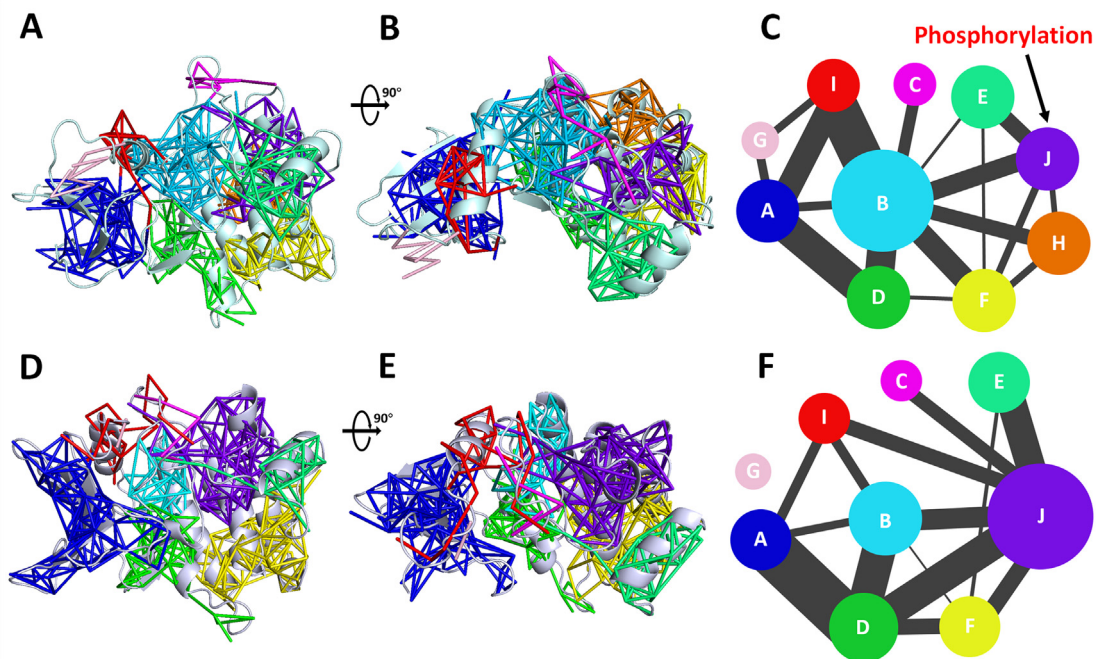


Fig. 7. Map of community network of (A-C) the active and (D-F) the inactive SIK1 kinase domain. Each community was represented by a circle in different colors, with its radius proportional to the number of residues in it, which are listed in Table S1. Inter-community connection was visualized by stick, the thickness of which was proportional to the maximum edge betweenness of the two connected communities.

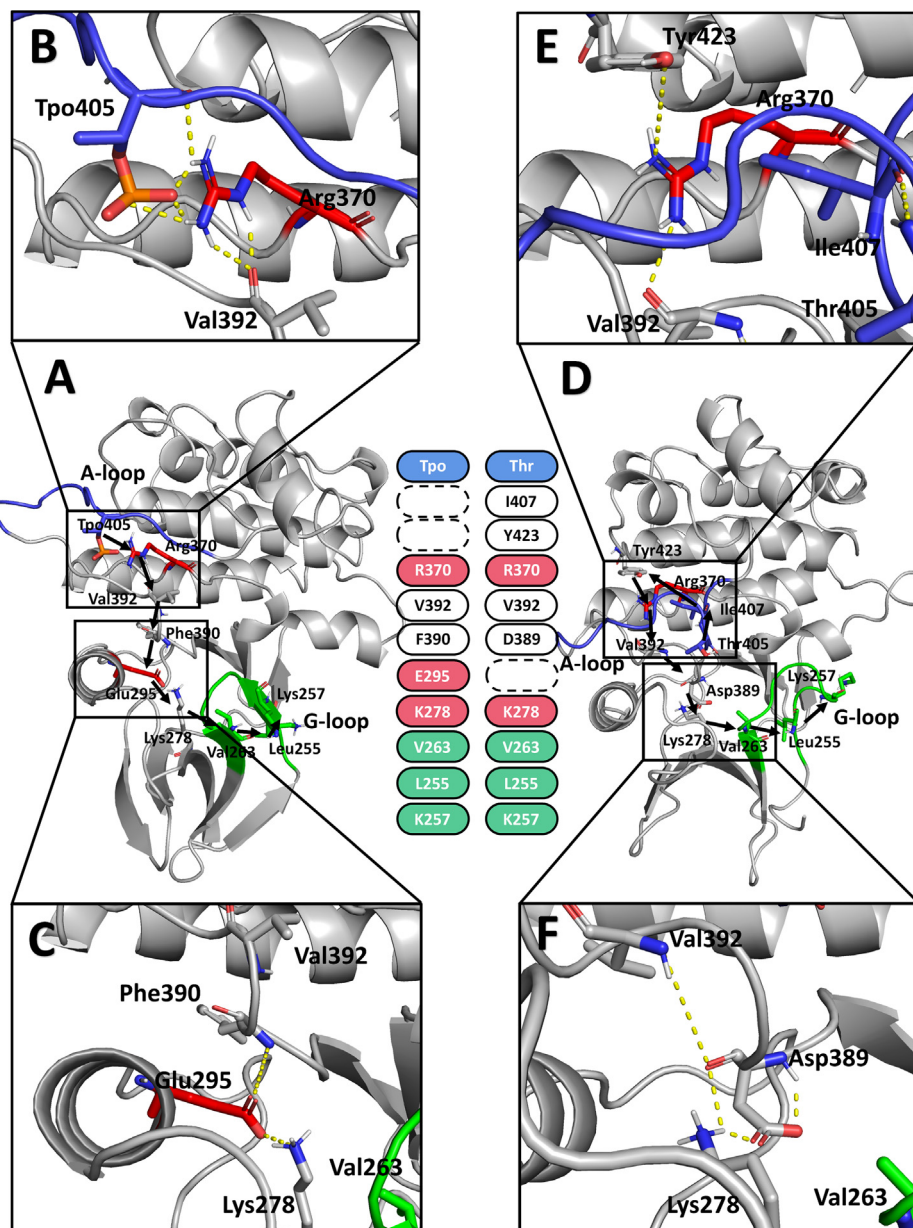


Fig. 8. Comparison between the two shortest correlation paths in SIK1 kinase domain in the different phosphorylation states. The structures of the phosphorylated (A) and non-phosphorylated (D) kinase domains are also shown beside. The starting residues (Tpo405 and Thr405) are colored in blue, same as the backbone of A-loop in the structures. Similarly, the downstream residues on the G-loop as well as the backbone of the G-loop are colored in green. Two key residues (Arg370 and Glu295) validated by kinase assay are in red. (B) The detailed interactions (in yellow) around one of the critical residues Arg370 (in red) in the allosteric pathway of active SIK1. (C) The detailed interactions (in yellow) around another key residues Glu295 (in red) in the allosteric pathway of active SIK1. (E) The detailed interactions (in yellow) around Arg370 (in red) in the pathway of inactive SIK1. (F) The detailed interactions (in yellow) around the mid-stream of the pathway of inactive SIK1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

placeable to conduct the allosteric effect from the A-loop to the G-loop to activate SIK1, as well as to attract the negative groups of Asp389 and ATP together. The salt bridges and hydrogen bonds from the highly conserved, positively charged residues His369 and Lys373 could be essential in fixing the catalytic residues Asp389 and Asp371 in their active positions. Lastly, the positive guanidine of Arg370 may help connecting the phosphate on Tpo405 to the allosteric path.

To confirm the importance of these key residues, *in vitro* kinase assay was performed. We purified recombinant SIK1 Δ N and its mutants, K278A, E295A, H369A, R370A, and K373A from *E. coli*. The kinase activities were evaluated by analyzing autophosphorylation using an antibody-labeled assay, of which results are shown

in Fig. 9B-C. Clearly, the SIK1 Δ N protein could autophosphorylate itself, while SIK1 Δ N^{K278A}, SIK1 Δ N^{E295A}, SIK1 Δ N^{H369A}, and SIK1 Δ N^{K373A} completely lost the autophosphorylation activity. The autophosphorylation activity of SIK1 Δ N^{R370A} was significantly reduced. The essential function of Lys278, Glu295, His369, and Lys373, and the important function of Arg370 were therefore all validated by the kinase assay.

4. Conclusions

Although many crystal structures of human MAP4Ks have been reported, the details of dynamic changes in conformation remained elusive. In this study, we outlined an allosteric regulation pathway

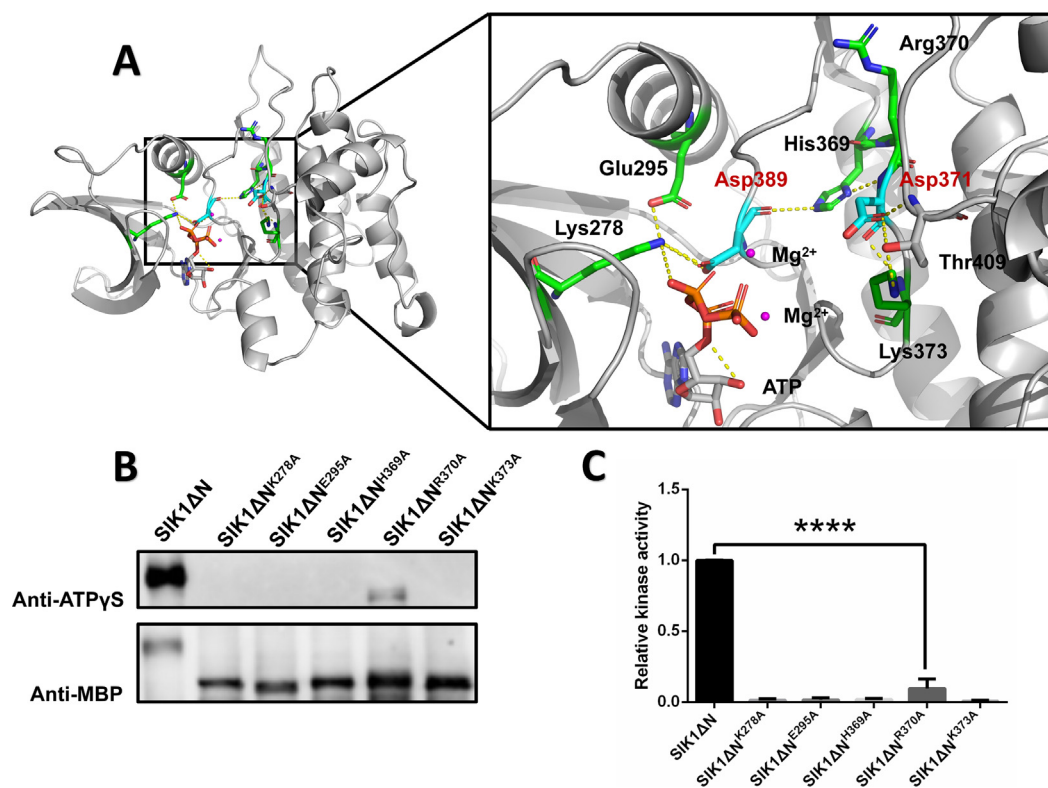


Fig. 9. Kinase activities of SIK1ΔN protein and its mutants. (A) The model with ATP-Mg²⁺-Mg²⁺ superposed in the catalytic pocket of SIK1 kinase domain. The key residues (green) around the catalytic residues (cyan) are highlighted. (B) Recombinant His-MBP-SIK1ΔN and five mutated versions of His-MBP-SIK1ΔN were expressed and purified from *E. coli*. The *in vitro* kinase assay was initiated by adding ATPγS along with ATP to the indicated recombinant proteins. Autophosphorylation was tested by an anti-Thiophosphate ester antibody that specifically recognizes ATPγS (upper panel), while the recombinant proteins were blotted with mouse monoclonal antibody to MBP tag (bottom panel). Similar results were obtained from three biological replicates. (C) Quantification of relative kinase activity from the grayscale in (A). The relative grayscale of SIK1ΔN protein in the upper panel versus that in the lower panel was set to 1. The autophosphorylation of SIK1ΔN^{E295A} was not detected whereas the autophosphorylation of SIK1ΔN^{R370A} was greatly reduced. The autophosphorylation of SIK1ΔN^{K278A}, SIK1ΔN^{H369A}, and SIK1ΔN^{K373A} were not detected. **** indicates $p < 0.0001$ in paired Student's *t*-test. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in SIK1 by molecular dynamic simulations and dynamic cross-correlation matrix analyses, which could be summarized into three points:

Firstly, SIK1 is activated by phosphorylation on the A-loop through conformational changes in the G-loop that open the catalytic pocket. From Fig. 3, we can see that the phosphorylation of the A-loop greatly improves the probability of pocket opening, which means that ATP would more likely enter the catalytic pocket and interact with catalytic sites (Asp371 and Asp389). As shown in Figs. 4 and 5, the opening of the pocket upon the phosphorylation mainly comes from the folding of the G-loop at the entry, while in the non-phosphorylation state the G-loop becomes so flexible that Lys257 could block the entry.

Secondly, an allosteric pathway coordinates the phosphorylation site and the entry of the catalytic pocket. Based on the residual DCCM map in Fig. 6, we calculated the network communities in Fig. 7 to illustrate the global structural changes upon phosphorylation, as well as the shortest paths in Fig. 8, which demonstrated the key roles of the residues Lys278, Glu295, and Arg370 in the allosteric regulation pathway from the phosphorylated Tpo405 on the A-loop to Lys257 on the G-loop, which was confirmed by experimental results from the *in vitro* kinase assay in Fig. 9. Unlike the long recognized key residue Lys278 (also validated in the kinase assay), these two residues have not been fully discussed in previous studies, although they are highly conserved in both sequence and structure of all MAP4Ks compared here.

Thirdly, some key residues with relatively conserved function in the 3D structure may not be conserved in the primary sequences, especially when they are on a flexible loop instead of

in the rigid catalytic pocket. Unlike His369 and Lys373 which may interact with the catalytic residues, or Lys278 and Glu295 which may affect ATP binding in the deep catalytic pocket, the terminal residues Thr405 and Lys257 of the allosteric pathway are not fully conserved in all the STE20-family kinases. Similarly, the importance of not fully conserved residues was reported in the activation of coagulation factor X [61]. Therefore, only searching for key residues according to conservations in high throughput sequence alignment may not be enough. To develop more tools involving three-dimensional structural information of the target proteins might be necessary, especially in the post-AlphaFold decade.

Considering that SIK1 and other STE20-family kinases are highly conserved in the kinase domain, the allosteric mechanism described here is likely shared by other STE20 kinases. The key residues reported in this study could serve as potential allosteric drug targets or protein design hotspots not only on SIK1 for plant biomass production and yield formation, but also on other MSTs for organ repair and cancer research.

CRediT authorship contribution statement

Junxi Mu: Investigation, Software, Data curation, Writing – original draft. **Jiali Zhou:** Investigation, Writing – original draft. **Qingqiu Gong:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Funding acquisition, Supervision. **Qin Xu:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Funding acquisition, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Dr. Pingping Zhang for discussions. This work is supported by the National Natural Science Foundation of China (91954102, 31871355, and 31671419) to QG and (31770772) to QX, and Joint Research Funds for Translational Medicine at Shanghai Jiao Tong University (ZH2018ZDA06).

References

- Dan I, Watanabe NM, Kusumi A. The Ste20 group kinases as regulators of MAP kinase cascades. *Trends Cell Biol* 2001;11(5):220–30.
- Colcombet J, Hirt H. Arabidopsis MAPKs: a complex signalling network involved in multiple biological processes. *Biochem J* 2008;413:217–26.
- Bae SJ, Luo X. Activation mechanisms of the Hippo kinase signaling cascade. *Biosci Rep* 2018;38.
- Chen R et al. STRIPAK integrates upstream signals to initiate the Hippo kinase cascade. *Nat Cell Biol* 2019;21(12):1565–77.
- Plouffe SW et al. Characterization of Hippo Pathway Components by Gene Inactivation. *Mol Cell* 2016;64(5):993–1008.
- Wu S et al. hippo encodes a Ste-20 family protein kinase that restricts cell proliferation and promotes apoptosis in conjunction with salvador and warts. *Cell* 2003;114(4):445–56.
- Harvey KF, Pfeleger CM, Hariharan IK. The Drosophila Mst ortholog, hippo, restricts growth and cell proliferation and promotes apoptosis. *Cell* 2003;114(4):457–67.
- Creasy CL, Chernoff J. Cloning and characterization of a human protein kinase with homology to Ste20. *J Biol Chem* 1995;270(37):21695–700.
- Ma S et al. The Hippo pathway: biology and pathophysiology. *Annu Rev Biochem* 2019;88(1):577–604.
- Zheng Y et al. Identification of Happyhour/MAP4K as alternative Hpo/Mst-like kinases in the Hippo kinase cascade. *Dev Cell* 2015;34(6):642–55.
- Meng Z et al. MAP4K family kinases act in parallel to MST1/2 to activate LATS1/2 in the Hippo pathway. *Nature Commun* 2015;6(1). <https://doi.org/10.1038/ncomms9357>.
- Li S et al. Overlapping functions of the MAP4K family kinases Hppy and Msn in Hippo signaling. *Cell Discovery* 2015;1(1). <https://doi.org/10.1038/celldisc.2015.38>.
- Li Q et al. The conserved Misshapen-Warts-Yorkie pathway acts in enteroblasts to regulate intestinal stem cells in drosophila. *Dev Cell* 2014;31(3):291–304.
- Xiong J et al. The Hippo/STE20 homolog SIK1 interacts with MOB1 to regulate cell proliferation and cell expansion in Arabidopsis. *J Exp Bot* 2016;67(5):1461–75.
- Zhang P et al. The Arabidopsis STE20/Hippo kinase SIK1 regulates polarity independently of PIN proteins. *Biochem Biophys Res Commun* 2021;549:21–6.
- Zhang M et al. The MAP4 kinase SIK1 ensures robust extracellular ROS burst and antibacterial immunity in plants. *Cell Host Microbe* 2018;24(3):379–391.e5.
- Guo Z, Yue X, Cui X, Song L, Cheng Y. AtMOB1 genes regulate jasmonate accumulation and plant development. *Plant Physiol* 2020;182(3):1481–93.
- Pan LX, De Smet I. Expanding the mitogen-activated protein kinase (MAPK) universe: an update on MAP4Ks. *Front Plant Sci* 2020;11.
- Karpov PA et al. Bioinformatic search of plant microtubule-and cell cycle related serine-threonine protein kinases. *BMC Genomics* 2010;11(S1). <https://doi.org/10.1186/1471-2164-11-S1-S14>.
- Jonak C et al. Complexity, cross talk and integration of plant MAP kinase signalling. *Curr Opin Plant Biol* 2002;5(5):415–24.
- Galan JA, Avruch J. MST1/MST2 protein kinases: regulation and physiologic roles. *Biochemistry* 2016;55(39):5507–19.
- Nemoto K et al. Autophosphorylation profiling of Arabidopsis protein kinases using the cell-free system. *Phytochemistry* 2011;72(10):1136–44.
- Glantschnig H, Rodan GA, Reszka AA. Mapping of MST1 kinase sites of phosphorylation - Activation and autophosphorylation. *J Biol Chem* 2002;277(45):42987–96.
- Ko T-P et al. Structures of human MST3 kinase in complex with adenine, ADP and Mn2+. *Acta Crystallogr Section D-Biol Crystallogr* 2010;66(2):145–54.
- Eswaran J, Lee WH, Debreczeni JE, Filippakopoulos P, Turnbull A, Fedorov O, et al. Crystal structures of the p21-activated kinases PAK4, PAK5, and PAK6 reveal catalytic domain plasticity of active group IIPAKs. *Structure* 2007;15(2):201–13.
- Record CJ, Chaikuad A, Rellos P, Das S, Pike ACW, Fedorov O, et al. Structural comparison of human mammalian Ste20-Like kinases. *PLoS ONE* 2010;5(8).
- Dosztányi Z et al. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol* 2005;347(4):827–39.
- Wang S et al. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput Biol* 2017;13(1).
- Jumper J et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596(7873):583–9.
- Tunyasuvunakool K et al. Highly accurate protein structure prediction for the human proteome. *Nature* 2021;596(7873):590–6.
- Waterhouse A, et al., SWISS-MODEL: homology modelling of protein structures and complexes. *Nucl Acids Res*, 2018. 46(W1): p. W296–W303.
- Bienert S et al. The SWISS-MODEL Repository-new features and functionality. *Nucl Acids Res* 2017;45(D1):D313–9.
- Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* 2009;30(S1):S162–73.
- Studer G, et al., QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics*, 2020. 36(6): p. 1765–1771.
- Bertoni M et al. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep* 2017;7(1). <https://doi.org/10.1038/s41598-017-09654-8>.
- Madeira F, et al., The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucl Acids Res*, 2019. 47(W1): p. W636–W641.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25(9):1189–91.
- Hao Q et al. Structural insights into regulatory mechanisms of MO25-mediated kinase activation. *J Struct Biol* 2014;186(2):224–33.
- Wu P et al. Hematopoietic Progenitor Kinase-1 Structure in a Domain-Swapped Dimer. *Structure* 2019;27(1):125–133.e4.
- Case DA et al. The Amber biomolecular simulation programs. *J Comput Chem* 2005;26(16):1668–88.
- Zhang Y et al. Well-balanced force field ff03CMAP for folded and disordered proteins. *J Chem Theory Comput* 2019;15(12):6769–80.
- Horn HW et al. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* 2004;120(20):9665–78.
- Khoury GA et al. Forcefield_NCAA: ab initio charge parameters to aid in the discovery and design of therapeutic proteins and peptides with unnatural amino acids and their application to complement inhibitors of the compstatin family. *ACS Synth Biol* 2014;3(12):855–69.
- Khoury GA et al. Forcefield_PTM: ab initio charge and AMBER Forcefield parameters for frequently occurring post-translational modifications. *J Chem Theory Comput* 2013;9(12):5653–74.
- Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 1977;23(3):327–41.
- Essmann U et al. A smooth particle mesh Ewald method. *J Chem Phys* 1995;103(19):8577–93.
- Berendsen HJC et al. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81(8):3684–90.
- Götz AW et al. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J Chem Theory Comput* 2012;8(5):1542–55.
- Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* 2013;9(7):3084–95.
- Feig M, Karanicolas J, Brooks CL. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graph Model* 2004;22(5):377–95.
- Tai K et al. Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. *Biophys J* 2001;81(2):715–24.
- Young MA et al. Dynamic coupling between the SH2 and SH3 domains of c-Src and hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* 2001;105(1):115–26.
- Hünenberger PH, Mark AE, van Gunsteren WF. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J Mol Biol* 1995;252(4):492–503.
- Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* 1991;11(3):205–17.
- Floyd RW. Algorithm 97: Shortest path. *Commun ACM* 1962;5(6):345. <https://doi.org/10.1145/367766.368168>.
- Sethi A et al. Dynamical networks in tRNA: protein complexes. *PNAS* 2009;106(16):6620–5.
- Ye W et al. Allosteric autoinhibition pathway in transcription factor ERG: dynamics network and mutant experimental evaluations. *J Chem Inf Model* 2017;57(5):1153–65.
- Yang J et al. Synergistic allosteric mechanism of fructose-1,6-bisphosphate and serine for pyruvate kinase M2 via dynamics fluctuation network analysis. *J Chem Inf Model* 2016;56(6):1184–92.

- [59] Schein CH, Noteborn MHM. Formation of soluble recombinant proteins in *Escherichia Coli* is favored by lower growth temperature. *Bio/Technology* 1988;6(3):291–4.
- [60] San-Miguel T, Pérez-Bermúdez P, Gavidia I. Production of soluble eukaryotic recombinant proteins in *E-coli* is favoured in early log-phase cultures induced at low temperature. *Springerplus* 2013;2(1). <https://doi.org/10.1186/2193-1801-2-89>.
- [61] Li F et al. The disulfide bond between Cys22 and Cys27 in the protease domain modulate clotting activity of coagulation factor X. *Thromb Haemost* 2019;119(06):871–81.