scientific reports

Check for updates

OPEN

Molecular dynamics simulations reveal membrane lipid interactions of the full-length lymphocyte specific kinase (Lck)

Dheeraj Prakaash[1,2], Charline Fagnen[1,2], Graham P. Cook[3], Oreste Acuto[4] & Antreas C. Kalli[1,2]✉

The membrane-bound lymphocyte-specific protein-tyrosine kinase (Lck) triggers T cell antigen receptor signalling to initiate adaptive immune responses. Despite many structure–function studies, the mode of action of Lck and the potential role of plasma membrane lipids in regulating Lck's activity remains elusive. Advances in molecular dynamics simulations of membrane proteins in complex lipid bilayers have opened a new perspective in gathering such information. Here, we have modelled the full-length Lck open and closed conformations using data available from different crystalographic studies and simulated its interaction with the inner leaflet of the T cell plasma membrane. In both conformations, we found that the unstructured unique domain and the structured domains including the kinase interacted with the membrane with a preference for PIP lipids. Interestingly, our simulations suggest that the Lck-SH2 domain interacts with lipids differently in the open and closed Lck conformations, demonstrating that lipid interaction can potentially regulate Lck's conformation and in turn modulate T cell signalling. Additionally, the Lck-SH2 and kinase domain residues that significantly contacted PIP lipids are found to be conserved among the Src family of kinases, thereby potentially representing similar PIP interactions within the family.

Activation of T cells is triggered by the engagement of the T cell receptor (TCR) with antigenic peptides presented by major histocompatibility complexes (pMHC)[1,2]. Upon pMHC binding, allosteric sites in the extracellular and transmembrane regions of the T cell receptor-CD3 complex (TCR-CD3)[3,4] promote exposure of immunoreceptor tyrosine-based activation motifs (ITAMs) in the cytoplasmic tails of CD3/ζ subunits. Further, ITAMs are promptly phosphorylated by Lck. Remarkably, non-activated T cells maintain a sizable fraction of constitutively activated Lck at the plasma membrane that is necessary and sufficient for ITAM phosphorylation upon ligand binding[5]. Additionally, imaging studies have suggested that Lck conformational states dictate its spatial distribution that may impact TCR-CD3 ITAM phosphorylation[6]. Phosphorylated ITAMs then provide stable binding sites for the tyrosine kinase ZAP-70[7,8] that is regulated by Lck to propagate signals required for T cell activation[9].

Understanding the role of Lck in molecular detail is important in deciphering the initial phases of T cell activation. To achieve this, it is key to obtain the full-length 3D structure of Lck which remains structurally unresolved until date. The full-length Lck contains the following domains (from the *N* to *C* terminus): the SH4 (first ~ 10 residues), unique domain (UD; following ~ 50 residues) both of which are likely to be devoid of secondary structure. The UD is followed by the SH3, SH2, and the kinase domains for which structural data is available. The X-ray crystallographic structure of the Lck-SH2 and SH3 domains combined is available at a resolution of 2.36 Å (PDB:4D8K), and the isolated kinase domain in its active state i.e., phosphorylated at Y394 is available at 1.7 Å resolution (PDB:3LCK)[10]. However, the structures of the SH4 and UD remain largely unresolved. Nonetheless, NMR data for Lck-UD in solution indicates that it lacks structure and has no significant influence on Lck-SH3[11]. In this work, we refer to the SH4 and the UD combined as the 'SH4-U' domain for simplicity. In the SH4 domain, G2, C3 and C5 undergo acylation as a post-translational acylation i.e., myristoylation at G2[12,13], and palmitoylation at C3 and C5[14]. As a result, the acyl chains or lipid tails covalently attached to these residues insert into the hydrophobic core of the membrane and aid in membrane localization of Lck[15].

[1]Leeds Institute of Cardiovascular and Metabolic Medicine, School of Medicine, University of Leeds, Leeds, UK. [2]Astbury Center for Structural Molecular Biology, University of Leeds, Leeds, UK. [3]School of Medicine, Leeds Institute of Medical Research, University of Leeds, Leeds, UK. [4]Sir William Dunn School of Pathology, University of Oxford, Oxford, UK. ✉email: a.kalli@leeds.ac.uk

Lipids in the inner leaflet of the plasma membrane have been reported to play an important role in interacting with Lck via its SH2 domain and in turn regulating TCR-CD3 signalling[16]. In particular, anionic lipids such as phosphatidylinositol-4,5-biphosphate (PI(4,5)P2 or PIP$_2$) and phosphatidylinositol-3,4,5-triphosphate (PIP$_3$) were suggested to aid in Lck interaction with the TCR-CD3 in a spatiotemporal manner. The Lck-SH2 domain was found to be key to lipid interaction by selectively contacting these PIP lipids via a cationic patch[16]. SH2 domains in other tyrosine kinases such as Zap70 have also been reported to direct signalling pathways by binding to PIP lipids[17]. Our previous studies suggested that the TCR-CD3 maintains an anionic lipid environment enriched in PIP lipids with the help of its cytoplasmic region[18]. Lck was also shown to possess high affinity for PIP lipids[16] and its clustering is driven by their open conformational state[6]. However, lateral interactions of Lck molecules may also include other Lck isoforms as suggested by a recent study[19]. Nevertheless, it is important to understand how the open and closed states of Lck interact with the membrane in molecular detail. This could further aid in our understanding of its interaction with ITAMs of stimulated TCR-CD3 complexes.

In this study, we use molecular modelling followed by molecular dynamics (MD) simulations to elucidate the lipid interactions of the open and closed states of the full-length Lck. MD simulations were previously used to study membrane interactions with peripheral membrane proteins with good agreement with biological experiments[20,21]. MD simulation studies of the PLCγ1 core domain revealed its orientation with the membrane. Mutation of the residues that were identified by the simulations to be important for the interaction of PLCγ1 core domain with the membrane reduced the binding of PLCγ1 to liposomes[22]. MD simulations have also been used to study the association of other peripheral membrane proteins to the membrane including the talin head domain (or parts of the talin head)[23,24], the KRas-4B[25] and PH domains[20].

Here, we modelled the full-length Lck by predicting the structure of the SH4-U domain and integrating it with the experimentally resolved structures of the SH2, SH3 and kinase domains. Further, we performed coarse-grained molecular dynamics simulations over a cumulative time of 100 microseconds for each of the open and closed states of Lck in a complex symmetric bilayer whose lipid headgroup composition resembles the inner leaflet of the T cell plasma membrane[26]. From these simulations, we study the conformational dynamics and lipid interactions of the open and closed conformations of the full-length Lck.

Results and discussion

Modelling the full-length Lck in its open and closed states.
To obtain a model of the 3D structure of full-length Lck, we first modelled the SH4-U domain since its structure is unknown. To achieve this, we used two independent 3D structure prediction tools i.e., I-Tasser[27] and Robetta[28], and obtained multiple 3D models from each. Then, the PSIPRED Protein Analysis Workbench[29] was used to calculate the secondary structure of the SH4-U domain (Fig. S1A). The best 3D structural models, one from I-Tasser and one from Robetta (representing the highest prediction confidence score and agreement with secondary structure predictions) were subjected to 250 ns atomistic molecular dynamics (ATMD) simulations in solution neutralized by 0.15 M Na$^+$ and Cl$^-$ ions to allow optimization of the predicted structures. At the end of the ATMD simulations, we used two criteria to select the best model: (i) agreement with secondary structure predictions, and (ii) agreement with structural information revealed by an NMR study[30] where the UD contained a hairpin-like loop region (Fig. 1A left). This loop region was found to be responsible for binding CD4 and CD8 co-receptors via a coordinating Zn^{2+} ion[30], though data have indicated that not all Lck molecules are bound to co-receptors. Using these criteria, the model derived from the Robetta server was selected. However, due to the absence of a zinc ion in our model, a disulphide bond was formed. In addition, our model suggests that residues E10, D11, D12, E15, E21 (Fig. 1A) form an anionic patch that potentially interacts with cationic regions of the TCR-CD3 cytoplasmic region. Moreover, the D12N mutation was shown to reduce binding with CD3ε BRS by NMR experiments[31].

Following modelling of the SH4-U domain, the crystal structures of SH2, SH3 (PDB:4D8K), and kinase domains (PDB:3LCK) along with the SH4-U model were assembled in a linear manner to model the full-length Lck open conformation (Fig. 1B) using UCSF Chimera[32]. Missing residues located in the linker region were predicted to be unstructured and hence modelled as loops between the domains using Modeller 9.2[33]. The structures of the different domains were assembled sufficiently far from each other to avoid bias in protein–protein interactions at the beginning of the simulations. Note that the positioning of the SH3 and SH2 domains relative to each other were not altered and were used as obtained from the crystal structure (PDB:4D8K).

The crystal structure of the closed conformation of Hck, an Src family member of kinases, resolved at 1.65 Å (PDB:5H0B)[34] was used as a template to model the Lck closed state using Modeller 9.2[33]. The resultant homology model of the Lck closed state (containing SH2, SH3, kinase domains) was then conjoined with the SH4-U model (as shown in Fig. 1A) to obtain the full-length Lck (Lck-FL) in its closed state (Fig. 1C). This modelling used a multiple sequence alignment of Hck and Lck produced by Clustal Omega[35]. The structures of the SH2, SH3, kinase domains of Lck were also individually aligned with those respective domains of Hck (Fig. 1D) using the 'super' aligning method in PyMOL (pymol.org) indicating their structural similarities i.e., RMSD = 0.582, 1.135, 0.92 Å respectively.

Finally, to both the Lck-FL open and closed models, post-translational modifications were added to the N-terminal residues i.e., G2 was myristoylated and, C3 and C5 were palmitoylated prior to the simulations. The initiator Met1 residue was removed during this process since they are known to be cleaved in mature eukaryotic proteins[16,17]. These models were then coarse-grained using the *martinize* script and the Martini 2.2 forcefield[36], and placed in a complex symmetric membrane environment using the *insane* tool[37] such that the post-translational modifications of the Lck N-terminus were partially inserted in the membrane. Following this, coarse-grained molecular dynamics (CGMD) simulations were conducted (see Table 1 for details).
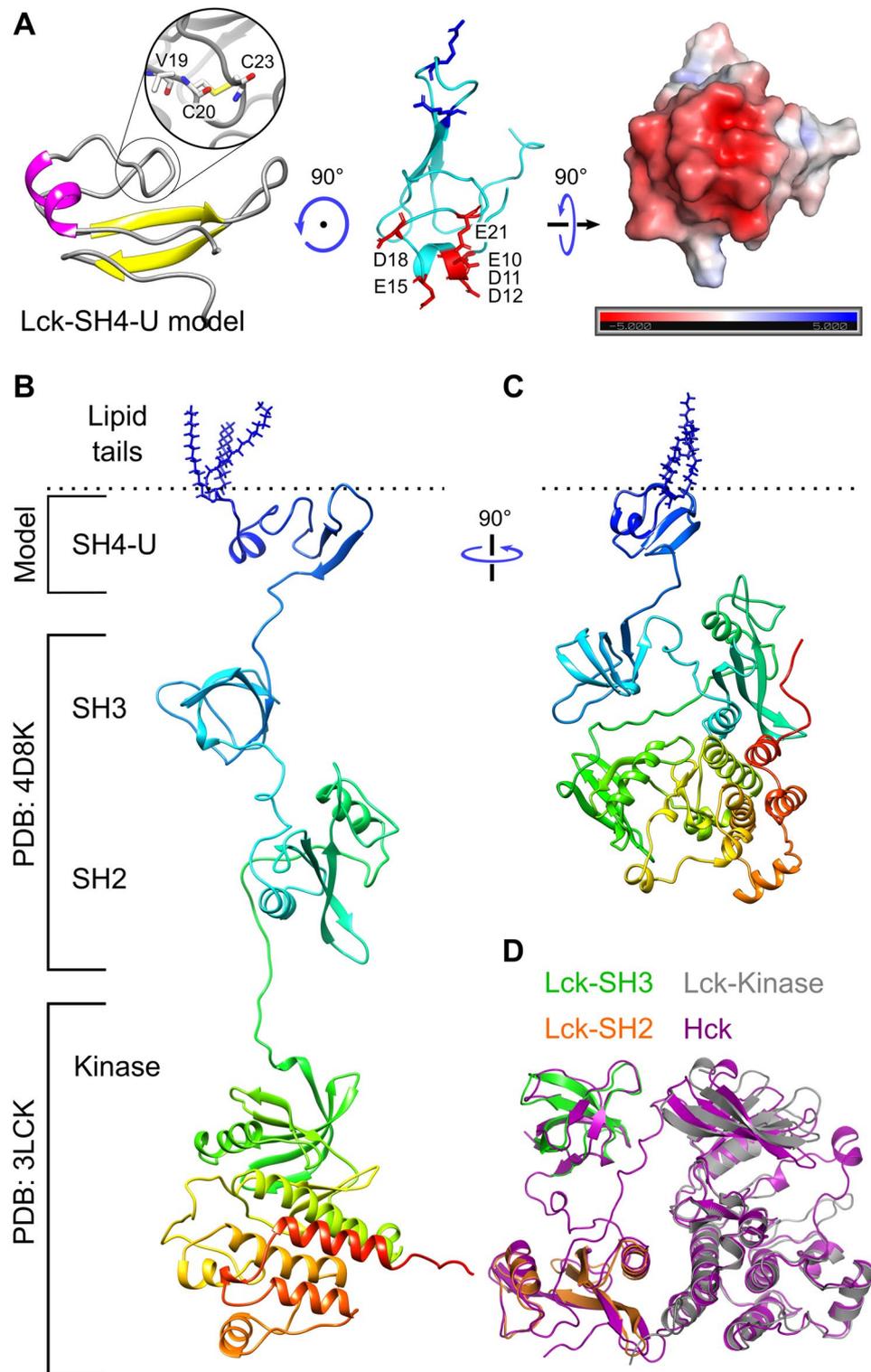
Figure 1. Model of the SH4-U domain and, the open and closed full-length Lck conformations. (**A**) The model of the SH4-U domain (residues 2 to 63) used in this study. The region responsible for coordinating a Zn^{2+} ion is magnified (left). Residues forming the anionic patch in our model are shown as red sticks and labelled. PIP lipid binding cationic residues (R39, R45), as suggested in this study, are shown as blue sticks are located on the opposite side of the anionic patch (middle). The electrostatic profile of the anionic patch (right) shown was calculated in the ± 5 kT/e range and at pH 7.0 using the PDB2PQR[38] and APBS[39] tools. Electronegative and electropositive regions are indicated by red and blue intensities respectively. (**B**) The model of the open Lck-FL conformation and (**C**) the closed Lck-FL conformation used in this study. (**D**) Isolated SH2, SH3 and kinase domains aligned to the closed state of Hck (PDB:5H0B)[34].

Simulations	Membrane	Particles	Simulation box (X × Y × Z axis) nm	Duration	Replicas
Lck-SH4-U	complex	25,526	12 × 12 × 17	1 µs	20
Lck-SH3	complex	25,686	12 × 12 × 20	1 µs	20
Lck-SH2	complex	25,412	12 × 12 × 20	1 µs	20
Lck-SH2,3,4-U	complex	53,576	16 × 16 × 23	5 µs	20
Lck-FL open	complex	83,137	19 × 19 × 26	5 µs	20
Lck-FL open	complex without PIPs	79,099	19 × 19 × 26	5 µs	20
Lck-FL open mut5	complex	83,194	19 × 19 × 26	5 µs	20
Lck-FL closed	complex	46,187	16 × 16 × 20	5 µs	20
Lck-FL closed	complex without PIPs	43,488	16 × 16 × 20	5 µs	20

Table 1. Summary of CGMD simulations conducted in this study. (Symmetric) membrane composition: complex: POPC/POPE/POPS/Chol/PIP$_2$/PIP$_3$ = 12/40/20/20/6/2; complex without PIPs: POPC/POPE/POPS/Chol = 20/40/20/20.

Membrane association and lipid interaction of the full-length Lck.

To assess the association of the Lck-FL models with the membrane (see Table 1 for membrane composition), we performed CGMD simulations. At the beginning of these simulations, the post-translational modifications (lipid tails) of both the Lck-FL open and closed models were made to partially penetrate the membrane surface to mimic the fact that the lipid tails are expected to penetrate the membrane upon binding of Lck to the membrane. 20 individual simulations for 5 µs were performed for both the open and closed models. Calculation of the average distance versus time of the center of mass (COM) of the initial protein model to the COM of the membrane along the vertical (Z) axis showed that both Lck-FL models closely associated with the membrane within 1 µs simulation time (Fig. 2A). Analysis of Lck's preference in lipid contacts was performed by normalizing the contacts of lipids by the number of respective lipids available in the membrane. This showed an increase in Lck's preference for PIP$_2$ and PIP$_3$ in both its open and closed conformations (Fig. 2B). This preference created an anionic lipid annulus around Lck and was retained for the remaining time of the simulations. For both models, a normalized radial distribution function (RDF) analysis also showed that PIP$_2$ and PIP$_3$ were preferred by Lck over other lipids (Fig. 2C). In addition, experimental studies suggest that the SH2 domain prefers to interact with PIP$_3$ compared to PI(4,5)P2[16]. Interestingly, these lipid species were found to also cluster around the TCR-CD3 and interact with its cytoplasmic region in our previous studies of the complete TCR-CD3 complex[18].

Given the potential of strong electrostatic interactions between Lck and PIP lipids, and also between the TCR-CD3 cytoplasmic tails and PIP lipids of the inner leaflet of the membrane[18], it is possible that TCR-Lck association and ITAM phosphorylation occurs proximal to the inner leaflet of the plasma membrane.

PIP lipid binding sites.

Analysis of the interactions of the Lck-FL open conformation (final snapshot shown in Fig. 3A left) with PIP lipids showed that the SH2 domain made significant contacts via a primary binding site (K182 > R184 ~ R134 > K135 ~ K179). The SH4-U domain also made significant contacts by preferring to bind to PIP lipids via R39 and R45, followed by cholesterol interactions via myristoylated G2, palmitoylated C3 and C5, followed by residues H24, Y25, P26, V44, R45, D46 (Fig. 3B). The SH3 domain made less contact with the membrane, interacting mostly via K84. Interestingly, the kinase domain of Lck also showed significant PIP lipid interactions; the most interactive residues were R455, R458, R474, K478 (Fig. 3B). These residues are situated at the bottom surface of the C-terminal lobe of the kinase domain and constitute a flat cationic area acting as a PIP lipid binding site (Fig. S1C). Furthermore, in our simulations, residues including and neighbouring A160 were observed to interact with lipids, but not as significantly as K182 and R184 (Fig. 3B). This observation is consistent with mutation studies which revealed that A160K reduces dissociation of Lck-SH2 from plasma membrane-mimetic vesicles[16].

In the simulations of the closed Lck-FL, the lipid interaction profile of the SH4-U, SH3, and kinase domains remained fairly similar to the simulations of the open Lck-FL. However, in the closed Lck-FL, the SH2 domain exhibited a distinct PIP lipid binding site (referred hereon as secondary PIP binding site) consisting of residues H148, R219, and R222 (Fig. 3C). Note that K182 and R184 present in the primary binding site interacted with PIP lipids in both open and closed Lck-FL. However, in the closed Lck-FL, PIP interaction was mostly observed in the secondary binding site via residues H148, R219, R222. Finally, we note that both open and closed models of the full-length Lck showed that PIP$_3$ exhibits a greater preference compared to PI(4,5)P$_2$ in our CGMD simulations. This adds to previous findings by SPR analysis that the SH2 domain prefers to interact with PIP$_3$ followed by PI(4,5)P$_2$[16].

Cholesterol interactions can be observed with Lck-SH4-U in all the Lck-FL models (Fig. 3B–D; magnified by 2.5x in Fig. S2A). These cholesterol interactions were primarily driven by the post-translational modifications i.e., the myristoylated G2, the palmitoylated C3 and C5 residues. In the absence of PIPs in the membrane, the Lck-SH2 PIP binding site and its preferred membrane-binding orientation was disrupted whereas a new interaction i.e., Lck-SH2 R196 with POPS lipids was observed but not within the primary or secondary PIP binding site region as described in this paper. In addition, the interactions of cholesterol and POPS with Lck-SH4-U increased and the residues identified to interact with PIP in the kinase domain interacted with POPS but to a lower extent (Fig. S2B). This suggests that the rate of electrostatic interactions of Lck with the membrane reduces
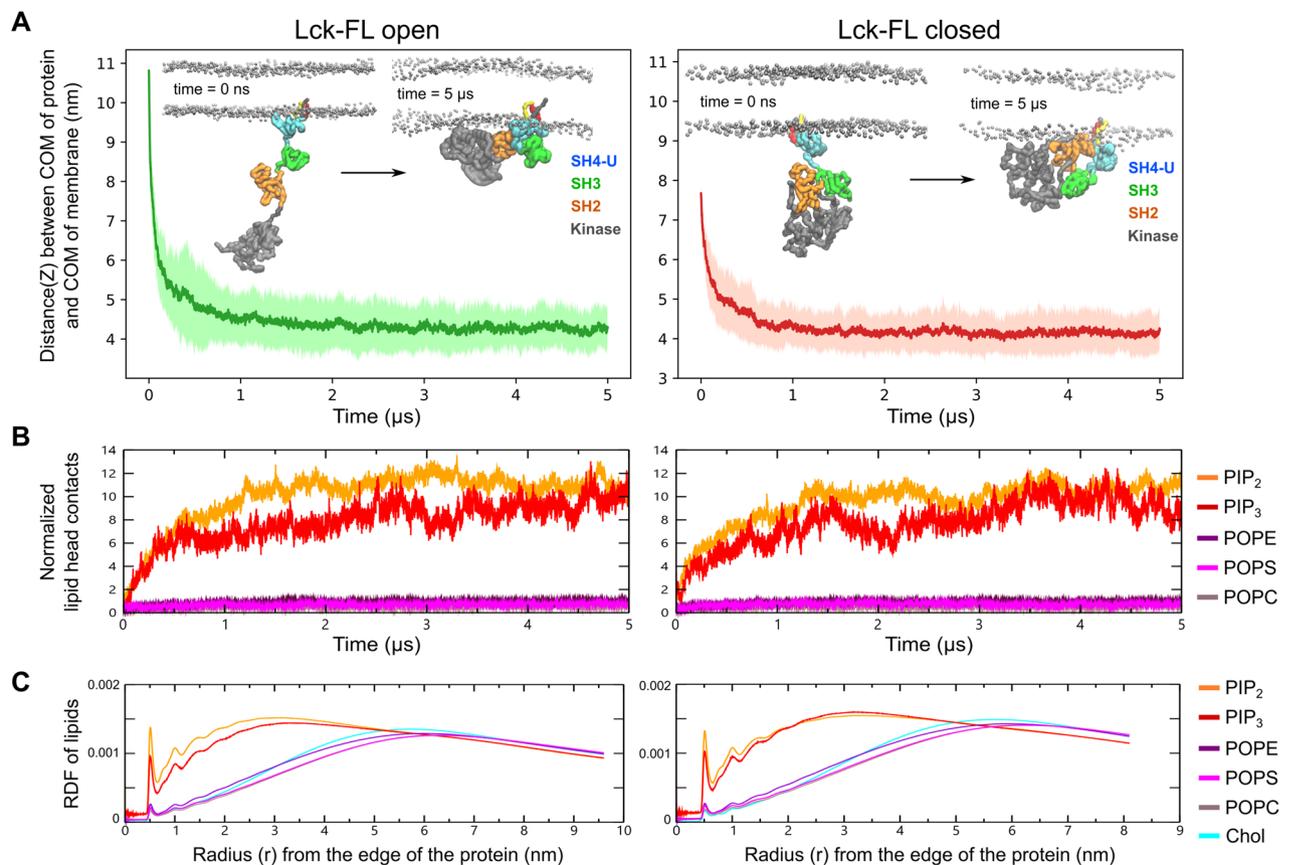
Figure 2. Membrane association and lipid interactions of the open and closed Lck-FL conformations. (**A**) Association of open and closed conformations of Lck-FL with the membrane is indicated by the reduction in distance between the center of mass (COM) of Lck-FL and COM of the membrane versus time. The dark green/red lines indicate the average distance from all 20 CGMD simulations while the light green/red region indicate standard deviation. See Fig. S2C for distance vs time plots from individual simulations. (**B**) Normalized number of contacts between Lck and lipid headgroups versus time averaged from all 20 CGMD simulations. The number of headgroup interactions of each phospholipid type is normalized by the number of lipids of the respective lipid type in the membrane. See Fig. S1B for non-normalized contacts of lipid headgroups with Lck. (**C**) The radial distribution function (RDF) of all lipid types around Lck calculated throughout the simulation time from all 20 CGMD simulations. The RDF is normalized by the total number of lipids in the membrane of that system to enable comparison between the open and closed conformations of Lck-FL. Note: The Lck-FL open and closed systems contain different number of lipids in the membrane due to different sizes of the bilayers used in our simulations.

in the absence of PIPs, thereby leading to weaker membrane attachment. In the absence of PIPs in the membrane, our Lck-FL models bound to the membrane within a similar time frame as in the simulations with PIPs in the membrane (Fig. S2C). However, our radial distribution function analyses combined with normalized contact analyses indicate that cholesterol and POPS lipid interactions were preferred upon removal of PIP lipids from the membrane (Fig. S2D).

Lck-SH2 adopts a secondary PIP lipid binding site upon mutation of the primary binding site. To investigate the significance of the primary PIP lipid binding site of the open Lck-FL identified above, we mutated its residues i.e., R134A, K135A, K179A, K182A, R184A. This mutation (referred as mut5) in the open Lck-FL led to the loss of PIP interaction via the primary binding site but formed contacts with PIPs via the secondary binding site (Fig. 3D), thereby resembling the lipid interaction profile of the closed Lck-FL (Fig. 3C). The lipid interactions and orientations of the other domains remained unaffected by this mutation in the SH2 domain.

The fact that the secondary binding site of Lck-SH2 (H148, R219, R222) is located on the opposite side of the primary binding site (Fig. 3E) and dominated PIP interaction in the closed Lck-FL suggests that the closed conformation potentially alters the preferred/primary membrane-binding orientation of Lck-SH2 to some degree. This suggests that Lck-SH2 can attain a secondary membrane-bound conformation but is less preferred and potentially weaker. This secondary binding site was observed frequently in the closed state of Lck.

Note that the orientation of Lck-SH2 was not retained for the entire simulation once bound to the membrane. In most individual simulations, the SH2 domain was able to tilt relative to the membrane and explore its conformational space in both the open and closed states of Lck-FL (Figs. S3, S4). Taking all simulations in
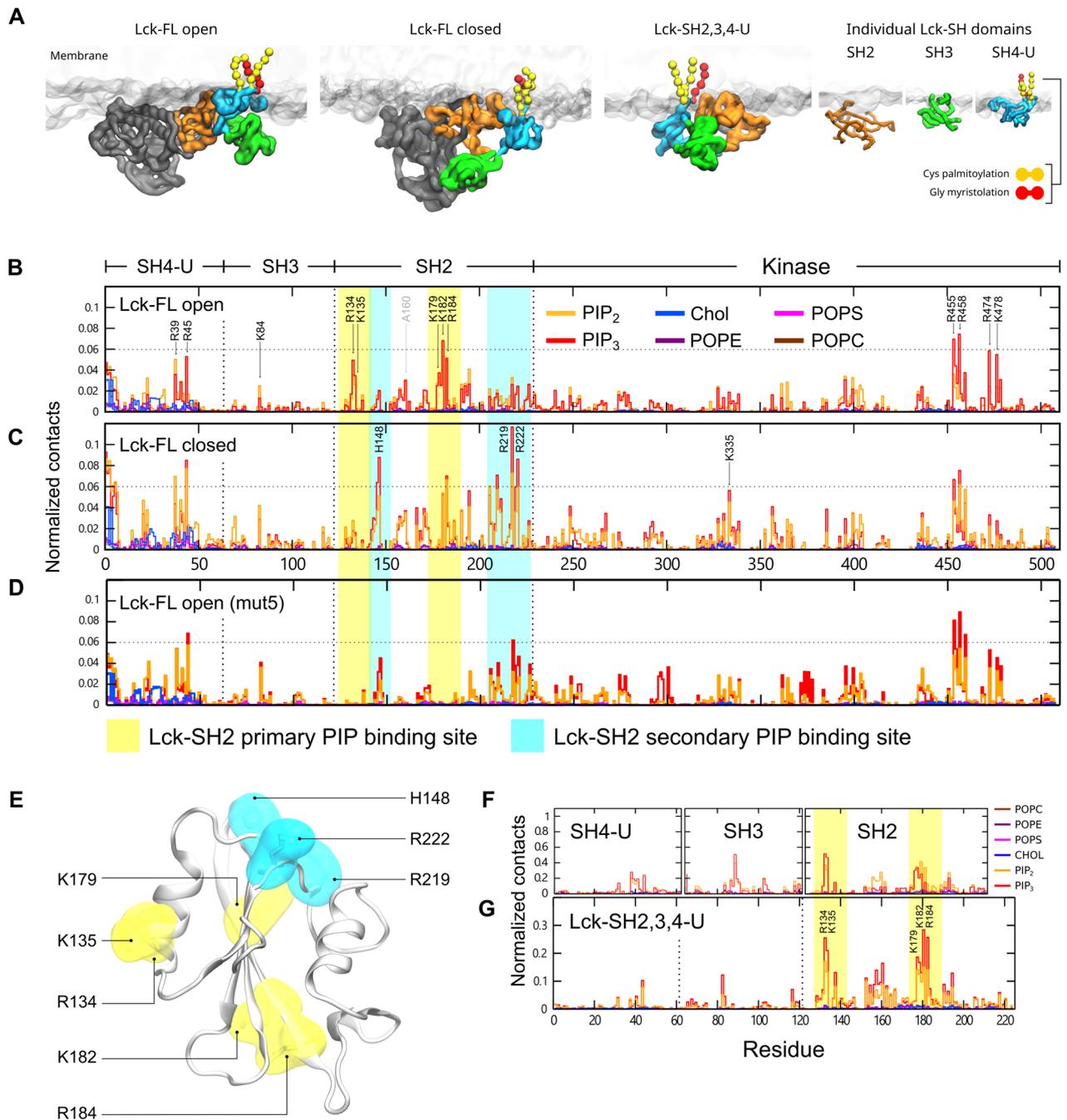
Figure 3. Snapshots of the initial simulation setup and PIP lipid binding sites. (**A**) The final snapshots from one of the simulations of the open and closed conformations of Lck-FL. Also, the final frames of one of the simulations of the Lck-SH domains combined (SH2, SH3, SH4-U) and of the individual Lck-SH domains where the protein structure was initially placed in solution ~ 6 nm away from the membrane. (**B**) Normalized lipid interactions of Lck-FL open, (**C**) Lck-FL closed, (**D**) Lck-FL open when mutated (mut5 i.e., R134A, K135A, K179A, K182A, R184A), (**E**) Residues constituting the primary (yellow) and secondary (cyan) PIP lipid binding sites of Lck-SH2 as observed in CGMD simulations of Lck-FL open and closed. See Fig. S2A for a magnified view of lipid interactions other than PIPs. (**F**) Normalized lipid interactions of the SH4-U, SH3, SH2 domains when individually simulated, and (**G**) of the SH4-U, SH3, SH2 domains simulated in conjunction (Lck-SH2,3,4-U).

account, Lck-SH2 preferred to take up the primary PIP binding site in the open Lck-FL (Fig. 3B, Fig. S3A), and the secondary PIP binding site in the closed Lck-FL (Fig. 3C, Fig. S4A).

It was shown that Lck exhibits lesser membrane binding if its preferred PIP lipid binding site K182/R184 is altered[16]. The fact that Lck-membrane binding was reduced and not completely diminished indicates that this

secondary PIP lipid binding site may aid membrane association to a certain degree but reduce colocalization with stimulated TCR-CD3 due to change in its orientation. During spatial re-organization of Lck with TCR-CD3 upon activation[40], this alteration of SH2 domain orientation and PIP lipid binding site may also reduce its competence with the preferred open Lck conformation as previously suggested by other studies[6,41].

Simulations of the isolated domains reveal similar interaction with PIP lipids. Following our investigation of lipid interactions of the Lck-FL open and closed states, we also simulated the Lck-SH2, SH3, SH4 domains individually to be able to analyse their lipid interactions independently of the influence of the other domains. Given that the kinase domain is the largest domain, constituting greater than half of the Lck-FL sequence, we also simulated the Lck-SH domains combined (Lck-SH2,3,4-U) to assess their lipid interactions without the influence of the kinase domain. In all the individual Lck-SH and SH2,3,4-U simulations, the protein structure was placed ~ 6 nm away from the bilayer (Fig. 3A) to allow it to explore all possible orientations in solution before binding to the membrane.

These simulations suggested that R134, K135, K179, K182, R184 of Lck-SH2 were the most interactive residues with PIP lipids, the first two preferring PIP$_3$ over PIP$_2$ (Fig. 3F,G) as suggested by previous experimental studies[16] and by our simulations with the Lck-FL models in this study. Lck-SH2 was also found to bind to the membrane within 700 ns of simulation time in the majority of simulations (Fig. S1D). The SH3 domain required a similar amount of simulation time to bind to the membrane. R89 of the Lck-SH3 was the most interactive residue with PIP lipids in the simulations with the isolated Lck-SH3, but K84 was the most interactive in the simulations in which Lck-SH3 was conjoined with other domains including the Lck-FL models. Lck-SH4-U made a very small number of contacts when simulated individually (Fig. 3F) due to the lack of a strong PIP lipid binding site, and because the myristoylated and palmitoylated lipid tails failed to insert into the membrane in some of the simulations (Fig. S1D). As a result, we found a smaller fraction of SH4-U bound to the membrane compared to the SH2 and SH3 (Fig. S1E). Note that, in some individual SH4-U simulations where its lipid tails inserted into the membrane, the SH4-U domain stayed membrane-bound for the rest of the simulation time (S1 Movie).

In the Lck-SH2,3,4-U simulations, the Lck-SH2 dominated the interactions with PIP lipids (with R134, K135, K182, R184), while those of SH3 (K84) and SH4-U (R45) were observable but not significant (Fig. 3G). This indicated that Lck-SH2 lipid interactions were not influenced by the other SH domains. Note that, although the protein had achieved a membrane bound state via the SH2 domain (by ~ 1.5 μs in all simulations) (Fig. S1D bottom), the SH4 lipid tails had not inserted in the membrane. Instead, the lipid tails were found binding to a small cavity near the SH2-SH3 linker before the protein attained its membrane-bound state (Fig. S1F) and did not insert into the membrane despite lipid binding initiated by the SH2 domain. This is presumably due to strong hydrophobic interactions between the lipid tails and the SH2-SH3 linker region, and possibly energetically unfavourable to switch to a membrane inserted state. However, it is important to note that membrane insertion may be achieved given more simulation time. Consistent with this observation, in vitro studies have reported that the N-terminal myristoyl group in c-Src binds to the SH3 domain while in solution and modulates membrane anchoring[42].

Simulations of Lck-FL indicate flexibility of the kinase domain in the open conformation. We deduced the most observed conformations of the membrane bound Lck-FL in its open state from the CGMD simulations using clustering analysis and with a 0.35 nm RMSD cut-off. In the top three most observed conformations of Lck-FL, we observed that Y394 and Y505 often switched positions i.e., in one conformation, Y394 is proximal to the SH2 domain whilst in another conformation, Y505 is proximal to the SH2 domain (Fig. 4A). This indicates that the kinase domain can rotate and re-orient relative to the SH2 domain. We also performed atomistic MD (ATMD) simulations of the Lck-FL open state in solution (250 ns × 3 replicas) and found similar activity of the kinase domain, where pY394 and Y505 switched positions alternating their proximity to the SH2 domain (Fig. 4B).

This flexibility of the kinase domain in the Lck-FL open conformation is potentially key to the dynamics of its catalytic activity. Note that, in the ATMD simulations, despite taking up positions near the SH2 domain, neither pY394 nor Y505 contacted the Lck-SH2 PIP lipid binding site (Fig. 4C) suggesting that Lck-SH2 is free to bind to the membrane in the open state, unlike in the closed state where pY505 interacted with some PIP lipid binding residues of Lck-SH2 (Fig. 4D).

Atomistic simulations of the Lck-SH4-U domain reveal its anionic patch. In addition to the open Lck-FL in solution, we performed ATMD simulations of the closed Lck-FL in solution (250 ns × 3 replicas). From these simulations, clustering analyses were performed on the SH4-U domain alone. As a result, we obtained a structure of the most observed conformations of the Lck-SH4-U in the open and closed states of Lck. The same analysis was performed for the CGMD simulations to enable comparison with the ATMD simulations. We then calculated their electrostatic profiles and compared them with that of the initial SH4-U model (Fig. 5). This revealed that the SH4-U domain, despite its dynamic nature, maintained an anionic patch (E10, D11, D12, E15, D18, E21) which was independent of the open and closed Lck-FL conformations. Interestingly, this anionic patch was identified on the opposite side of its PIP lipid binding surface (R39, R45) indicating its availability to bind to cationic residues of other proteins, especially those of the BRS motifs of CD3ε and ζ subunits of the TCR-CD3 complex as suggested by NMR experiments[31].
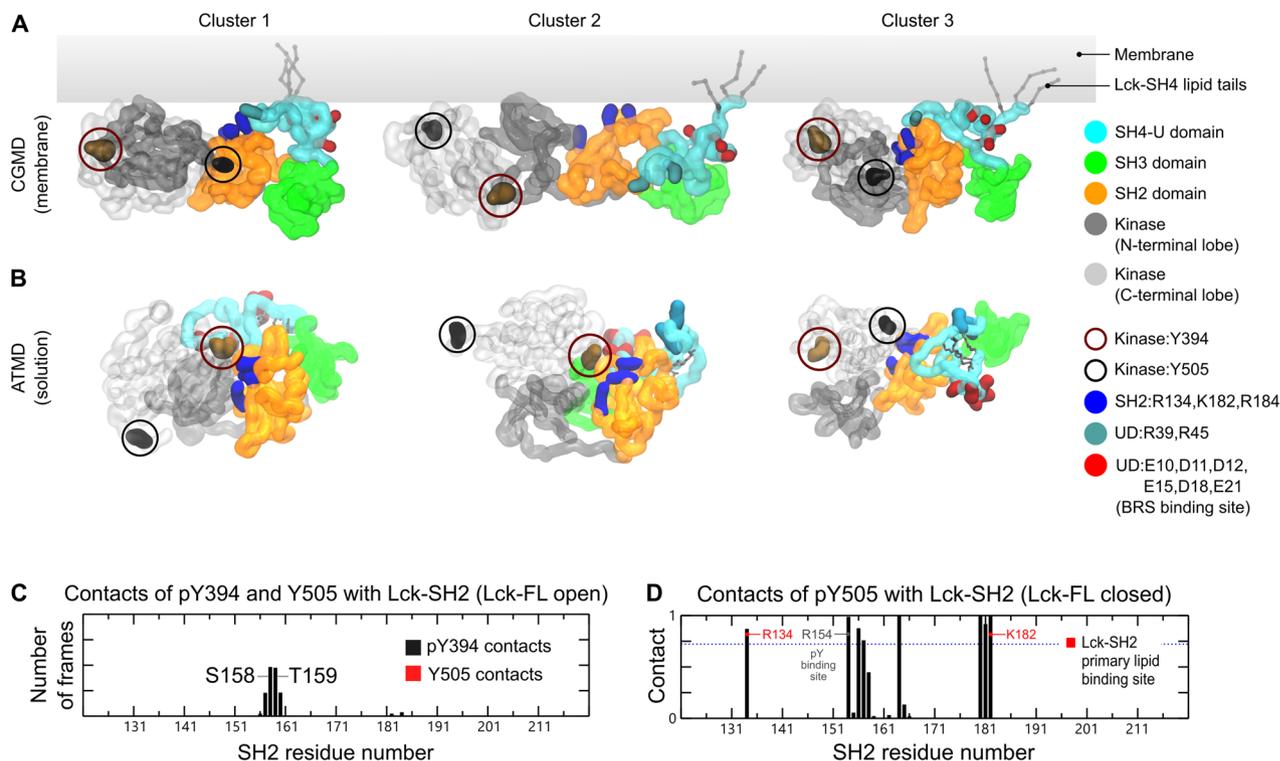
Figure 4. The flexibility of the kinase domain highlighted by the top three representative conformations of Lck-FL open (**A**) from CGMD (top) and (**B**) ATMD simulations (bottom) derived from clustering analyses (see main text for details). The kinase domain is made transparent to clarify the positions of Y394 and Y505. Y505 and Y394 which are shown in circles exchange positions relative to Lck-SH2, in both the membrane-associated form of Lck (CGMD) and in solution (ATMD) during our simulations. This indicates the rotation of the kinase domain relative to the SH2 domain that demonstrates the flexibility of the kinase domain. (**C**) Normalized average number of contacts of pY394 and Y505 with Lck-SH2 in the open Lck-FL state. (**D**) Normalized average number of contacts of pY505 with Lck-SH2 domain in the Lck-FL closed state in ATMD simulations. Normalization was done by dividing the number of contacts by the number of simulation frames thereby obtaining a scale of 0 to 1.

Conclusion

In this study, we have revealed lipid interactions of the full-length post-translationally modified Lck in both its open and closed conformations and highlighted its PIP lipid binding sites. Our key finding concerning PIP lipid binding was that the SH2 domain adopts a secondary PIP binding site in the closed state of Lck-FL compared to its open state. Although this secondary PIP binding site may aid in membrane localization to some degree, it may be less competent compared to the primary PIP binding site of the Lck-SH2 domain during spatial organization of open Lck during T cell activation. Moreover, from our simulations of the Lck-SH2,3,4-U, we observed that electrostatic interactions with the membrane is primarily driven by the primary PIP binding site of the Lck-SH2 domain. These simulations also suggest that the insertion of the N-terminal post-translational modifications of Lck into the membrane may depend on Lck-SH2-driven membrane association. Upon binding to the membrane, we show that the full-length Lck is surrounded by a pool of negatively charged lipid headgroups creating an anionic environment, whereas it is also observed that the TCR-CD3 cytoplasmic tails maintain an anionic lipid environment[18]. This sheds light on the potential significance of lipids during TCR-Lck association[19].

In this study, we suggest that the residues R134, R135, and K179 contribute to the primary PIP lipid binding site in addition to those previously reported i.e., K182 and R184[16]. Moreover, all five of these residues are found to be conserved among Src family members (Fig. S5A) indicating that all their SH2 domains are likely to localize to the membrane with the same orientation. The PIP interactions of Lck-SH3, although not significant, consistently interacted via K84. Upon lipid tail insertion into the membrane, the Lck-UD also showed significant interaction with PIPs via R39 and R45. Furthermore, we present a structural model of the Lck-UD that reveals an anionic patch with which it could bind to basic-rich motifs of the TCR-CD3 subunits and aid in TCR-Lck association.

We also found that the kinase domain interacts with PIPs in the membrane via residues R455, R458, R474, K478 which form a cationic patch at the bottom of its C-terminal lobe. These residues, along with other potentially PIP contacting residues, were also found to be conserved among other Src family members (Fig. S5A) suggesting similar lipid interactions and membrane-bound orientations of the kinase domains in the Src family. Further, the lipid interactions of the kinase domain implies that it is likely to be situated proximal to the membrane surface. Therefore, given that the TCR-CD3 cytoplasmic region is also closely associated with the membrane surface[18], it is likely that Lck kinase-mediated phosphorylation of TCR-CD3 ITAMs and of other
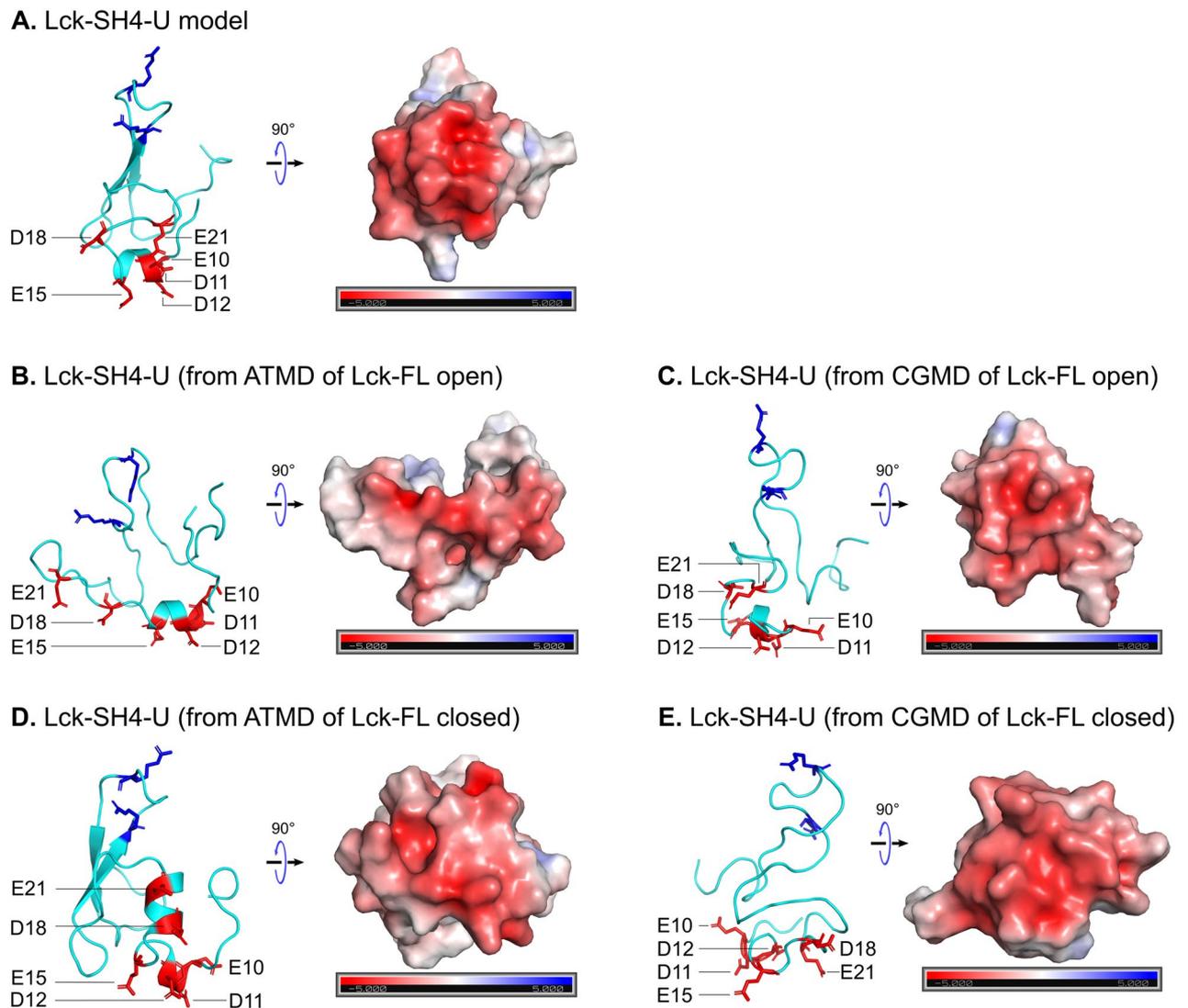
A. Lck-SH4-U model

B. Lck-SH4-U (from ATMD of Lck-FL open)

C. Lck-SH4-U (from CGMD of Lck-FL open)

D. Lck-SH4-U (from ATMD of Lck-FL closed)

E. Lck-SH4-U (from CGMD of Lck-FL closed)

Figure 5. The electrostatic profiles of the Lck-SH4-U domain as in the initial model (**A**), and according to clustering analyses after simulating the Lck-FL open state (**B,C**), and the Lck-FL closed state (**D,E**). Their electrostatic profiles were calculated in the ± 5 kT/e range and at pH 7.0 using the PDB2PQR[38] and APBS[39] tools. Electronegative and electropositive regions are indicated by the red and blue intensities respectively. The residues forming the anionic patch are shown as red sticks and labelled. PIP lipid binding residues (R39, R45) are shown as blue sticks for reference.

downstream signalling proteins during the initial phase of T cell activation is carried out proximal to the plasma membrane. Similarly, given the conservation of important PIP binding sites in the SH2 and kinase domains, ITAM phosphorylation mediated by other members of the Src family may also occur close to the surface of the plasma membrane.

It is also important to consider some limitations of this study. Here, we performed CGMD simulations using the Martini 2.2 forcefield[43] which involved elastic network restraints[44] within each domain of the open state of Lck in order to maintain their tertiary structures as suggested by experiments (PDB:4D8K, 3LCK). Unlike our ATMD simulations which contained phosphorylated tyrosines (pTyr) 394 and 505 in the open and closed conformations respectively, our CGMD simulations did not contain pTyr residues. Therefore, in the closed state of Lck, we used the elastic network to mimic the interaction between the kinase pY505 residue and the SH2 domain throughout the simulation. This interaction was modelled based on the crystal structure of Hck (PDB:5H0B)[34] which exhibits the highest identity with Lck in the Src kinase family (Fig. S5B). The limitation arising from this model is that we assume that the pY505-SH2 interaction is retained throughout the simulation. We could not observe how this interaction will break and what conformational change it would result in. On the other hand, in the CGMD simulations of the open Lck, we did not include inter-domain elastic network restraints to allow each domain to freely associate or dissociate from each other. Although these simulations were performed using Martini 2.2, the Martini 3 forcefield may help in optimizing some of the interactions observed in this study[45,46]. In our simulations, we observe a preference of Lck to interact with PIP lipids compared to POPS lipids, but due to the lower resolution of PIP lipids in coarse-grained simulations it is challenging to distinguish specificity

between PIP_2 and PIP_3. Free energy calculations may be performed in atomistic resolution to obtain specificity of PIP_2 vs PIP_3[47]. Nonetheless our data is consistent with experimental evidence that Lck has a higher preference to interact with PIP_3 compared to PIP_2[16]. Finally, the SH4-U domain was modelled based on predictions and limited experimental evidence, and conjoined with the rest of the Lck structure in both its open and closed conformations.

Despite these limitations, our study demonstrates that lipid interactions potentially regulate the conformation of Lck and thereby regulate T cell signalling. Furthermore, the Lck-SH2 and kinase domain residues observed as PIP lipid binding sites in this study are conserved among the Src family of kinases. This suggests a pattern in plasma membrane interactions within this family of T cell signalling molecules.

Methods

Molecular modelling. To obtain a model of the 3D structure of Lck-SH4-U, the PSIPRED secondary structure prediction tool[29] along with 3D structure predictions by the I-Tasser[27] and Robetta[28] servers were used. The sequence of the SH4-U domain was obtained from UniprotKB (P06239). Post-translational acylations/lipid tails were added using CHARMM-GUI[48]. Modeller 9.2[33,49] and UCSF Chimera[32] were used to conduct modelling of the open Lck-FL. Homology modelling of the closed Lck-FL was conducted based on available structural data of Hck using Modeller 9.2.

Coarse-grained molecular dynamics (CGMD) simulations. All models were coarse-grained using the Martini 2.2 forcefield[36] and the *martinize* script. To coarse-grain the lipid tails along with the rest of the protein, the *martinize* script, and the Martini 2.2 amino acid topology were modified to include published parameters[50], and made publicly available (https://github.com/DJ004/martini_mod).

CGMD simulations were set up using the *insane* tool[37] and Gromacs 5.0. Elastic network restraints[44] with a 1000 kJ/mol/nm² force constant and 0 to 0.7 nm cut-off distance was applied. However, the restraints were applied only within each domain to maintain their tertiary structure and not between domains to allow unbiased inter-domain interactions. Membrane lipid compositions used to set up each CGMD simulation are shown in Table 1. In all CGMD simulations, each lipid contained one saturated acyl chain and one mono-unsaturated acyl chain, while their headgroup composition is based on the composition of TCR-CD3 activation domains in the T cell plasma membrane[26]. The solvent was neutralized with 0.15 M Na^+ and Cl^- ions. All systems were energy minimized using the steepest descent algorithm until the maximum force converged to 1000 kJ/mol/nm² and equilibrated for 2.5 ns with the protein position-restrained. The equilibrated system was then used to generate differing initial velocities for twenty production simulations run for 5 μs each with a 20 fs time-step. The NPT ensemble was used to conduct equilibration and production simulations. Co-ordinates were saved at 200 ps intervals. A semi-isotropic Parrinello-Rahman barostat (1 bar)[51] and V-rescale thermostat (323 K)[52] were used for production simulations along with a 3×10^{-4}/bar compressibility.

Atomistic molecular dynamics (ATMD) simulations. CHARMM-GUI[48] was used with the CHARMM36 forcefield[53] to setup ATMD simulations of the initial Lck-SH4-U model, the Lck-FL open model and closed model in solution using the TIP3 water as solvent neutralized with 0.15 M Na^+ and Cl^- ions. All systems were energy minimized using the steepest descent algorithm using Gromacs 2016 until the maximum force converged to 1000 kJ/mol/nm², followed by isotropic (NPT) equilibration at 323 K where the protein backbone was position-restrained. The equilibrated system was used to generate differing initial velocities for three production simulations run 250 ns each using a 2 fs time-step. Co-ordinates were saved at 40 ps intervals. The V-rescale thermostat (323 K) was used[52]. Parrinello-Rahman isotropic barostat (1 bar)[51] was used with a compressibility of 4.5×10^{-5}/bar. The LINCS algorithm[54] applied constraints on hydrogen bond lengths and the Particle Mesh Ewald algorithm[55]. Coulombic and van der Waals interactions were defined by a 1.2 nm distance cut-off.

Data analysis and visualization. Protein-lipid and protein–protein interactions in all simulations were calculated using the *gmx mindist* command where a contact was defined by a 0.55 nm distance cut-off in CGMD simulations and 0.4 nm distance cut-off in ATMD simulations. All contact analyses results represent merged data from all simulation replicates. Clustering analyses used the *gmx cluster* command and the gromos method[56] with an RMSD cut-off of 0.35 nm. For this, all trajectories were concatenated using *gmx trjcat*, the protein was extracted using *gmx trjconv* and RMSD calculations were run skipping 5 frames for both CGMD and ATMD simulations. Distance versus time and radial distribution function calculations were done using the *gmx distance* and *gmx rdf* commands respectively. The membrane orientation data of the Lck-SH2 domain was obtained from *gmx distance* (Z axis data) and *gmx rotmat* combined. This was then processed and plotted using an in-house python script. VMD was used for visualization and rendering[57]. The APBS[39] plugin of PyMOL 2.4 (pymol.org) was used to calculate electrostatics. Xmgrace (https://plasma-gate.weizmann.ac.il/Grace/) and Matplotlib 3.3 (https://doi.org/10.5281/zenodo.3948793) were used for plotting.

Data availability

The molecular dynamics simulation datasets generated in the current study are available in the Research Data Leeds Repository, https://archive.researchdata.leeds.ac.uk/968/.

References

1. Courtney, A. H., Lo, W. L. & Weiss, A. TCR Signaling: Mechanisms of initiation and propagation. *Trends Biochem. Sci.* **43**, 108–123. https://doi.org/10.1016/j.tibs.2017.11.008 (2018).
2. Mariuzza, R. A., Agnihotri, P. & Orban, J. The structural basis of T-cell receptor (TCR) activation: An enduring enigma. *J. Biol. Chem.* **295**, 914–925. https://doi.org/10.1074/jbc.REV119.009411 (2019).
3. He, Y. *et al.* Peptide–MHC binding reveals conserved allosteric sites in MHC class I- and class II-restricted T cell receptors (TCRs). *J. Mol. Biol.* **432**, 166697. https://doi.org/10.1016/j.jmb.2020.10.031 (2020).
4. Lanz, A.-L. *et al.* Allosteric activation of T cell antigen receptor signaling by quaternary structure relaxation. *Cell Rep.* **36**, 109375. https://doi.org/10.1016/j.celrep.2021.109375 (2021).
5. Nika, K. *et al.* Constitutively active lck kinase in T cells drives antigen receptor signal transduction. *Immunity* **32**, 766–777. https://doi.org/10.1016/j.immuni.2010.05.011 (2010).
6. Rossy, J., Owen, D. M., Williamson, D. J., Yang, Z. & Gaus, K. Conformational states of the kinase Lck regulate clustering in early T cell signaling. *Nat. Immunol.* **14**, 82–89. https://doi.org/10.1038/ni.2488 (2013).
7. Hatada, M. H. *et al.* Molecular basis for interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor. *Nature* **377**, 32–38. https://doi.org/10.1038/377032a0 (1995).
8. Katz, Z. B., Novotná, L., Blount, A. & Lillemeier, B. F. A cycle of Zap70 kinase activation and release from the TCR amplifies and disperses antigenic stimuli. *Nat. Immunol.* **18**, 86–95. https://doi.org/10.1038/ni.3631 (2017).
9. Palacios, E. H. & Weiss, A. Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation. *Oncogene* **23**, 7990–8000. https://doi.org/10.1038/sj.onc.1208074 (2004).
10. Yamaguchi, H. & Hendrickson, W. A. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature* **384**, 484. https://doi.org/10.1038/384484a0 (1996).
11. Briese, L. & Willbold, D. Structure determination of human Lck unique and SH3 domains by nuclear magnetic resonance spectroscopy. *BMC Struct. Biol.* **3**, 1–7. https://doi.org/10.1186/1472-6807-3-3 (2003).
12. Wingfield, P. N-terminal methionine processing. *Curr. Protoc. Protein Sci.* https://doi.org/10.1002/CPPS.29 (2017).
13. Udenwobele, D. I. *et al.* Myristoylation: An important protein modification in the immune response. *Front. Immunol.* https://doi.org/10.3389/fimmu.2017.00751 (2017).
14. Yurchak, L. K. & Sefton, B. M. Palmitoylation of either Cys-3 or Cys-5 is required for the biological activity of the Lck tyrosine protein kinase. *Mol. Cell Biol.* **15**, 6914–6922. https://doi.org/10.1128/MCB.15.12.6914 (1995).
15. Md, R. Myristylation and palmitoylation of Src family members: the fats of the matter. *Cell* **76**, 411–413. https://doi.org/10.1016/0092-8674(94)90104-X (1994).
16. Sheng, R. *et al.* Lipids regulate Lck protein activity through their interactions with the Lck Src homology 2 domain. *J. Biol. Chem.* **291**, 17639–17650. https://doi.org/10.1074/jbc.M116.720284 (2016).
17. Park, M. J. *et al.* SH2 domains serve as lipid-binding modules for pTyr-signaling proteins. *Mol. Cell* **62**, 7–20. https://doi.org/10.1016/j.molcel.2016.01.027 (2016).
18. Prakaash, D., Cook, G. P., Acuto, O. & Kalli, A. C. Multi-scale simulations of the T cell receptor reveal its lipid interactions, dynamics and the arrangement of its cytoplasmic region. *PLoS Comput. Biol.* **17**, e1009232. https://doi.org/10.1371/journal.pcbi.1009232 (2021).
19. Porciello, N. *et al.* A role of Lck annular lipids in the steady upkeep of active Lck in T cells. *bioRxiv* https://doi.org/10.1101/2022.03.18.484902 (2022).
20. Le Huray, K. I. P., Wang, H., Sobott, F. & Kalli, A. C. Systematic simulation of the interactions of pleckstrin homology domains with membranes. *Sci. Adv.* **8**, eabn6992. https://doi.org/10.1126/sciadv.abn6992 (2022).
21. Yamamoto, E., Kalli, A. C., Yasuoka, K. & Sansom, M. S. P. Interactions of pleckstrin homology domains with membranes: Adding back the bilayer via high-throughput molecular dynamics. *Structure* **24**, 1421–1431. https://doi.org/10.1016/j.str.2016.06.002 (2016).
22. Le Huray, K. I. P., Bunney, T. D., Pinotsis, N., Kalli, A. C. & Katan, M. Characterization of the membrane interactions of phospholipase Cγ reveals key features of the active enzyme. *Sci. Adv.* **8**, eabp9688. https://doi.org/10.1126/sciadv.abp9688 (2022).
23. Kalli, A. C., Campbell, I. D. & Sansom, M. S. P. Conformational changes in talin on binding to anionic phospholipid membranes facilitate signaling by integrin transmembrane helices. *PLoS Comput. Biol.* **9**, e1003316. https://doi.org/10.1371/journal.pcbi.1003316 (2013).
24. Arcario, M. J. & Tajkhorshid, E. Membrane-induced structural rearrangement and identification of a novel membrane anchor in talin F2F3. *Biophys. J.* **107**, 2059–2069. https://doi.org/10.1016/j.bpj.2014.09.022 (2014).
25. Li, Z. L. & Buck, M. Computational modeling reveals that signaling lipids modulate the orientation of K-Ras4A at the membrane reflecting protein topology. *Structure* **25**, 679-689.e2. https://doi.org/10.1016/j.str.2017.02.007 (2017).
26. Zech, T. *et al.* Accumulation of raft lipids in T-cell plasma membrane domains engaged in TCR signalling. *EMBO J.* **28**, 466–476. https://doi.org/10.1038/emboj.2009.6 (2009).
27. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8. https://doi.org/10.1038/NMETH.3213 (2015).
28. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, W526–W531. https://doi.org/10.1093/NAR/GKH468 (2004).
29. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkz297 (2019).
30. Kim, P. W., Sun, Z. Y. J., Blacklow, S. C., Wagner, G. & Eck, M. J. A zinc clasp structure tethers Lck to T cell coreceptors CD4 and CD8. *Science* **301**, 1725–1728. https://doi.org/10.1126/science.1085643 (2003).
31. Li, L. *et al.* Ionic CD3–Lck interaction regulates the initiation of T-cell receptor signaling. *Proc. Natl. Acad. Sci. USA* **114**, E5891–E5899. https://doi.org/10.1073/pnas.1701990114 (2017).
32. Pettersen, E. F. *et al.* UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612. https://doi.org/10.1002/jcc.20084 (2004).
33. Webb, B. & Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* https://doi.org/10.1007/978-1-4939-0366-5_1 (2014).
34. Yuki, H. *et al.* Activity cliff for 7-substituted pyrrolo-pyrimidine inhibitors of HCK explained in terms of predicted basicity of the amine nitrogen. *Bioorg. Med. Chem.* **25**, 4259–4264. https://doi.org/10.1016/J.BMC.2017.05.053 (2017).
35. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145. https://doi.org/10.1002/PRO.3290 (2018).
36. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824. https://doi.org/10.1021/jp071097f (2007).
37. Wassenaar, T. A., Ingólfsson, H. I., Böckmann, R. A., Tieleman, D. P. & Marrink, S. J. Computational lipidomics with *insane*: A versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.* **11**, 2144–2155. https://doi.org/10.1021/acs.jctc.5b00209 (2015).
38. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667. https://doi.org/10.1093/NAR/GKH381 (2004).

39. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041. https://doi.org/10.1073/pnas.181342398 (2001).
40. Rossy, J., Williamson, D. J. & Gaus, K. How does the kinase Lck phosphorylate the T cell receptor? Spatial organization as a regulatory mechanism. *Front. Immunol.* https://doi.org/10.3389/fimmu.2012.00167 (2012).
41. Hilzenrat, G. *et al.* Conformational states control Lck switching between free and confined diffusion modes in T cells. *Biophys J.* **118**, 1489–1501. https://doi.org/10.1016/j.bpj.2020.01.041 (2020).
42. Le Roux, A. L. *et al.* A myristoyl-binding site in the SH3 domain modulates c-Src membrane anchoring. *iScience* **12**, 194–203. https://doi.org/10.1016/J.ISCI.2019.01.010 (2019).
43. De Jong, D. H. *et al.* Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.* **9**, 687–697. https://doi.org/10.1021/ct300646g (2013).
44. Periole, X., Cavalli, M., Marrink, S.-J.J. & Ceruso, M. A. Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* **5**, 2531–2543. https://doi.org/10.1021/ct9002114 (2009).
45. Srinivasan, S., Zoni, V. & Vanni, S. Estimating the accuracy of the MARTINI model towards the investigation of peripheral protein–membrane interactions. *Faraday Discuss.* **232**, 131–148. https://doi.org/10.1039/D0FD00058B (2021).
46. Souza, P. C. T. *et al.* Martini 3: A general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **18**, 382–388. https://doi.org/10.1038/s41592-021-01098-3 (2021).
47. Naughton, F. B., Kalli, A. C. & Sansom, M. S. P. Modes of interaction of pleckstrin homology domains with membranes: Toward a computational biochemistry of membrane recognition. *J. Mol. Biol.* **430**, 372–388. https://doi.org/10.1016/j.jmb.2017.12.011 (2018).
48. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865. https://doi.org/10.1002/jcc.20945 (2008).
49. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* https://doi.org/10.1002/0471140864.ps0209s50 (2007).
50. Atsmon-Raz, Y. & Tieleman, D. P. Parameterization of palmitoylated cysteine, farnesylated cysteine, geranylgeranylated cysteine, and myristoylated glycine for the martini force field. *J. Phys. Chem. B* **121**, 11132–11143. https://doi.org/10.1021/acs.jpcb.7b10175 (2017).
51. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190. https://doi.org/10.1063/1.328693 (1981).
52. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 14101. https://doi.org/10.1063/1.2408420 (2007).
53. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145. https://doi.org/10.1002/jcc.23354 (2013).
54. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3c1463::AID-JCC4%3e3.0.CO;2-H (1997).
55. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593. https://doi.org/10.1063/1.470117 (1995).
56. Daura, X. *et al.* Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.* **38**, 236–240. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2%3c236::AID-ANIE236%3e3.3.CO;2-D (1999).
57. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph* **14**, 33–8. https://doi.org/10.1016/0263-7855(96)00018-5 (1996).

Acknowledgements

Author contributions

D.P. and C.F. performed simulations and analyses, and prepared figures, D.P. wrote the original manuscript. All authors revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-25603-6.

Correspondence and requests for materials should be addressed to A.C.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.