

Supporting Information Appendix

Structural mechanism for Bruton's tyrosine kinase activation at the cell membrane

T1. Molecular Dynamics Simulation Analysis of an Individual Apo PH-TH Module

To our knowledge, there are no crystal structures of an individual PH-TH module, and all PH-TH modules that have been crystallized are part of Saraste dimers. To simulate an individual module, we thus performed molecular dynamics (MD) simulations using the PH-TH module conformation taken from the dimer crystal structure (PDB ID: 1BTK).¹ The trajectories ranged in length from 10 to 100 μ s (see Methods for details). The overall conformation of the PH-TH module at the end of the trajectories is similar to the initial crystal structure (SI Appendix Figure S9A). The average root-mean-square deviation (RMSD) between the crystal structure and instantaneous structures from the trajectories, aligned on the entire PH-TH module, was ~ 3.5 Å (SI Appendix Figure S9B).

The simulations indicate that the Saraste dimer interface is unstable in an individual PH-TH module, as has been previously observed in sub-microsecond MD simulations using a similar simulation setup.² The average RMSD at the Saraste dimer interface was ~ 5.0 Å (SI Appendix Figure S9B). Significant deviations were observed at the $\beta 3$ – $\beta 4$ hairpin and the $\alpha 1$ helix, which constitute the Saraste dimer interface (SI Appendix Figure S9A). Within the first 100 ns of simulation, the $\beta 3$ – $\beta 4$ hairpin and the $\alpha 1$ helix bent towards the core of the PH-TH module. Phe 44 and Tyr 42, located at the center of the Saraste dimer interface, changed their side-chain

conformations and packed against interface residues Ile 92 and Ile 95. This conformational change closed the dimer interface, burying the interface residues. Over the full course of our simulations, the $\alpha 1$ helix sampled different conformations, ranging from folded, helical conformations like those seen in crystal structures to unfolded, extended conformations. This structural change further disrupted the Saraste dimer interface by moving interface residues Ile 92 and Ile 95 away from positions compatible with forming the Saraste dimer (SI Appendix Figure S9A). The closure of the Saraste dimer interface in our simulations did not have any obvious structural effects at the canonical binding site or the peripheral binding site. In addition to the Saraste interface, we also observed significant deviations in the IP₄-binding loop and the loop that connects the PH and TH domains (SI Appendix Figure S9C).

T2. Reproducing Features of IP₄ and IP₆ Binding to an Individual PH-TH Module

We tested the ability of our simulations to capture experimentally known behavior of the PH-TH module by simulating its interactions with the soluble inositol phosphates IP₄ and IP₆. We first performed MD simulations of an individual PH-TH module in the presence of IP₄. IP₄ is known to bind to the canonical binding site of the PH-TH module with ~30 nM affinity, and has no detectable affinity at the peripheral binding site.^{2,3} We started simulations by placing two IP₄ molecules ~10 Å away from the PH-TH module, on the opposite side from the canonical binding site. Within 2 μ s of simulation time, one IP₄ molecule spontaneously diffused to the canonical PIP₃ binding site and bound there until the end of the simulations (SI Appendix Figure S10A). We did not observe IP₄ binding stably in the peripheral binding site in our simulations (SI Appendix Figure S10C).

We next showed that IP₆ could bind at the canonical and the peripheral binding sites of the PH-TH module in our simulations. The affinities of IP₆ at the canonical and peripheral binding sites are known to be 300 nM and 10 μM, respectively.² We started simulations by placing two IP₆ molecules on the opposite side of the PH-TH module from the binding sites, as described above. Within 2 μs of simulation, the two IP₆ molecules spontaneously diffused to the canonical and peripheral binding sites. In contrast to the stable binding of IP₆ in the canonical binding site, we observed multiple binding and unbinding events at the peripheral binding site (SI Appendix Figure S10B), and no one binding pose predominated: A number of distinct binding poses, including the one seen in a crystal structure of an IP₆-bound PH-TH module, were observed (SI Appendix Figures S10B and S10D). The less stable binding of IP₆ in the peripheral binding site is expected given the weaker binding affinity of IP₆ for this site.

The simulations described above reproduced essential experimentally determined features of PH-TH module binding to soluble inositol phosphates. We note that the interactions with the inositol phosphatases appear somewhat weaker in the simulations (presumably due to force field shortcomings), but that the simulations correctly identify both binding sites and correctly order the strengths of interactions.

T3. The Saraste Dimer Interface Is Not Stable in Solution-Based Simulations

Saraste dimer stability was much lower in solution than on the membrane in our simulations. In conventional MD simulations, Saraste dimers in solution underwent a transition from the crystal structure-like conformation to the pre-Saraste dimer conformation within 10 μs (SI Appendix Figure S5A). In some simulations, the interface of the pre-Saraste dimer further deteriorated,

forming a cluster of similar metastable dimer structures which we term the *partially broken* Saraste dimer conformation (SI Appendix Figures S5B and S5C). In these partially broken dimers, the hydrophobic packing interactions between key interface residues (Phe 44 of one module and Ile 92 and Ile 95 of the other), characteristic of the membrane-bound Saraste dimer, were broken (SI Appendix Figure S5C). This led to an increased distance between the modules and produced a $\sim 20^\circ$ angle between the two $\beta 3$ – $\beta 4$ hairpins (SI Appendix Figure S5B). These partially broken Saraste dimer conformations, which frequently appeared in our simulations of solution-based dimers, contributed to the large (>4.5 Å) RMSD values at the Saraste interface observed for the solution-based dimers. By comparison, the membrane-bound Saraste dimers had interface RMSDs in the range of 2.0 Å–4.5 Å.

T4. The Membrane Colocalization Effect Is Not Sufficient to Explain the Frequent Formation of Saraste Dimers on the Membrane

Although we did not observe the formation of Saraste dimers in solution-based simulations, we frequently observed their formation on the membrane. One possible explanation for this difference is that the local concentration of PH-TH modules is higher when anchored to the membrane than in solution.⁴ If the higher frequency of Saraste dimer formation arises solely from this membrane colocalization effect, attaching the PH-TH modules to the membrane should always promote Saraste dimer formation, regardless of the method by which the modules are attached. Our investigation has thus far focused on the PH-TH module attaching to the membrane by binding PIP₃, but there are many alternative membrane-attachment strategies: Small G proteins, such as Kirsten rat sarcoma viral oncogene homolog (KRAS), attach to the membrane through covalent lipid modification.^{5,6} The C terminus of KRAS is farnesylated at its cysteine residue in the cytosol, allowing KRAS to attach to the cell membrane.^{5,6}

To test whether non-PIP₃-dependent membrane-attachment methods would also promote Saraste dimer formation, we generated a PH-TH-KRAS fusion protein, in which the PH-TH module of Btk is fused to the farnesylated C-terminal tail of KRAS (SI Appendix Figure S11A). We used tempered binding simulations to study the encounter of two PH-TH-KRAS fusion proteins on a piece of membrane containing only POPC lipids. The effective concentration of the fusion protein in these simulations was comparable to that of the wild-type PH-TH module in our tempered binding simulations that included PIP₃-containing membrane. Nevertheless, the Saraste dimer did not form over the course of an aggregated ~500 μ s of simulation time under various tempering conditions, which indicates that the increased protein concentration caused by membrane colocalization is not sufficient to explain the frequent formation of Saraste dimers on the PIP₃-containing membranes.

T5. PIP₃ Binding at the Canonical and Peripheral Sites Orients the PH-TH Module in a Way that Favors the Formation of the Saraste Dimer

We analyzed the instantaneous MD structures in the encounter simulations of the PH-TH-KRAS fusion proteins, and found that although membrane attachment indeed brought the two proteins into close proximity, they were still able to sample a variety of relative orientations comparable to the range they sampled in our solution-based simulations of wild-type PH-TH modules. A number of structural factors enabled this flexibility, including the fact that there were no stable interactions between the PH-TH module of the fusion protein and the membrane surface, and the fact that the linker between the farnesylated residue and the last residue of the PH-TH module was flexible and did not significantly restrain the protein's orientation. Furthermore, we found that some of the transiently stable dimers that formed in solution-based simulations of wild-type PH-TH modules, and some of the dimers of the PH-TH-KRAS fusion proteins, would not have

been structurally possible if both PH-TH modules were anchored to the membrane by PIP₃ at both the canonical and peripheral sites.

In the fully formed, membrane-bound Saraste dimer, the long axis of the $\alpha 2$ helix of each PH-TH module is parallel to the plane of the membrane. In simulations of an individual PH-TH module anchored by a single PIP₃ molecule, bound at the canonical site, this angle was restricted to 0°–45°. In simulations with multiple PIP₃ molecules, the peripheral site was also occupied, and the angle between the $\alpha 2$ helix of the PH-TH module and the membrane surface was restricted further, to 0°–25°. The two-PIP₃-anchored PH-TH modules thus only sampled orientations that are very close to the one that is compatible with Saraste dimer formation. These simulations indicate that PIP₃ binding to PH-TH modules makes dimerization into the Saraste interface more likely by restricting the relative orientations the modules can assume to ones that favor dimer formation (SI Appendix Figure S11B).

T6. Binding of PIP₂ in the Peripheral and Bridging Sites

We performed additional landing simulations using dimerized PH-TH modules and a piece of membrane that contained 94% PC and 6% PIP₂. The PH-TH modules landed on the membrane with the canonical binding sites occupied by PIP₂ lipids, despite their much less stable binding poses compared to those of PIP₃. Even at 6% PIP₂, the occupancy rates of the bridging sites and peripheral sites were only <1% and <10%, respectively (by comparison, the occupancy rate of the bridging sites and peripheral sites by PIP₃ were 30% and ~100%, respectively, on membranes containing 94% PC and 6% PIP₃). The lower stability of PIP₂ binding is consistent with our structural observation that for the representative binding poses of PIP₃, having an additional phosphate at position 3 of the inositol ring increases binding stability with Btk at both the

peripheral and bridging sites. This is similar to the binding of PIP at the canonical binding site, where PIP₃ is a stronger binder than PIP₂

Methods

General simulation details

All production simulations were run on Anton,⁷ a specialized machine for MD simulations, using the Amber ff99SB-ILDN force field for proteins,^{8–10} the CHARMM TIP3P model for water,¹¹ and the CHARMM36 force field for POPC lipids.¹² The initial parameters of soluble inositol phosphates and PIP₃ lipids were generated with the generalized Amber force field (GAFF),¹³ with the atomic charges refit by restrained electrostatic potential fitting.

The initial protein structure used for simulations of the wild-type PH-TH module was obtained from the crystal structure of apo Btk (PDB ID: 1BTK).¹ The mutation Cys 28 was changed back to Arg using PyMOL.¹⁴ The missing side chains were automatically built using the program Maestro.¹⁵ The N terminus and the C terminus, as well as the terminus of the missing loop in the structure, were capped and neutral. Cys 154, Cys 155, and Cys 165, which coordinate Zn²⁺ in the TH domain, were set to be deprotonated. No positional restraints were, however, applied to the Zn²⁺ ion, and its sp³ tetrahedral coordination remained intact during all MD simulations.

The loop containing residues 80–88 is missing in the crystal structure of Btk (PDB ID: 1BTK),¹ and for our simulations we manually added this missing loop using Coot.¹⁶ In our simulations, loop residues Arg 81 and Arg 82 frequently interacted with the membrane surface. We were initially concerned that these interactions might influence encounter simulation efficiency, so we

deleted residues Pro 80, Arg 81, and Arg 82 and performed the rest of our simulations using this three-residue-deleted construct. Later on, we observed that sampling efficiency is not, in fact, sensitive to the identity of this loop, since we saw dimer formation whether or not the loop was included (SI Appendix Table S1: simulations 3–6).

A typical simulation box with and without membrane contained 151,924 atoms ($115 \text{ \AA} \times 115 \text{ \AA} \times 115 \text{ \AA}$) and 87,650 atoms ($90 \text{ \AA} \times 90 \text{ \AA} \times 100 \text{ \AA}$), respectively. The system was neutralized and salted with NaCl, with a final concentration of 0.15 M. The system was equilibrated in the NPT ensemble for 100 ns using gDesmond¹⁷ on a commodity GPU cluster. Production runs were subsequently performed in the NVT ensemble from the final frame of the NPT relaxation simulation¹⁸ by coupling the system to a variant¹⁹ of the Nosé-Hoover thermostat²⁰ at 310 K with a relaxation time of 1 ps. A RESPA²¹ integrator was used, with a time step of 2.5 fs. The long-range electrostatic forces were calculated in k-space using a grid-based method with Gaussian spreading to the grid every 7.5 fs.²²

Conventional MD simulations of inositol phosphates binding to the PH-TH module

We placed two copies of, depending on the simulation, IP₄ or IP₆ molecules $\sim 10 \text{ \AA}$ away from the PH-TH module and on the opposite side from the canonical and peripheral binding sites. We ran three 10- μ s trajectories for each system, varying the initial velocity assignment on the atoms. Binding was measured using the minimum distance between any heavy atom in the IP₄ or IP₆ molecule and any heavy atom in the residues that constitute the canonical binding site (Arg 28, Tyr 39, Asn 24, Lys 12) and the peripheral binding site (Arg 49, Lys 52, Tyr 40, Lys 36).

Membrane equilibration

We replaced the POPC lipids from one leaflet of a 256-POPC-containing membrane with PIP₃ lipids to create the PIP₃-containing membranes. Simulations with membranes were performed under a semi-isotropic pressure condition. PIP₃ molecules diffused freely in simulations. Dimers of PIP₃ lipids, which are mediated by the hydrogen bonds between their inositol rings, occasionally appeared, but we observed no large clusters of PIP₃ in our simulations. Membranes with different numbers of PIP₃ lipids were equilibrated for 10 μ s, and the last frames of the simulations were extracted and used as the starting points of the simulations of PH-TH modules that included membrane.

Conventional MD membrane-binding simulations

We equilibrated a crystal structure of the PH-TH module in aqueous solution for 1 μ s, and extracted the last frame of the trajectory as the starting conformation for the membrane-binding simulations. We then placed the PH-TH module ~ 10 Å away from a membrane containing eight PIP₃ molecules in one of the leaflets. The system was then neutralized and salted by NaCl, with a final concentration of 0.15 M. Simulations with membranes containing two and four PIP₃ lipids were set up similarly. Simulations were performed in triplicate, with different initial velocities. The length of trajectories varied between 10 μ s and 100 μ s. The binding of PIP₃ was measured using the minimal distance between any atom in the head group of PIP₃ and any atom in the residues that constitute the canonical binding site (Arg 28, Tyr 39, Asn 24, Lys 12) and the peripheral binding site (Arg 49, Lys 52, Tyr 40, Lys 36, Ser 55).

Tempered binding simulations

We prepared our simulations by placing two PH-TH modules in solution with the crystallographic dimer interface of each module facing away from the other, with ~ 10 Å between the surfaces of the modules. We then performed 10 conventional MD simulations of 2 μ s each. We extracted the last frame of each trajectory, and only used those in which both PH-TH modules had landed on the membrane, with a PIP3 molecule bound to the canonical site of at least one of the modules. We then assigned new velocities to the system based on the temperature of simulation (310 K) and used these frames to start a new set of tempered binding simulations.

In these tempered binding simulations of the membrane-bound PH-TH modules, the TH domain can dissociate from the PH domain and undergo partial unfolding on long timescales (it was only observed to happen after 50 μ s of simulation time). We observed this in two trajectories out of 20. Because no experimental data suggest that this happens in nature, and because the PH and TH domains are known experimentally to be an integrated module, we concluded that this phenomenon was an artefact of our simulations. We thus applied torsional backbone corrections:

$$U = k \sum_{m=1}^M (-1)^{m-1} \left[1 + \cos\left(\frac{m(\varphi - \varphi')}{m!}\right) \right]$$

(centered at $\varphi' = \varphi_{\text{xtal}} - 180^\circ$; $M = 6$ and $k = 0.01$ – 1.0 kcal mol $^{-1}$) to the protein in encounter simulations on the membrane. We found that sampling efficiency in the encounter simulations is not sensitive to the backbone correction for the strengths we used (from 0.01– 1.0 kcal mol $^{-1}$) (SI Appendix Table S1).

To enhance the sampling in our PH-TH module encounter simulations, we used an approach called “tempered binding,”²³ which tempered the system’s Hamiltonian (energy function) as opposed to its temperature, as is done in conventional tempering methods. Tempered binding dynamically scales various atomic interactions during an MD simulation by a factor, λ , that is updated among a ladder of discrete values, λ_i . (Our tempered binding simulations had 120 rungs, subdivided into 3 overlapping regions of 60 rungs each.) The atomic interactions are unscaled at the lowest rung of the ladder (rung 0). The structures sampled at rung 0 are sampled from the same Boltzmann distribution as a conventional MD simulation (with the advantage that the tempered binding simulations more rapidly explore conformational space); we thus sometimes, where indicated, show data for just the portion of the simulation trajectories when the Hamiltonian is in this “unscaled” rung.

Our tempered binding simulations linearly scaled the near electrostatic interactions between the positively charged atoms on one protein and the negatively charged atoms on the other protein, and vice versa (interactions within a protein, and between protein and water, were not scaled). We varied the maximum tempering strength (i.e., the scale factor reached at the highest rung; the scale factor at the lowest rung is 1) from 0.991 to 0.997, and found that the efficiency of dimer formation was not sensitive to the maximum tempering strength in this range. Further decreasing the maximum tempering strength to 0.980 resulted in no dimers of any kind forming during simulations.

Clustering of dimers formed on membranes

All frames from all replicas of 24 runs (for a total of 6.55 ms of simulation time) were extracted. For each frame, the structure was described by a 338-dimension vector of binary values,

representing each residue in the dimer. Each value in the vector was set to 1 if any atom in the corresponding residue (for a given module) was within 3 Å of any atom of the other module. The residues that were not in contact with the other module had their values set to 0. In total, there were 5,571,555 such vectors. Vectors in which there was no dimer structure (a vector of all zeros) were removed. The filtered data set contained $N = 4,379,672$ vectors.

The set of vectors was normalized by bringing each dimension of the vectors to a mean of 0 and a standard deviation of 1, independently of other dimensions. The 10 top principal components of the normalized matrix were calculated using the randomized singular value decomposition (SVD) method. The projected matrix was subsampled at every 10th vector to generate a 437968×10 matrix. These 10-dimensional vectors were clustered using hierarchical density-based spatial clustering of applications with noise (HDBSCAN).²⁴ The minimum spanning tree was computed with Boruvka's Kd-trees algorithm,²⁵ with a minimum cluster size of 3,000 and a minimum core neighborhood size of 300. The entire $N \times 10$ matrix was then assigned cluster labels through an approximate fit of the subsampled clusters. The labels were ordered by cluster size.

A total of 5 clusters were identified, with 1,132,602 (26%) of the points analyzed categorized as noise. The primary cluster contained 2,679,781 points (61%). The second cluster was reclustered by rerunning the principal component analysis on the normalized $NS \times 338$ vectors followed by hierarchical DBSCAN with the same parameters. This reclustered resulted in more pronounced separation of a previously heterogeneous mixture of conformations. For each major cluster, an "ideal" representative 338-dimensional binary vector was calculated by setting 1 at positions where more than 50% of the vectors in the cluster had 1s, and setting 0s elsewhere. The cluster member that gave the smallest L1 distance to the "ideal" vector was taken as the best

approximate representation of the cluster. The structures for each cluster's best approximate representation vector were identified and visualized.

Equilibrium simulations between the Saraste conformation and the pre-Saraste conformation

We started these simulations by placing the crystal structure of the Saraste dimer in solution, ~ 5 Å from a membrane surface with 6% PIP₃ content. We first equilibrated the system by running conventional MD simulations for 2 μ s to let the Saraste dimers land on the membrane. We then extracted the last frames from the 2 μ s simulations, assigning new velocities to the system based on the temperature (310 K), and started the tempered binding simulation. The maximum tempering strength was 0.995, and the torsional backbone correction strength was 1.0 kcal mol⁻¹. All frames at rung 0 were extracted, the extracted RMSD values per frame were binned by 0.1 Å, and the distribution curve was smoothed using a Gaussian kernel density estimator to produce the histogram shown in SI Appendix Figures S4 and S6.

The population fractions of the tightly packed Saraste conformation and of the pre-Saraste conformation were calculated as follows: First, the RMSD values for all frames at rung 0 were calculated with respect to the crystal structure, as in the other analyses. Only frames with RMSD < 7 Å were considered. We classified frames with RMSD < 3.1 Å as being in the tightly packed conformation, and frames with RMSD > 3.1 (and < 7) as being in the pre-Saraste conformation. The percentage of the population for each conformation was then calculated as the fraction of the total frames associated with that conformation.

Calculation of the free energy transition between the Saraste and pre-Saraste conformations

Simulation trajectories with more than 10 transition events between the Saraste and pre-Saraste conformations were used for the free energy calculation, using the equation $\Delta G =$

$-RT \ln \frac{N_{pre-Saraste}}{N_{Saraste}}$. The final free energy value reported was averaged from 5 (for the 8-PIP₃

condition) and 3 (for the 2-PIP₃ condition) independent simulations. The total numbers of transition events for the 8-PIP₃ and 2-PIP₃ conditions were 86 and 81, respectively.

Bioinformatics analysis

The Btk sequences used for evolution analysis were compiled using a series of protein-protein BLAST searches in the NCBI non-redundant protein database, using the human Btk PH-TH module (residues 1–172) as the reference.^{26,27} The initial search generated 772 sequences with a minimal sequence identity of 48%. Btk sequences of mammals, birds, and fish were then extracted from the initial search results based on taxonomy IDs. The sequence pool was then filtered based on the following rules: 1) duplicate entries were removed; 2) sequences labeled as “partial” or “synthetic” were removed; 3) sequences annotated as “BMX kinase” were removed. The final number of sequences used for multiple-sequence alignment was 334. The multiple-sequence alignment of the 334 sequences was performed in Clustal Omega^{28,29} with default settings, and sequence logos were generated using the online tool WebLogo3.^{30,31} The final figures were manually adjusted from the WebLogo3 result to show only the residues at the sites of interest.

Supporting References

1. Hyvönen, M. & Saraste, M. Structure of the PH domain and Btk motif from Bruton's tyrosine kinase: Molecular explanations for X-linked agammaglobulinaemia. *EMBO J.* **16(12)**, 3396–3404 (1997).
2. Wang, Q., Vogan, E.M., Nocka, L.M., Rosen, C.E., Zorn, J.A., Harrison, S.C., and Kuriyan, J. (2015). Autoinhibition of Bruton's tyrosine kinase (Btk) and activation by soluble inositol hexakisphosphate. *eLife* **4**, e06074.
3. Baraldi, E., Carugo, K.D., Hyvönen, M., Surdo, P.L, Riley, A.M., Potter, B.V., O'Brien, R., Ladbury, J.E., and Saraste, M. (1999). Structure of the PH domain from Bruton's tyrosine kinase in complex with inositol 1,3,4,5-tetrakisphosphate. *Structure* **7(4)**, 449–460.
4. Groves, J. T. & Kuriyan, J. Molecular mechanisms in signal transduction at the membrane. *Nat. Struct. Mol. Biol.* **17(6)**, 659–665 (2010).
5. Zhang, F. L. & Casey, P. J. Protein prenylation: molecular mechanisms and functional consequences. *Annu. Rev. Biochem.* **65(1)**, 241–269 (1996).
6. Casey, P. J., Solski, P. A., Der, C. J. & Buss, J. E. p21ras is modified by a farnesyl isoprenoid. *Proc. Natl. Acad. Sci. U. S. A.* **86(21)**, 8323–8327 (1989).
7. Shaw, D. E., Grossman, J.P., Bank, J.A., Batson, B., Butts, J.A., Chao, J.C., Deneroff, M.M., Dror, R.O., Even, A., Fenton, C.H., Forte, A., Gagliardo, J., Gill, G., Greskamp, B., Ho, C.R., Ierardi, D.J., Iserovich, L., Kuskin, J.S., Larson, R.H., Layman, T., Lee, L.-S., Lerer, A.K., Li, C., Killebrew, D., Mackenzie, K.M., Mok, S.Y.-H., Moraes, M.A., Mueller, R., Nociolo, L.J., Peticolas, J.L., Quan, T., Ramot, D., Salmon, J.K., Scarpazza, D.P., Schafer, U.B., Siddique, N., Snyder, C.W., Spengler, J., Tang, P.T.P., Theobald, M., Toma, H., Towles, B., Vitale, B., Wang, S.C., and Young, C. (2014). Raising the bar for performance and programmability in a special-purpose molecular dynamics

supercomputer. In *Proc. International Conference for High Performance Computing, Networking, Storage and Analysis (SC14)* New York, NY: IEEE.

8. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amberff99SB protein force field. *Proteins* **78(8)**, 1950–1958.
9. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65(3)**, 712–725.
10. Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074.
11. MacKerell, A. D. Jr., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102(18)**, 3586–3616.
12. Klauda, J. B., Venable, R.M., Freites, J.A., O'Connor, J.W., and Tobias, D.J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A.D., Jr., and Pastor, R.W. (2010). Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *J. Phys. Chem. B* **114(23)**, 7830–7843.
13. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general Amber force field. *J. Comp. Chem.* **25(9)**, 1157–1174 (2004).

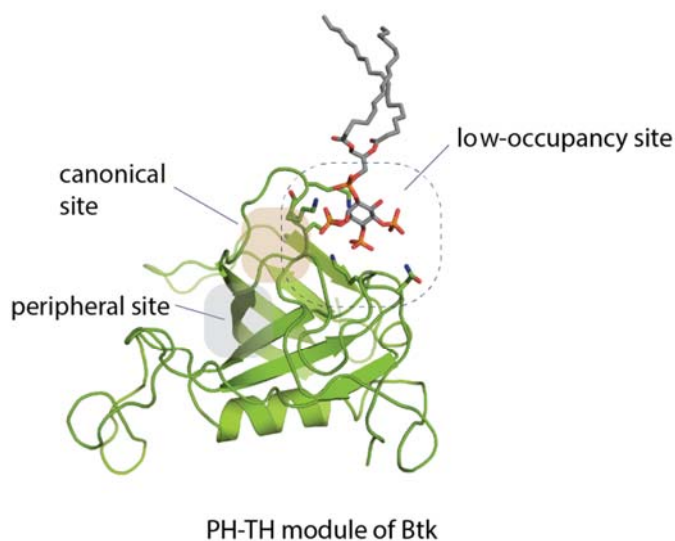
14. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.
<http://www.pymol.org>.
15. Maestro, Version 11 Schrödinger, LLC. <http://www.schrodinger.com/maestro>.
16. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta. Crystallogr. D: Biol. Crystallogr.* **66(4)**, 486–501 (2010).
17. Bowers, K. J., Chow, E., Xu, H., Dror, R.O., Eastwood, M.P., Gregersen, B.A., Klepeis, J.L., Kolossváry, I., Moraes, M.A., Sacerdoti, F.D., Salmon, J.K., Shan, Y., and Shaw, D.E. (2006). Scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proc. 2006 ACM/IEEE conference on Supercomputing*, 84
<https://doi.org/10.1145/1188455.1188544>.
18. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101(5)**, 4177–4189 (1994).
19. Lippert RA, Predescu C, Ierardi DJ, Mackenzie KM, Eastwood MP, Dror RO, Shaw DE. (2013). Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. *J Chem Phys* **139(16)**, 164106.
20. Hoover, W.G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A.* 1985, 31(3), 1695–1697.
21. Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97(3)**, 1990–2001 (1992).
22. Shan, Y., Klepeis, J. L., Eastwood, M. P., Dror, R. O. & Shaw, D. E. Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *J. Chem. Phys.* **122(5)**, 054101 (2005).

23. Pan, A. C., Jacobson, D., Yatsenko, K., Sritharan, D., Weinreich, T. M. & Shaw, D. E. Atomic-level characterization of protein-protein association. Preprint at <https://www.biorxiv.org/content/early/2018/04/17/303370> (2018).
24. Campello, R. J. G. B., Moulavi, D. & Sander, J. Density-based clustering based on hierarchical density estimates. In *PAKDD 2013: Advances in Knowledge Discovery and Data Mining* 160–172 https://doi.org/10.1007/978-3-642-37456-2_14 (2013).
25. March, W. B., Ram, P. & Gray, A. G. Fast euclidean minimum spanning tree: Algorithm, analysis, and applications. In *Proc. 16th ACM SIGKDD International Conf. Knowledge Discovery and Data Mining* 603–612 <https://doi.org/10.1145/1835804.1835882> (2010).
26. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215(3)**, 403–410 (1990).
27. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33(suppl. 1)**, D501–D504 (2005).
28. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., and Higgins, D.G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7(539)**, 539.
29. Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38(suppl. 2)**, W695–699.
30. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14(6)**, 1188–1190 (2004).

31. Schneider, T. D. & Stephens, R.M. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* **18(20)**, 6097–6100 (1990).
32. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372(3)**, 774–797 (2007).

Supporting Figures

A



B

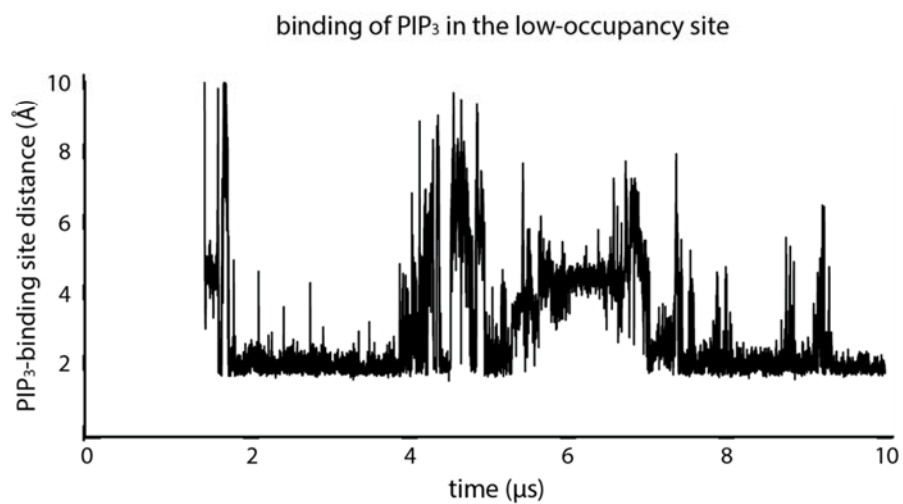


Figure S1. PIP₃ binding at a third, low-occupancy binding site on an individual PH-TH module.

A. An instantaneous structure from a conventional MD simulation ($t = 10 \mu\text{s}$) showing the binding of a PIP₃ molecule at a third, low-occupancy binding site. B. Minimum distance between atoms of the low-occupancy site and PIP₃ atoms in a membrane-binding trajectory.

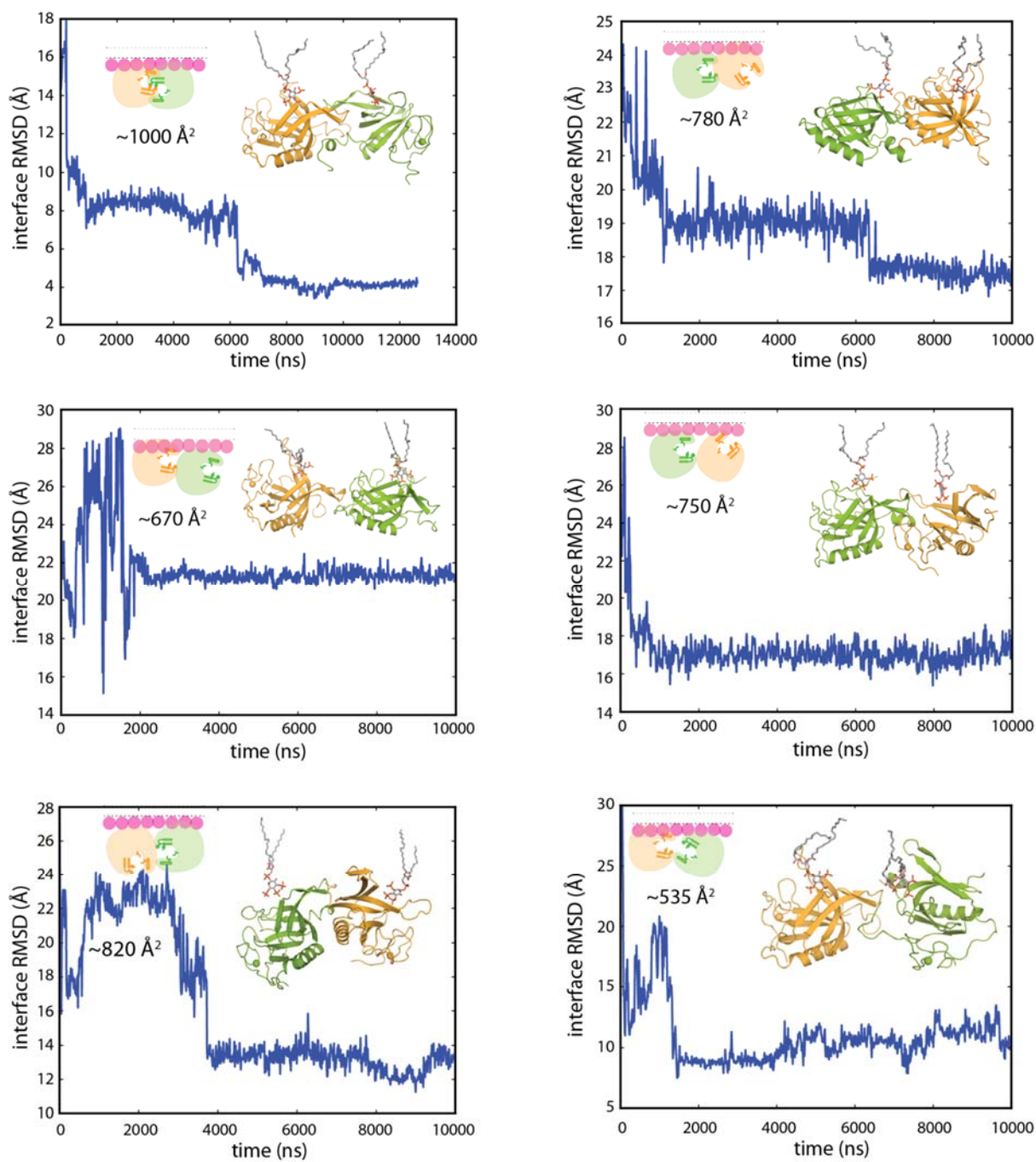


Figure S2. Representative trajectories of membrane-bound PH-TH dimers that formed in conventional MD simulations. The average interface area per module is calculated using the web server PISA.³²

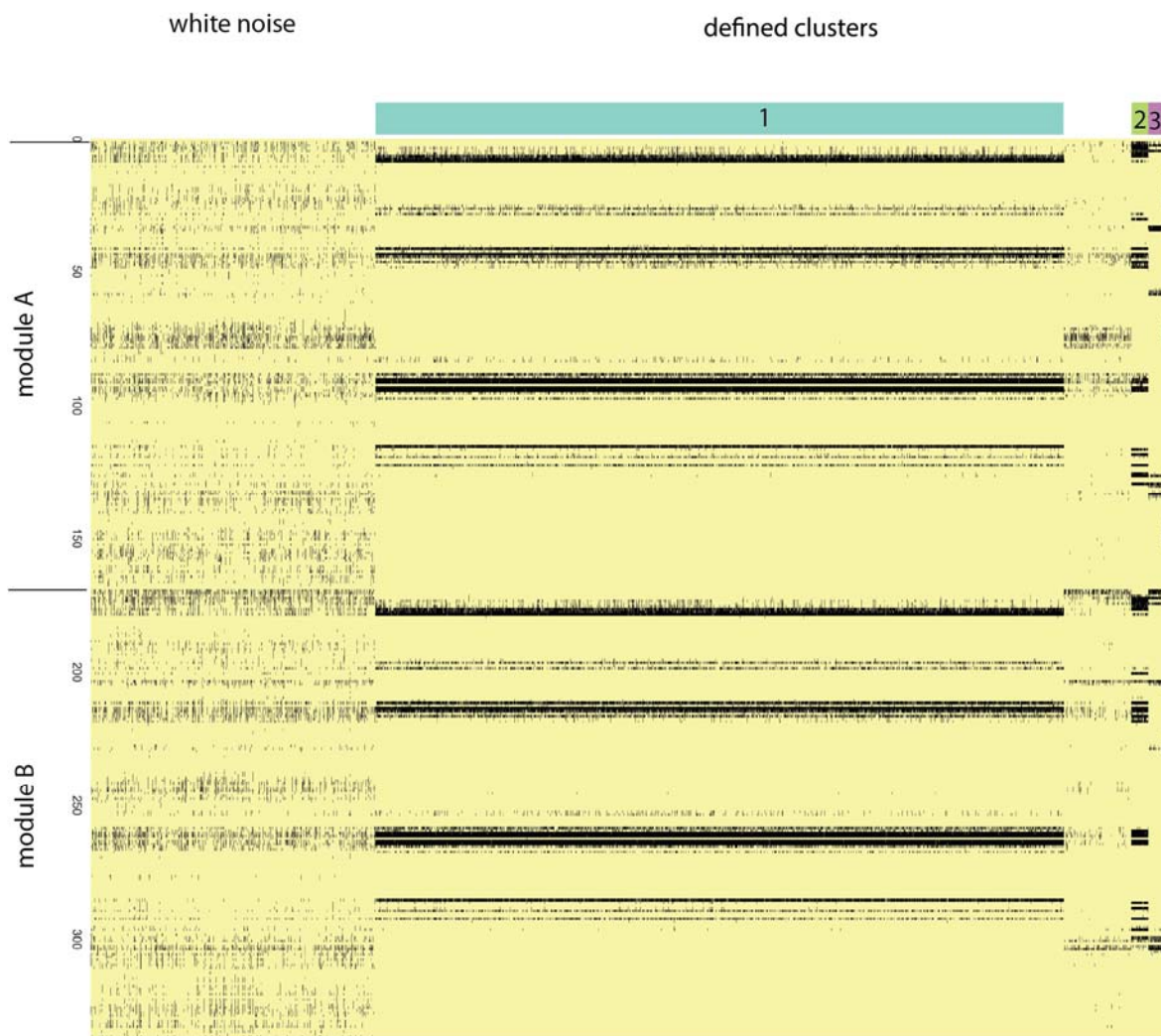


Figure S3. Cluster analysis of PH-TH dimers that formed on the membrane. Hierarchical DBSCAN clustering of PH-TH dimers from all replicas of 24 runs (a total of 6.55 ms simulation time). Five clusters were identified, with 1,132,602 (26%) of the points analyzed categorized as noise. The primary cluster contained 2,679,781 points (61%). The labels were ordered by cluster size (see Methods for details).

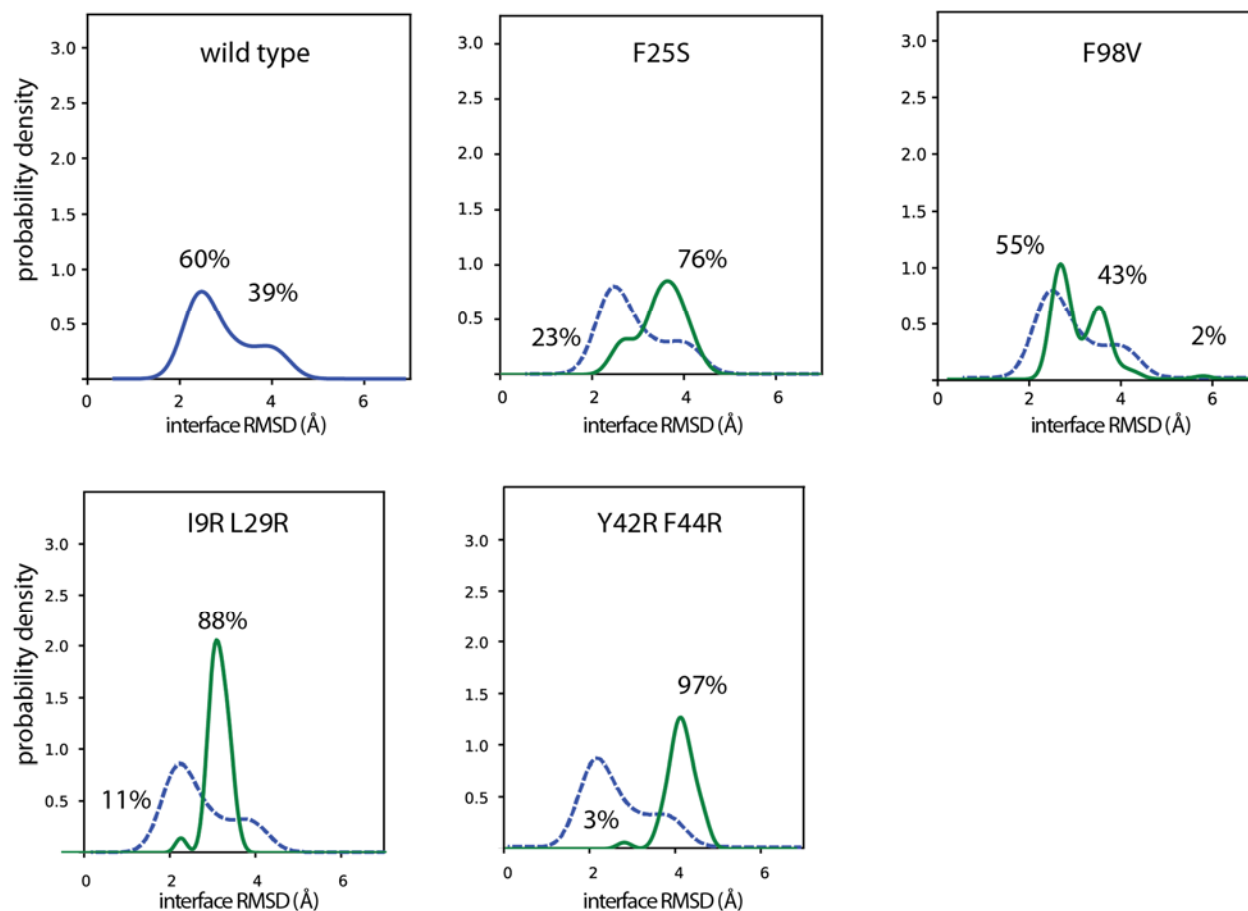


Figure S4. Stability analysis of the Saraste interface in wild-type and mutant PH-TH dimers.

The conformational equilibria at the dimer interface of various PH-TH dimer mutants are shown, each compared to the wild-type equilibrium (top left and dashed lines). The fractions of conformations were quantified by calculating the number of instantaneous structures sampled in the Saraste and pre-Saraste conformations from the rung-0 state, and smoothed using a Gaussian kernel density estimator to produce the distribution graph (see Methods for details).

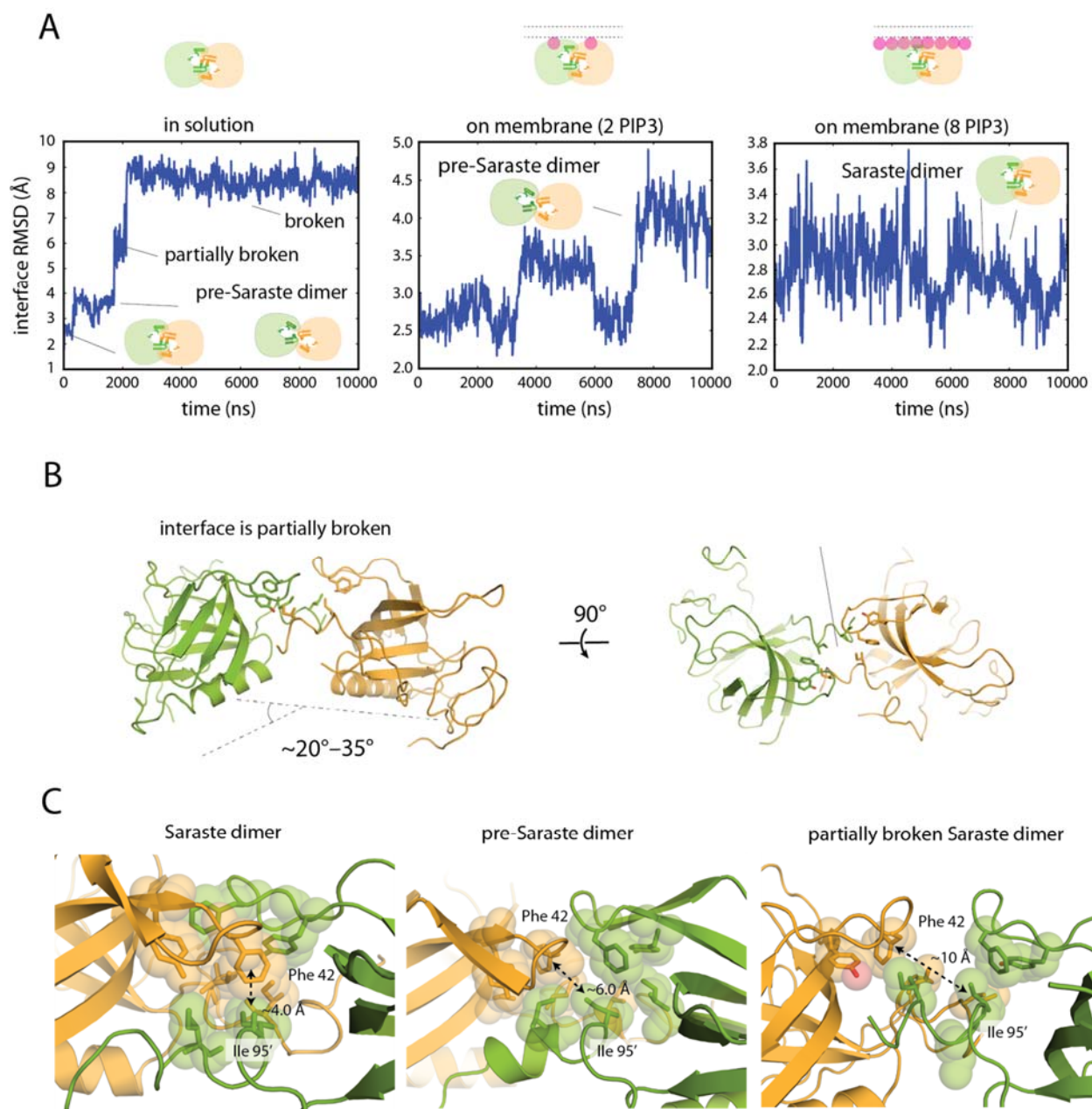


Figure S5. Saraste dimer instability in aqueous solution. A. Fluctuations at the Saraste interfaces of PH-TH dimers in conventional MD simulations with and without membranes. The simulations started from the crystal structure conformation of the Saraste dimer (PDB ID: 1BTK). The RMSD was calculated for the Saraste dimer interface residues of both modules

(after aligning on these same residues). B. A representative instantaneous MD structure showing the metastable partially broken conformation of the Saraste dimer seen in solution-based simulations. C. Hydrophobic packing at the Saraste interface seen in representative conformations of Saraste dimers.

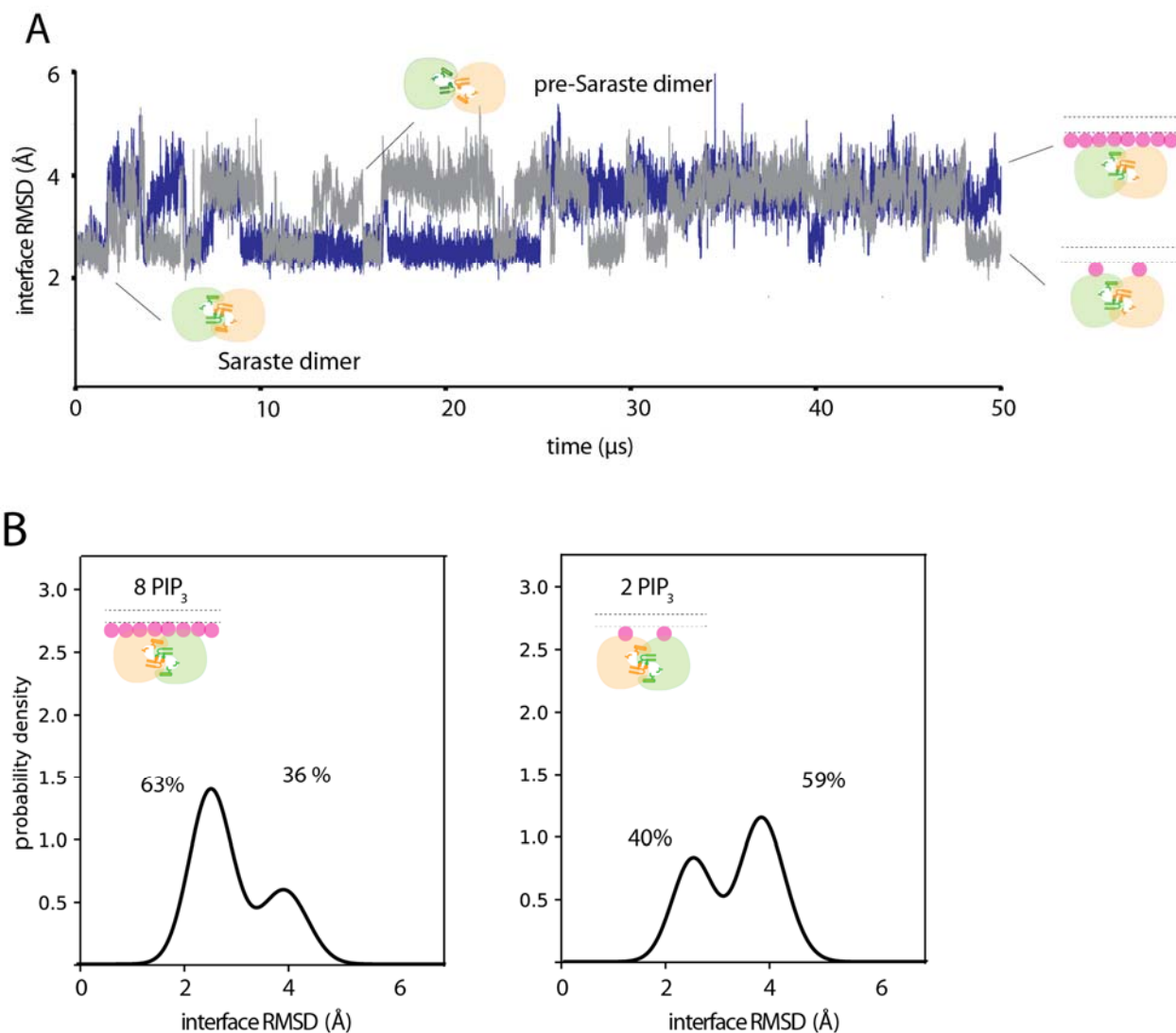
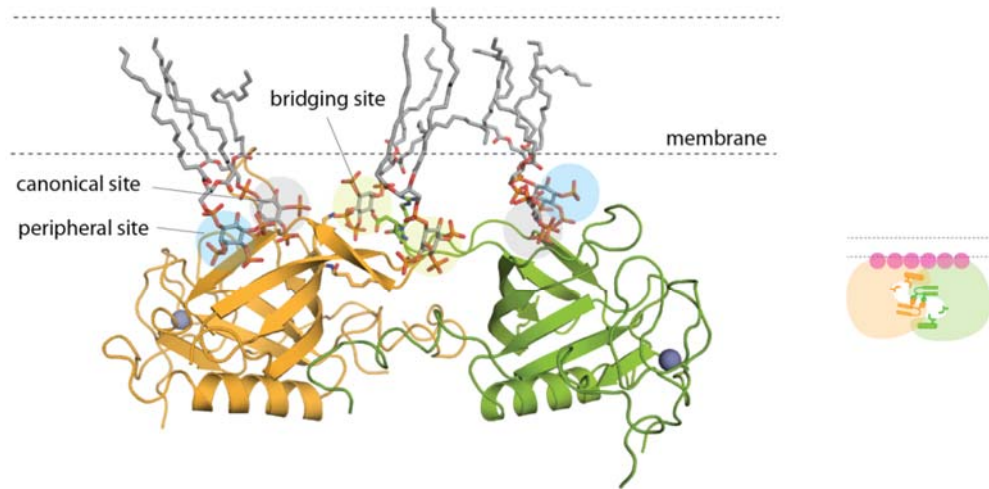


Figure S6. Reversible transitions between the Saraste interface and the pre-Saraste interface with differing membrane PIP₃ content. **A.** Representative tempered binding simulation trajectories showing multiple transitions between the pre-Saraste interface conformation and the Saraste interface conformation in two individual trajectories with membranes containing 1.5% PIP₃ (2 PIP₃ molecules) and 6% PIP₃ (8 PIP₃ molecules). **B.** Conformational equilibria at the dimer interface of the wild-type PH-TH module on membranes containing 6% PIP₃ and 1.5% PIP₃, respectively. The fractions of conformations were quantified by calculating the

number of instantaneous structures sampled in the Saraste and pre-Saraste conformations from the rung-0 state, and smoothed using a Gaussian kernel density estimator to produce the distribution graph (see Methods for details).

A



B

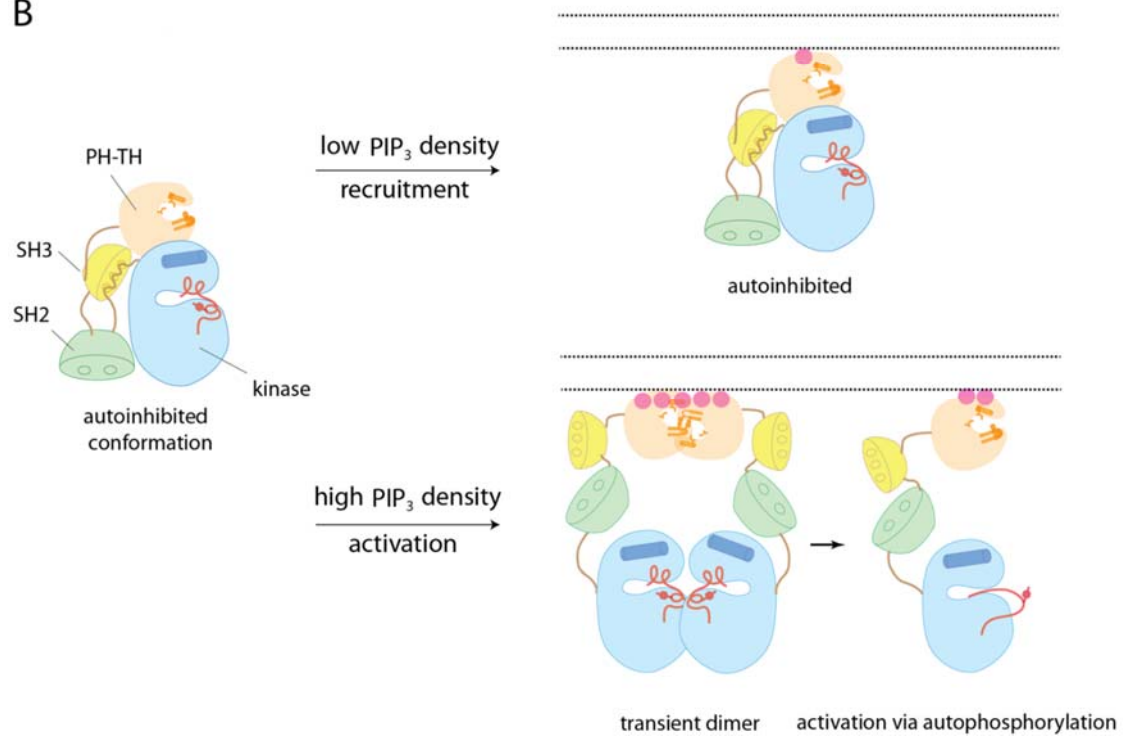


Figure S7. Model for PIP₃-dependent activation of Btk. A. An instantaneous structure from a tempered binding simulation showing a PH-TH module bound to multiple lipids. B. Cartoon illustration of the PIP₃-dependent activation of Btk on a membrane.

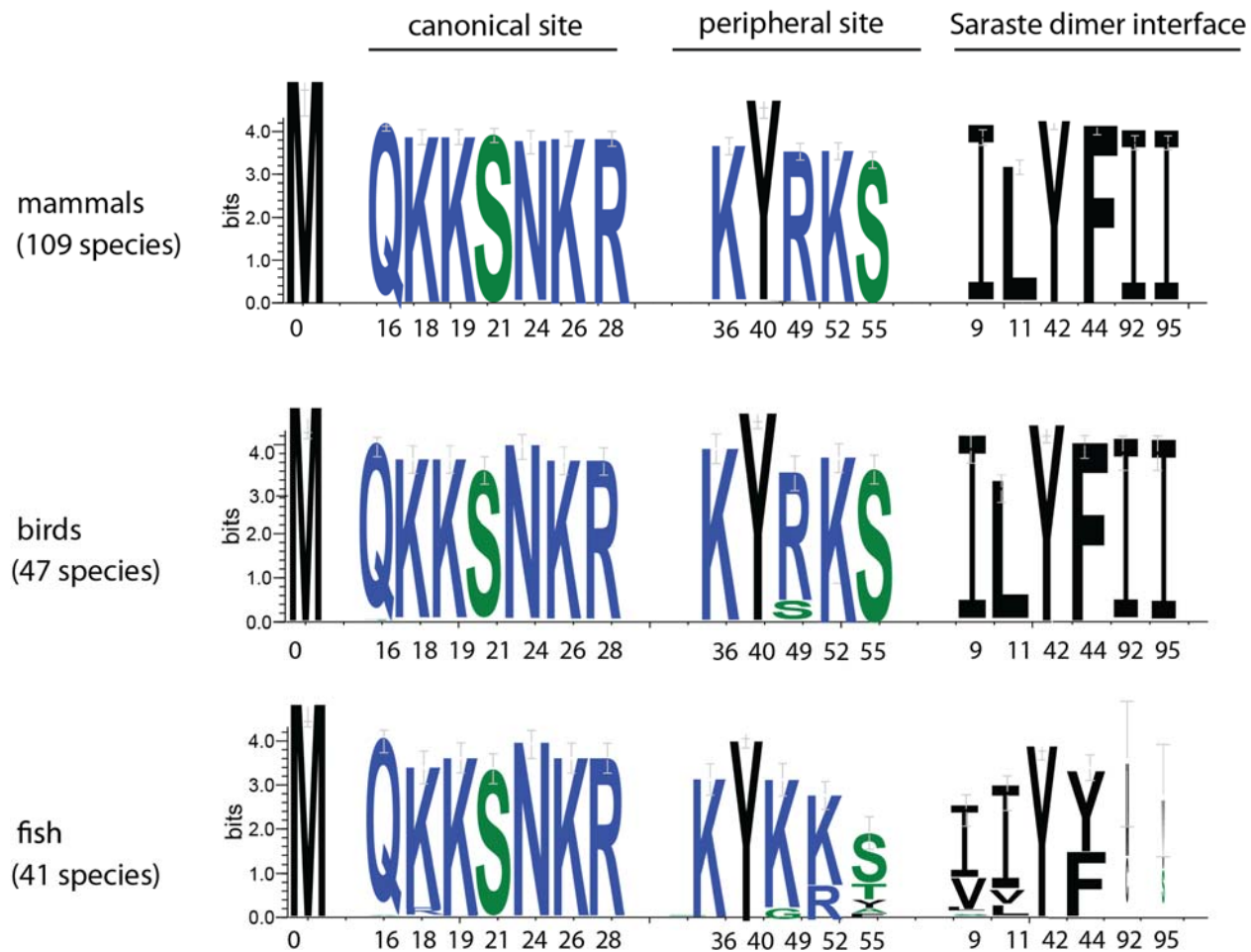


Figure S8. Sequence conservation analysis across various Btk species. Sequence logos showing the conservation at canonical site, peripheral site and Saraste dimer interface. Human Btk PH-TH sequence (residues 1–172) was aligned with 197 species from fish, birds, and mammals. Only the residues that constitute the sites of interest are shown, for clarity.

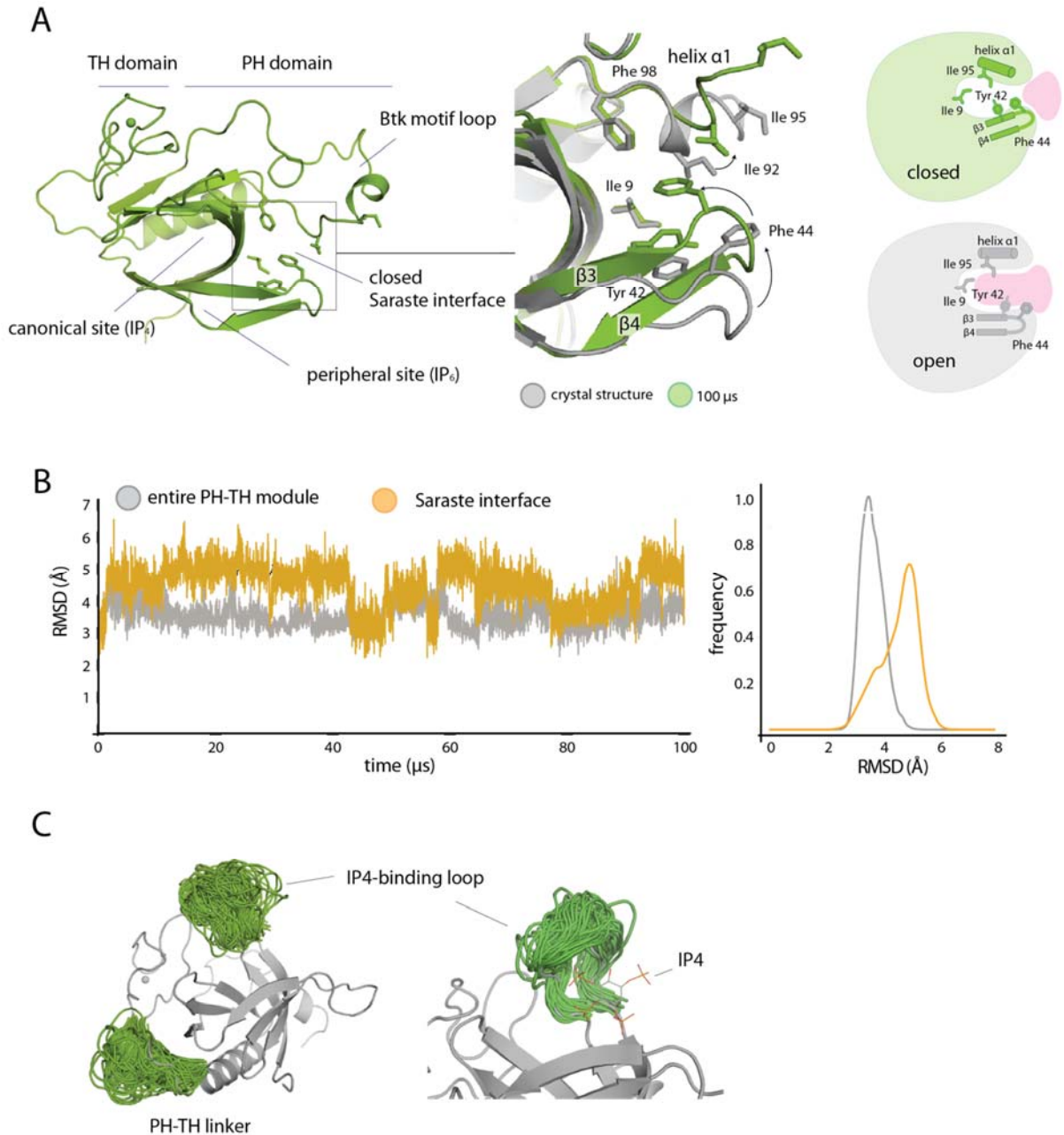


Figure S9. Conformational dynamics of an individual PH-TH module of Btk. A. Overlay of an instantaneous structure ($t = 100 \mu\text{s}$) from a simulation of the crystal structure of the PH-TH module of Btk (PDB ID: 1BTK).¹ B. Fluctuations in the Btk PH-TH module and in the Saraste interface during a $100 \mu\text{s}$ MD simulation. The RMSD was calculated using the crystal structure

of the PH-TH module of Btk as a reference (see Methods for details). C. Conformational fluctuations of the IP₄-binding loop and the PH-TH linker in simulations of an individual PH-TH module. 100 instantaneous structures from the 100- μ s simulation were extracted at an equal time step (1 μ s) and overlaid together. The structural overlay of the IP₄-binding loop conformation in the IP₄-bound crystal structure and the MD instantaneous structure of the PH-TH module is shown at right.

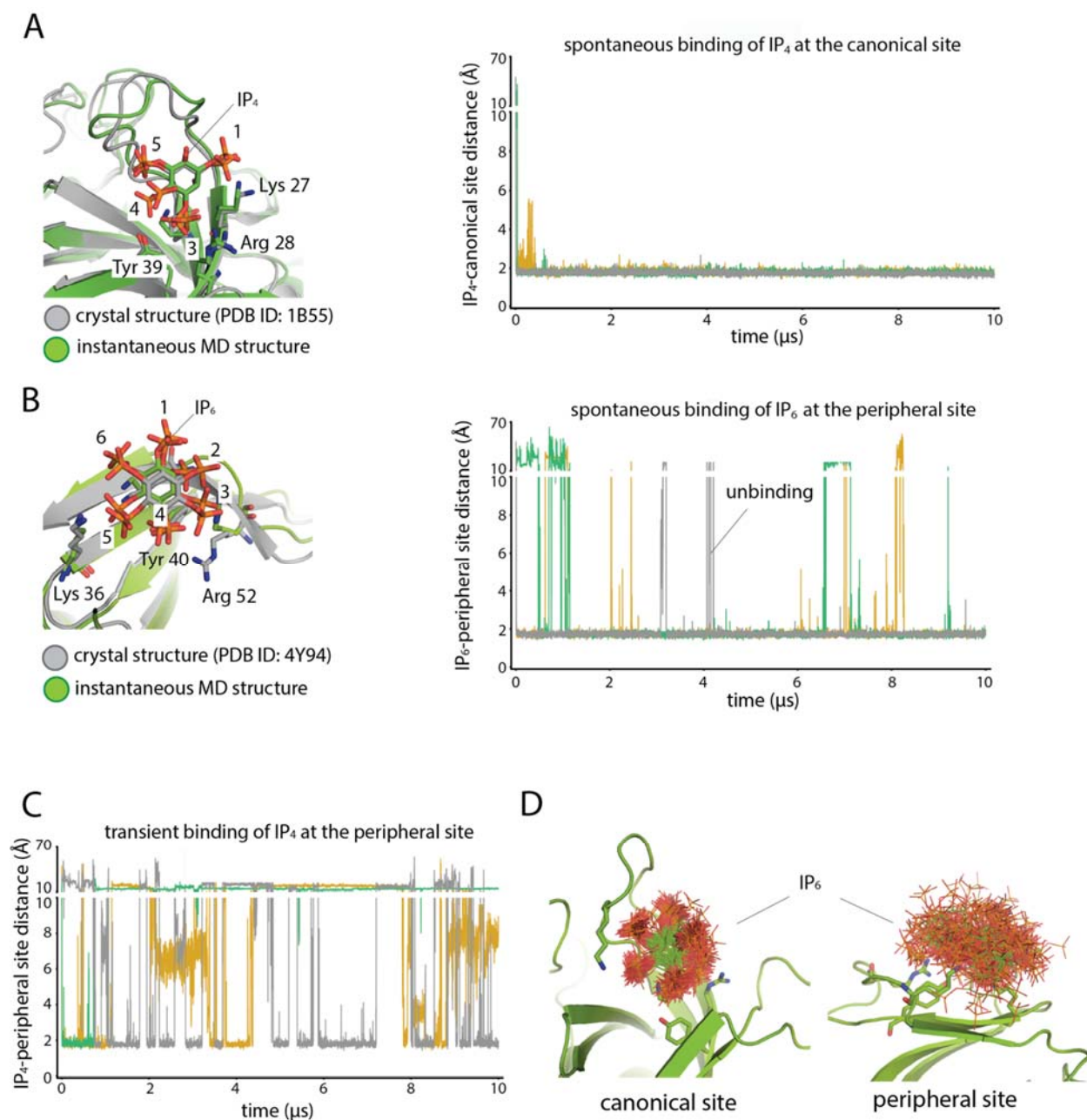


Figure S10. Spontaneous binding of IP₄ and IP₆ to the PH-TH module. A. Minimum distance between IP₄ atoms and canonical binding site atoms in three simulation trajectories (see Methods for details). An overlay of an instantaneous structure ($t = 10 \mu\text{s}$) from one of the simulations onto the crystal structure of the IP₄-bound PH-TH module of Btk (PDB ID: 1B55)³ is shown on

the left. Phosphate groups (from both MD frame and crystal structure) are shown in orange and red. B. Simulation trajectories of IP₆ binding to the peripheral site of the PH-TH module. An overlay of an instantaneous structure ($t = 5 \mu\text{s}$) from one of the simulations onto the crystal structure of the IP₆-bound PH-TH module of Btk (PDB ID: 4Y94)² is shown on the left. C. Simulation trajectories showing transient binding of IP₄ at the peripheral site of the PH-TH module. D. Overlay of 100 instantaneous structures from the 10- μs spontaneous binding simulations showing the structural fluctuations of IP₆ at the canonical and peripheral binding sites. Structures were extracted at an equal time step (100 ns); unbound structures are not shown.

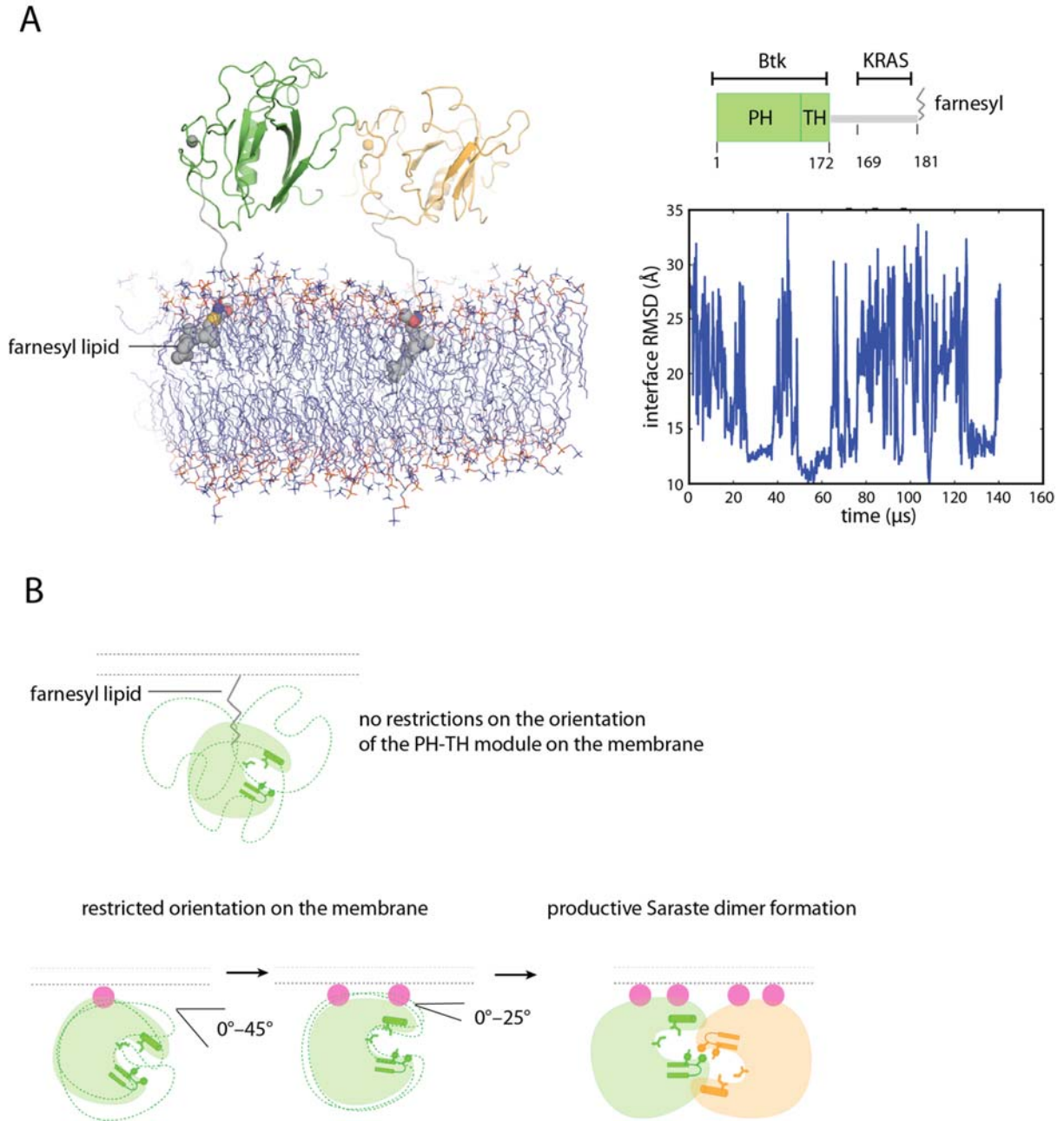


Figure S11. Encounter simulations of the PH-TH-KRAS fusion protein on the membrane. A. Illustration of the encounter simulation setup and the design of the fusion protein. A representative simulation trajectory is shown on the right. B. A comparison of PH-TH orientations on the membrane under different membrane-recruitment strategies.

Supplemental Tables

| Sim. ID | PH-TH construct (1-172) | # PIP3 | Tempering strength | Backbone patch (kcal mol⁻¹) | Simulation time (μs) |
|----------------|------------------------------------|---------------|-------------------------------|---|--|
| 1 | $\Delta(80-88)$ | 8 | 0.997 | 1.0 | 100 |
| 2 | $\Delta(80-88)$ | 8 | 0.997 | 1.0 | 100 |
| 3 | $\Delta(80-88)$ | 8 | 0.997 | 1.0 | 100 |
| 4 | $\Delta(80-86)$ | 8 | 0.997 | 1.0 | 100 |
| 5 | $\Delta(80-82)$ | 8 | 0.997 | 1.0 | 50 |
| 6 | $\Delta(82-88)$ | 8 | 0.997 | 1.0 | 50 |
| 7 | $\Delta(80-82)$ | 8 | 0.995 | 1.0 | 100 |
| 8 | $\Delta(80-82)$ | 8 | 0.993 | 1.0 | 100 |
| 9 | $\Delta(80-82)$ | 8 | 0.991 | 1.0 | 100 |
| 10 | $\Delta(80-82)$ | 8 | 0.997 | 1.0 | 100 |

| | | | | | |
|----|-----------------|---|-------|------|-----|
| 11 | $\Delta(80-82)$ | 8 | 0.997 | 0.1 | 100 |
| 12 | $\Delta(80-82)$ | 8 | 0.997 | 0.01 | 400 |
| 13 | $\Delta(80-82)$ | 2 | 0.997 | 1.0 | 100 |
| 14 | $\Delta(80-82)$ | 2 | 0.997 | 1.0 | 100 |
| 15 | $\Delta(80-82)$ | 2 | 0.997 | 1.0 | 100 |
| 16 | $\Delta(80-82)$ | 2 | 0.991 | 0.01 | 100 |
| 17 | $\Delta(80-82)$ | 2 | 0.991 | 0.01 | 100 |
| 16 | $\Delta(80-82)$ | 2 | 0.991 | 0.01 | 100 |
| 17 | $\Delta(80-82)$ | 2 | 0.993 | 0.01 | 100 |
| 18 | $\Delta(80-82)$ | 2 | 0.995 | 0.01 | 100 |
| 19 | $\Delta(82-82)$ | 2 | 0.997 | 0.01 | 140 |
| 20 | $\Delta(80-82)$ | 2 | 0.997 | 0.1 | 100 |
| 21 | $\Delta(80-82)$ | 2 | 0.997 | 0.1 | 100 |

| | | | | | |
|----|-----------------|---|-------|-----|-----|
| 22 | $\Delta(80-82)$ | 2 | 0.997 | 0.1 | 100 |
| 23 | $\Delta(80-82)$ | 2 | 0.995 | 0.1 | 100 |
| 24 | $\Delta(80-82)$ | 2 | 0.995 | 0.1 | 100 |
| 25 | $\Delta(80-82)$ | 1 | 0.997 | 0.1 | 100 |
| 26 | $\Delta(80-82)$ | 2 | 0.997 | 0.1 | 100 |
| 27 | $\Delta(80-82)$ | 4 | 0.997 | 0.1 | 100 |
| 28 | $\Delta(80-82)$ | 8 | 0.997 | 0.1 | 100 |
| 29 | $\Delta(80-82)$ | 8 | 0.995 | 1.0 | 100 |
| 30 | $\Delta(80-82)$ | 8 | 0.995 | 1.0 | 100 |
| 31 | F98V | 8 | 0.995 | 1.0 | 20 |
| 32 | F98V | 8 | 0.995 | 1.0 | 20 |
| 33 | E41K | 8 | 0.995 | 1.0 | 100 |
| 34 | E41K | 8 | 0.995 | 1.0 | 100 |

| | | | | | |
|----|------|---|-------|-----|----|
| 35 | F25S | 8 | 0.995 | 1.0 | 20 |
| 36 | F25S | 8 | 0.995 | 1.0 | 20 |

Table S1. List of tempered binding simulations performed in this work. Each tempered binding simulation included three replicas; the simulation length given is per replica.

| Figure number | Simulation IDs |
|----------------------|-----------------------|
| Figure 2B | 1–24 |
| Figure 2C | 19 |
| Figure 4A | 29, 30 |
| Figure 4B | 31, 32 |
| Figure 4C | 33, 34 |
| Figure 5A | 12 |
| Figure 5B | 10, 13 |
| Figure 6 | 25–28 |

Table S2. Tempered binding simulations used in the main text figures.