

Deep mutational scanning of the multi-domain phosphatase SHP2 reveals mechanisms of regulation and pathogenicity

Received: 19 December 2024

Accepted: 28 May 2025

Published online: 01 July 2025

Ziyuan Jiang¹, Anne E. van Vlimmeren^{1,2}, Deepti Karandur³,
Alyssa Semmelman¹ & Neel H. Shah^{1,4} ✉

Multi-domain signaling enzymes are often regulated through extensive inter-domain interactions, and disruption of inter-domain interfaces by mutations can lead to aberrant signaling and diseases. For example, the tyrosine phosphatase SHP2 contains two phosphotyrosine recognition domains that auto-inhibit its catalytic domain. SHP2 is canonically activated by binding of these non-catalytic domains to phosphoproteins, which destabilizes its auto-inhibited state, and several mutations at the main auto-inhibitory interface have been shown to hyperactivate SHP2 in cancers and developmental disorders. Hundreds of clinically observed mutations in SHP2 have not been characterized, but their locations suggest alternative modes of dysregulation. Here, we perform deep mutational scanning on full-length SHP2 and the isolated phosphatase domain to dissect mechanisms of dysregulation. Our analysis reveals mechanistically diverse mutational effects and identifies key intra- and inter-domain interactions that contribute to SHP2 activity, dynamics, and regulation. Our datasets also provide insights into the potential pathogenicity of previously uncharacterized clinical variants.

The majority of human signaling proteins contain multiple domains with distinct biochemical functions. These domains facilitate catalytic activity, such as phosphorylation and dephosphorylation, or binding to ligands, including short linear protein motifs¹, phosphorylated protein residues², lipids³, and nucleic acids⁴. Regulatory interactions between these functional domains and inter-domain linkers enable multi-domain signaling proteins to function as highly regulatable switches capable of accessing different conformations in response to specific signals^{5–7}. While this metastability is critical for regulation, it also makes multi-domain signaling proteins highly susceptible to dysregulation by mutations, which can disrupt their intra-domain allostery, inter-domain interactions, or linker structure and dynamics, leading to aberrant signaling and diseases⁸. A quintessential example is the protein tyrosine phosphatase (PTP) SHP2. This enzyme has two phosphotyrosine recognition (SH2) domains N-terminal to its catalytic

PTP domain (Fig. 1a). In signaling pathways such as the Ras/Erk and Jak/Stat pathways, SHP2 is recruited via its SH2 domains to phosphotyrosine-bearing sequences on receptors and scaffold proteins (Fig. 1b). Upon phosphoprotein binding, SHP2 switches from a closed, auto-inhibited state characterized by extensive N-SH2/PTP domain interactions, to an open, active state, allowing for the dephosphorylation of downstream substrates (Fig. 1b, c)⁹.

Missense mutations in the SHP2-coding gene (*PTPN11*) can cause cancers and developmental disorders¹⁰. The functional effects of mutations in SHP2 vary across disease contexts. For example, in the developmental disorders Noonan syndrome and Noonan syndrome with multiple lentigines, gain-of-function SHP2 mutants confer their pathogenicity by promoting Ras/Erk activation, but some loss-of-function mutants can have similar phenotypic effects¹¹. Numerous cases of gain-of-function SHP2 mutants are found in hematopoietic

¹Department of Chemistry, Columbia University, New York, NY 10027, USA. ²Department of Biological Sciences, Columbia University, New York, NY 10027, USA. ³Department of Biochemistry, Vanderbilt University, Nashville, TN 37232, USA. ⁴Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY 10032, USA. ✉e-mail: neel.shah@columbia.edu

cancers^{12,13}, and SHP2 also plays critical roles in the development of different types of solid tumors^{14–16}. However, SHP2 has also been demonstrated to function as a tumor suppressor in liver cancer and a few other cancer types^{17–19}. The spectrum of phenotypic effects caused by SHP2 variants is likely reflective of different axes of SHP2 function, including intrinsic catalytic activity, allosteric regulation, and protein-protein interactions. Given this, rigorous biochemical and biophysical characterization of disease-associated SHP2 mutants has substantially informed our understanding of SHP2 dysregulation and pathogenicity. For example, the extensive characterization of activating SHP2 mutants, including the high-frequency cancer mutation E76K, revealed key features of the N-SH2/PTP auto-inhibitory interface and also yielded the first active-state structure of SHP2^{20,21}. The Noonan Syndrome T42A mutation in the N-SH2 domain was found to alter SHP2 ligand affinity and specificity, sensitizing SHP2 to activators^{22,23}. The Y279C mutation, found in Noonan syndrome with multiple lentigines, has been shown to disrupt phosphoprotein binding in the PTP active site, thereby significantly diminishing catalytic activity²⁴. However, hundreds of disease-associated SHP2 mutations remain poorly studied²⁵. These mutations span all three domains of SHP2 and the linkers between them, and many of them lack obvious mechanistic explanations of pathogenicity, making their significance unclear (Fig. 1d). Characterization of these mutations will help elucidate unknown features of SHP2 structure, regulation, and pathogenicity.

Deep mutational scanning is a powerful method for characterizing protein mutants²⁶. By combining selection assays on pooled libraries with deep sequencing, this method provides a way to profile mutational effects across a protein with high throughput. Deep mutational scanning has been applied to reveal structure–function relationships²⁷, predict protein structures and dynamics^{28,29}, map drug resistance³⁰, and examine stability and expression^{31,32}. Molecular insights into structure and regulation derived from deep mutational scans have proven to be useful to rationalize pathogenicity^{33–35}. Here, we present a deep mutational scanning platform to characterize the effects of SHP2 mutations on phosphatase activity. We used this platform to examine comprehensive point mutant libraries of both full-length SHP2 (SHP2_{FL}) and its isolated phosphatase domain (SHP2_{PTP}). These experiments yielded the activity profile of over 11,000 SHP2 mutants, including previously uncharacterized variants documented in disease databases. A comparative analysis of the two datasets, supported by molecular dynamics (MD) simulations, revealed several classes of mutations with distinct mechanisms of dysregulation and pathogenicity. Our analysis also identified interactions that mediate the transition of SHP2 between its auto-inhibited and active states, providing new insights into SHP2 regulation.

Results

Rescue of yeast growth from tyrosine kinase toxicity enables SHP2 deep mutational scanning

We developed a yeast viability assay in which cell growth is dependent on the catalytic activity of SHP2. Lacking significant endogenous tyrosine kinase/phosphatase signaling, yeast (*S. cerevisiae*) proliferation is arrested when expressing an active tyrosine kinase, whereas co-expression of an active tyrosine phosphatase can rescue yeast growth^{36–38}. We co-expressed SHP2_{FL} variants known to have different levels of catalytic activity with two active versions of Src kinase, full-length viral Src (v-Src_{FL}) and the isolated c-Src kinase domain (c-Src_{KD}). In the presence of either kinase, the rate of yeast growth was dependent on the catalytic activity of the SHP2 variant (Supplementary Fig. 1a). Moreover, the activity of the tyrosine kinase dictates the selection pressure of the assay. With the highly active v-Src_{FL} kinase, the more active SHP2 variants were better differentiated, while with the less active c-Src_{KD} kinase, lower activity SHP2 variants were easier to differentiate (Supplementary Fig. 1b, c).

To comprehensively characterize SHP2 mutant activity, we conducted selection assays with SHP2_{FL} and SHP2_{PTP} saturation mutagenesis libraries co-expressed with each Src kinase variant (Fig. 2a). Libraries were constructed using the mutagenesis by integrated TILES (MITE) method, with SHP2_{FL} and SHP2_{PTP} divided into 15 and 7 sub-libraries (tiles), respectively (Supplementary Fig. 1d and Supplementary Data 1)³⁹. Each sub-library was separately introduced into yeast cells alongside plasmids encoding either v-Src_{FL} or c-Src_{KD} (Supplementary Fig. 1e). Cells were then subject to selection by induction of kinase and phosphatase expression, followed by a 24-h outgrowth phase. Before and after outgrowth, the SHP2-coding DNA was isolated and deep sequenced, allowing for the calculation of enrichment scores for each variant, relative to wild-type SHP2 (see Methods). Two to four replicates of high-quality selection data were acquired for each sub-library (each tile), with good correlation of enrichment scores between replicates (Supplementary Fig. 2a). The average enrichment scores of all variants across replicates were plotted as heatmaps. Due to the selection pressures of the two conditions and differences in intrinsic catalytic activity between SHP2_{FL} and SHP2_{PTP}, we observed that the SHP2_{FL} + c-Src_{KD} and SHP2_{PTP} + v-Src_{FL} selection assays provided the best dynamic range for both gain- and loss-of-activity mutants, and these datasets were used for subsequent analysis (Fig. 2b, c and Supplementary Fig. 2b–d).

To validate that our selection assay faithfully reports on SHP2 catalytic activity as a function of mutations, we purified several full-length SHP2 mutants and measured their basal catalytic activities (Supplementary Data 2). The catalytic efficiencies (k_{cat}/K_M) of these mutants correlated well with their enrichment scores in the SHP2_{FL} selection (Fig. 2d). Thus, while our mutational scans may be impacted by mutant-specific changes in expression level, protein-protein interactions, or substrate specificity, our results suggest that basal catalytic activity is the major determinant of enrichment in our selection assays. For further validation, we compared our datasets with known mutational effects. As expected, well-studied activating mutations at the N-SH2/PTP interface (e.g., E76, D61, and S502 substitutions) were highly enriched in the SHP2_{FL} selections (Fig. 2b, e). Furthermore, mutations at key catalytic residues (e.g., C459 and D425) were depleted in both the SHP2_{FL} and SHP2_{PTP} selections (Fig. 2b, c). In addition to the known mutations, we observed strong mutational hotspots in unexpected regions, including activating mutations in the core of the N-SH2 domain, inactivating mutations at the C-SH2/PTP interface, and activating mutations around the key catalytic WPD loop, which will be discussed in later sections (Fig. 2e).

Mutational scans reveal disease-specific profiles of SHP2 activity

Of the ~600 clinically observed SHP2 variants, only 20% are annotated as pathogenic (Supplementary Data 3). Thus, we used our SHP2_{FL} enrichment scores to gain insights into the functional effects of clinical variants. When compared to the full distribution of mutational effects, mutations annotated as pathogenic were more gain-of-function on average, however many reported pathogenic mutations did not enhance SHP2 activity (Fig. 3a, b). This is consistent with the presence of some known loss-of-activity SHP2 mutations in developmental disorders⁴⁰. High-frequency cancer mutations skewed further toward gain-of-activity, but even in this category, a few were neutral or even loss-of-activity (Fig. 3b). It is noteworthy that many low-frequency cancer mutations, reported in fewer than five cases across COSMIC and TCGA, were neutral or loss-of-activity in the SHP2_{FL} selection assay. Most of these mutations have not been characterized, but they may still drive oncogenic signaling through mechanisms that do not rely on phosphatase activity, such as scaffolding mediated by the SH2 domains.

We further parsed cancer mutations by the primary tissue type and were surprised to find differences in the distribution of SHP2

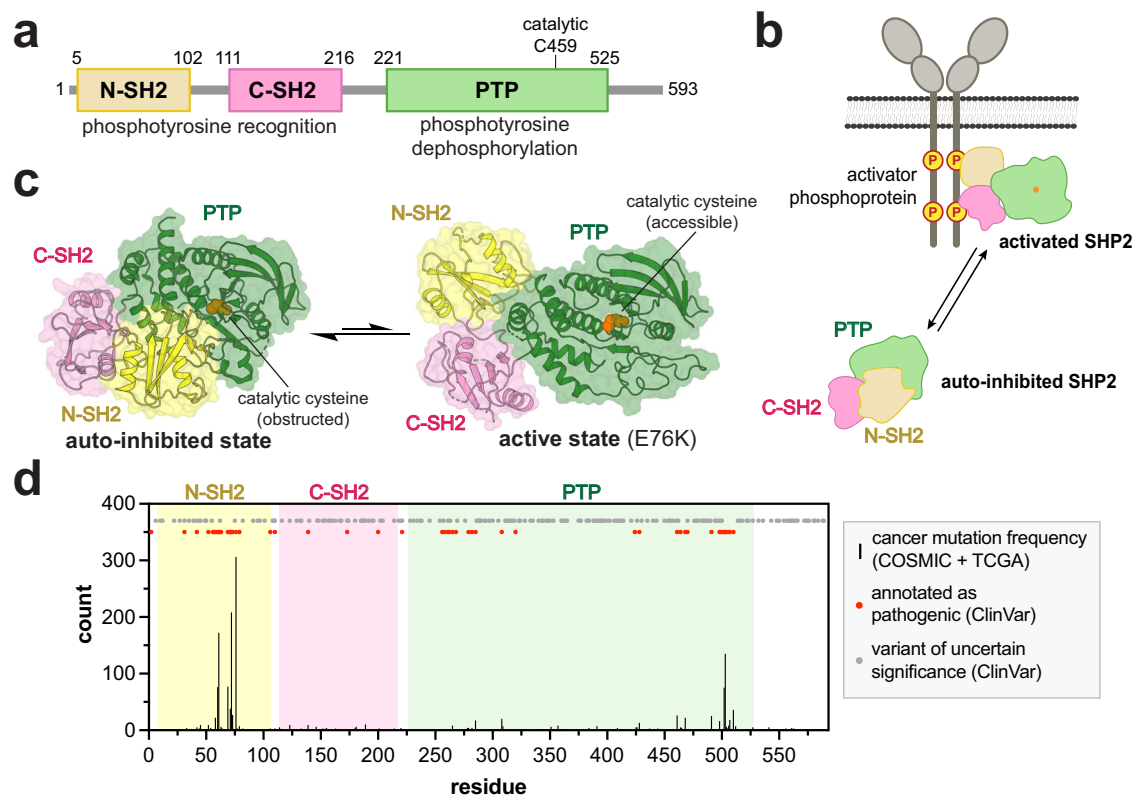


Fig. 1 | SHP2 activation and dysregulation by mutations. **a** Domain architecture diagram of SHP2. **b** Cartoon depiction of SHP2 activation by phosphoprotein binding. **c** Structures of the auto-inhibited (PDB code 4DGP) and active states of SHP2 (PDB code 6CRF). Note that the active state shown is that of the E76K mutant, and other active conformations likely exist in solution. **d** Positions and frequencies

of cancer hotspot mutations (black bars) in SHP2 along its 593-residue sequence, derived from the COSMIC and TCGA databases. Sites of other pathogenic mutations and variants of uncertain significance are labeled as red and gray dots, respectively, derived from the ClinVar database.

activity across cancer subtypes (Fig. 3c). As expected, many blood cancer mutations were highly activating in the SHP2_{FL} mutational scan^{11,41}. Gain-of-activity mutants were less common in other cancer types, but those observed were largely the same as the hallmark activating mutations observed in blood cancers (e.g., D61, A72, E76, and S502 substitutions). The prevalence of neutral and loss-of-function mutants in many solid tumor types raises the question of whether those mutations are drivers of pathogenic signaling or simply bystanders. A striking observation from our analysis that supports a pathogenic role for loss-of-function SHP2 mutations is that all documented SHP2 liver cancer mutants were neutral or loss-of-activity in our dataset (Fig. 3c and Supplementary Data 3). Mouse studies have shown that SHP2 knock-out in hepatocytes enhances pro-inflammatory Stat3 signaling and promotes liver cancer, however this has not been demonstrated with missense mutations¹⁷. Given that all liver cancer mutants were neutral or loss-of-activity, we hypothesize that mutants in this context are pathogenic through loss of SHP2 tumor suppressor activity. As such, these mutants may warrant further investigation using in vivo models.

Impaired catalytic activity is compensated by disrupted auto-inhibition in many disease mutants

Many unstudied SHP2 disease mutants are found in the PTP domain. Our SHP2_{PTP} selection assays show that the majority of these disease mutants reduce the intrinsic phosphatase activity of SHP2 (Fig. 3d–f). To better understand the molecular basis for pathogenicity of PTP domain mutations, we juxtaposed enrichment scores from SHP2_{PTP} and SHP2_{FL} selections (Fig. 3g). A monotonic relationship between the two datasets would suggest that the signal in both datasets depends solely on phosphatase domain activity. However, the enrichment

scores follow a funnel-like distribution, spanning all four quadrants of the two-dimensional space, allowing us to decouple mutational effects on intrinsic catalytic activity from those that alter inter-domain interactions (Fig. 3g). To narrow our analysis, we focused on significantly activating or deactivating mutants in either selection that fall more than two standard deviations from the mean of the distribution centered around wild-type (Fig. 3a, d, g). We binned these mutations into four distinct regions (Fig. 3g, colored boxes). Disease mutations in each of the regions dysregulate SHP2 through distinct mechanisms (Fig. 3g, circled points).

Region I encompasses mutations in the PTP domain that disrupt auto-inhibition but have either a neutral or negative effect on intrinsic phosphatase activity. Many well-studied mutations in this region exclusively disrupt the auto-inhibitory interface, with a nominal effect on SHP2_{PTP} activity (e.g., cancer mutations at R265 and S502, which interact with E76 in auto-inhibited SHP2). By contrast, our datasets reveal a number of previously unstudied disease mutants in region I that significantly reduce intrinsic phosphatase activity (depleted in SHP2_{PTP} dataset) but compensate for this perturbation by alleviating auto-inhibition (neutral or enriched in SHP2_{FL} dataset) (Supplementary Data 3). Many of them are positioned close to key catalytic loops that regulate both catalysis and auto-inhibition, such as D286, H287, T288, N306, and N308 mutants underneath the substrate-binding pTyr loop, or R501, M504, and V505 mutants in the Q loop region (Supplementary Fig. 3a, b). Frequent disease mutations at P491 also fall into this category, likely by disrupting C-SH2/PTP interactions in the closed conformation to alleviate auto-inhibition (Supplementary Fig. 3b, c). Previous studies have demonstrated that mutants with reduced auto-inhibition more readily interact with binding partners in signaling pathways, leading to increased signaling activity^{22,24,42}. The prevalence

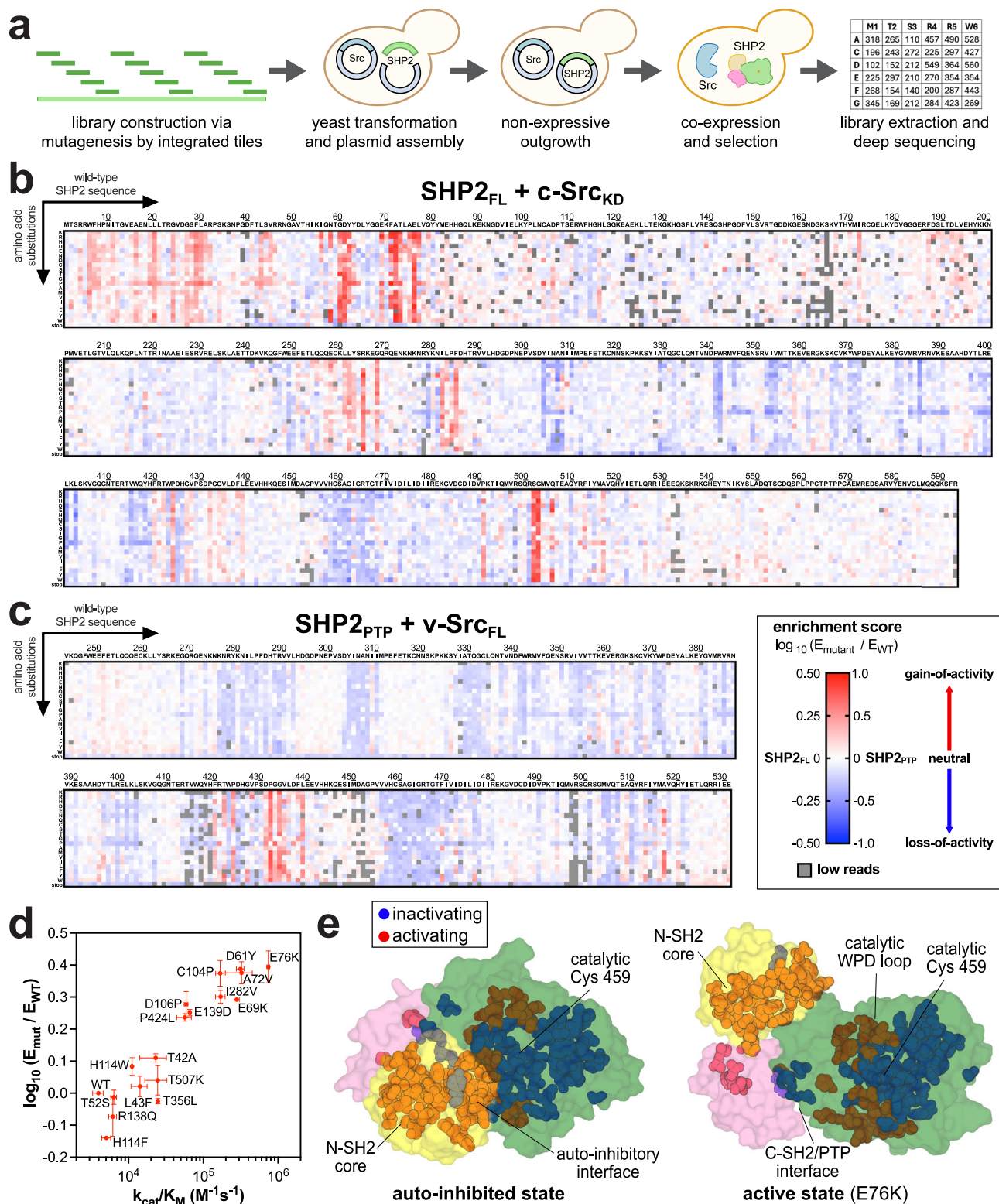


Fig. 2 | Deep mutational scanning of SHP2. a Workflow for SHP2 mutational scanning. SHP2 variant libraries were constructed using MITE, integrated into a yeast plasmid, co-expressed with Src, and subject to selection and deep sequencing. **b** Heatmap depicting the enrichment scores for SHP2_{FL} co-expressed with c-Src_{KD} ($n = 2-4$). **c** Heatmap depicting the enrichment scores for SHP2_{PTP} co-expressed with v-Src_{FL} ($n = 2$). **d** Correlation between SHP2 variant enrichment

scores in the SHP2_{FL} + c-Src_{KD} selection assay and measured catalytic activity against DiFUMP ($n = 3$ replicates for k_{cat}/K_M). The datapoints shown are average enrichment scores and catalytic efficiencies. Error bars represent standard deviations. **e** Sites in the SHP2_{FL} mutational scan that are substantially activating (red) and inactivating (blue) on average, mapped on the auto-inhibited and active state structures of SHP2 (PDB codes 4DGP and 6CRF, respectively).

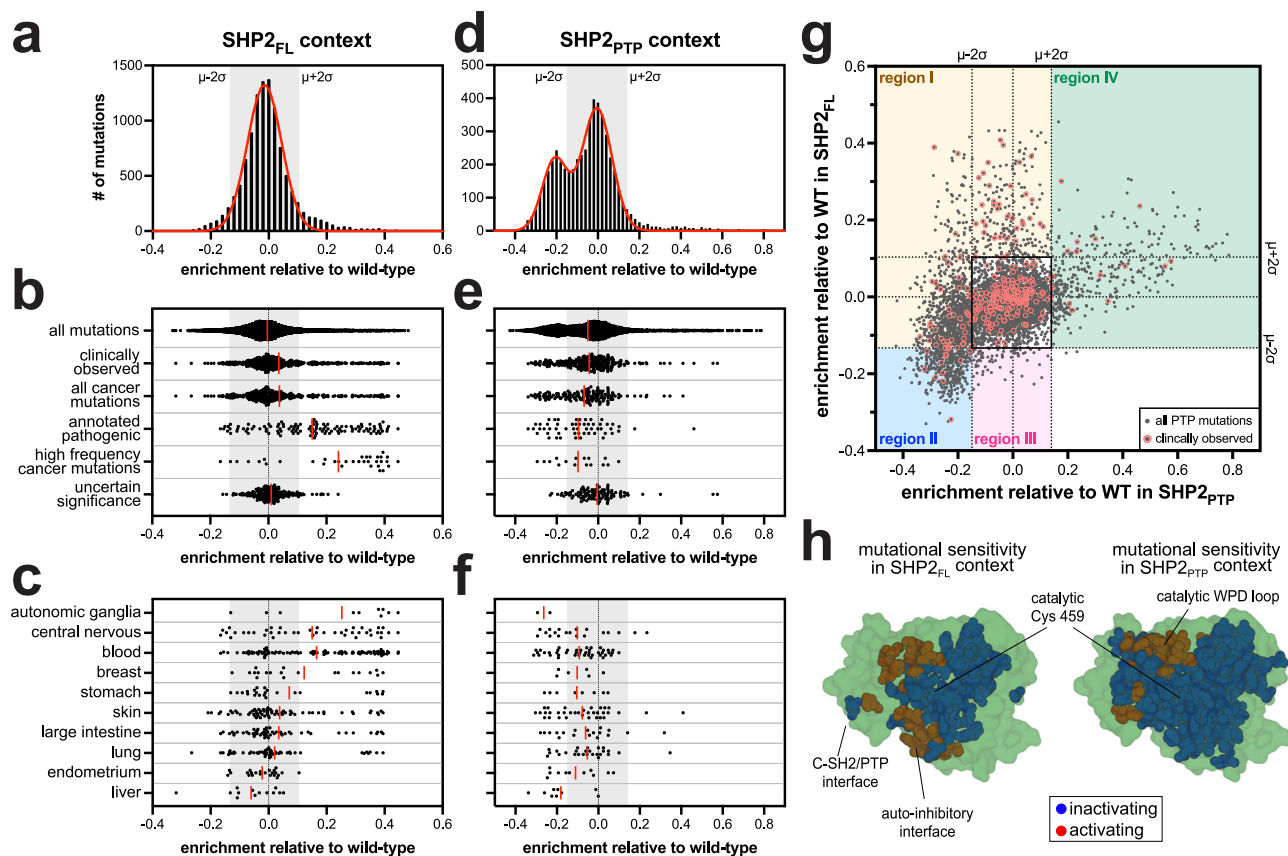


Fig. 3 | Differential phosphatase domain mutational effects in SHP2_{FL} and SHP2_{PTP} contexts. **a** Full distribution of enrichment scores from the SHP2_{FL} + c-Src_{KD} selection, fit to a single Gaussian curve. **b** Distribution of enrichment scores from the SHP2_{FL} + c-Src_{KD} selection for various sets of clinically relevant mutants. **c** Distribution of enrichment scores from the SHP2_{FL} + c-Src_{KD} selection for mutants found in different cancer subtypes. **d** Full distribution of enrichment scores from the SHP2_{PTP} + v-Src_{FL} selection, fit to the sum of two Gaussian curves. **e** Distribution of enrichment scores from the SHP2_{PTP} + v-Src_{FL} selection for various sets of clinically relevant PTP domain mutants. **f** Distribution of enrichment scores from the SHP2_{PTP} + v-Src_{FL} selection for PTP domain mutants found in different cancer

subtypes. Red lines in panels **b**, **c**, **e**, **f** correspond to mean values of the distributions. **g** Correlation between enrichment scores for PTP domain mutants in the SHP2_{FL} + c-Src_{KD} (y-axis) and SHP2_{PTP} + v-Src_{FL} (x-axis) selections. Mechanistically distinct regions are highlighted as different-colored boxes. The uncolored black box near the center of the graph encompasses mutations that have small or neutral functional effects in both selections. Clinically observed mutants are circled in red. **h** Sites in the SHP2_{FL} (left) and SHP2_{PTP} (right) mutational scans that are substantially activating (red) and inactivating (blue) on average, mapped on a structure of the SHP2 PTP domain (PDB code 3ZM0).

of disease-associated mutants in region I with a substantial loss of intrinsic phosphatase activity supports the idea that there are important roles of SHP2 in disease contexts that depend on protein-protein interactions but not protein dephosphorylation⁴³.

More dramatic decreases in PTP domain activity cannot be fully compensated by reduced auto-inhibition, as exemplified by previously studied region I mutants T468M and T507K^{42,44}. For these mutants, where SHP2_{PTP} activity is significantly impaired, disrupted auto-inhibition in SHP2_{FL} can only bring SHP2_{FL} activity back to wild-type levels (Supplementary Fig. 3a). Many region II mutants represent extreme cases where reduced PTP activity cannot be overcome by relieved auto-inhibition. For example, the disease-associated mutant Y279C at the region I/II boundary disrupts auto-inhibition, but it has such intrinsically low catalytic activity that it is still significantly depleted in the SHP2_{FL} selections (Supplementary Fig. 3a)^{24,45}. To corroborate this idea, we measured melting temperatures of several full-length mutants by differential scanning fluorimetry, which has been shown to report on SHP2 auto-inhibition⁴⁶. Indeed, Y279C, T468M, and T507K have lowered melting temperatures compared to wild-type, indicating less auto-inhibited conformations (Supplementary Fig. 3d). Several other clinically observed mutants in region II lie in the PTP active site and at the auto-inhibitory interface, including unstudied mutants T357M and I463L (Supplementary Fig. 3a). By analogy to well-characterized region I and II

mutations, we hypothesize that these uncharacterized loss-of-function mutants disrupt auto-inhibition and contribute to pathogenicity via phosphatase activity-independent signaling roles.

Residues at the C-SH2/PTP interface mediate exit from the auto-inhibited state of SHP2

Region III mutations are deactivating in the SHP2_{FL} context but neutral in SHP2_{PTP}. Many of them, most notably at E249, cluster on the α -helices facing the C-SH2 domain, and interactions at this interface differ between the auto-inhibited and active states (Supplementary Fig. 4a, b). We hypothesize that this interface mediates the transition from auto-inhibited to open conformations, even in the absence of any activating ligands. Two key residues on the SH2 side of this interface are R111 and H114. Most mutations at these residues deactivate SHP2_{FL}, and both residues harbor mutations reported in cancer (Fig. 4a and Supplementary Data 3). The interaction between R111 and E249 has been reported to stabilize an active conformation of SHP2⁴⁷, however the roles of these residues in SHP2 dynamics are unexplored, and H114 has not been implicated in SHP2 regulation.

To examine how the C-SH2/PTP interface governs SHP2 regulation, we performed molecular dynamics (MD) simulations on near full-length SHP2 (excluding the disordered tail), starting from three different conformational states: the auto-inhibited state (PDB code

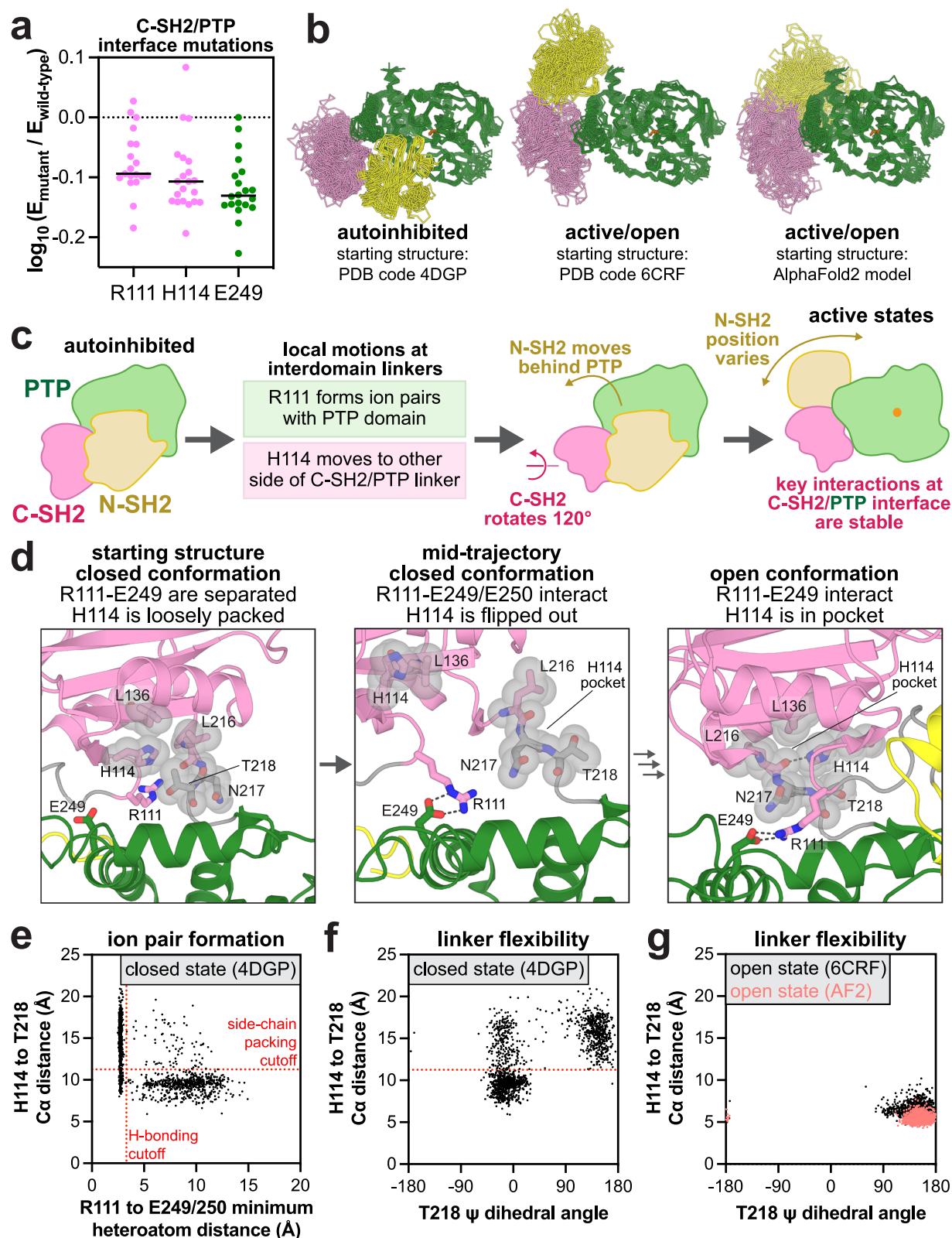


Fig. 4 | Structure and dynamics at the C-SH2/PTP interface. **a** Mutational effect at R111, H114, and E249 in the SHP2_{FL} selection with c-Src_{KD}. Black lines represent the median of enrichment scores in each distribution. **b** Conformational sampling across 18 MD simulations of SHP2 starting from three different conformational states. **c** Hypothesized sequence of events in the SHP2 closed-to-open transition. **d** Representative frames from MD simulations of SHP2 highlighting a lack of R111-E249 ion pairing and loose packing of H114 in simulations starting from the closed conformation (left), formation of the R111-E249 ion pair with concomitant exit of H114 from the C-SH2/PTP linker pocket in simulations starting from the closed

conformation (middle), and stable R111-E249 and H114-L216 interactions in simulations starting from the open conformation (right). The transition from the middle to the right frame is not observed in our simulations. **e** Correlation between R111-E249/E250 distance (shortest distance between one of the arginine terminal nitrogens and one of the four glutamate carbonyl oxygens) and H114-T218 Cα distance in the SHP2 closed conformation simulations. **f** Correlation between T218 ψ dihedral angle and H114-T218 Cα distance and in the SHP2 closed conformation simulations. **g** Correlation between R111-E249/E250 distance and H114-T218 Cα distance in the SHP2 open conformation simulations.

4DGP), a crystallographic open state of the E76K mutant (PDB code 6CRF), and an alternative open state from an AlphaFold2 model, which has the N-SH2 domain further behind the PTP domain when compared to the crystal structure (Fig. 4b). We used two open structures given the evidence that SHP2 adopts multiple active conformational states^{20,48,49}. Starting from each state, we conducted three 2.5- μ s simulations with both the wild-type sequence and the E76K mutant. In the open conformation simulations, starting from both the crystal structure and the AlphaFold2 model, the C-SH2/PTP interface is highly similar and quite stable. R111 frequently forms ion pairs with E249 and E250, or occasionally with E232 (Supplementary Fig. 4c, d). Furthermore, H114 docks in a pocket consisting of L136, L216, N217, and T218, stabilized by hydrophobic interactions as well as a hydrogen bond between its imidazole nitrogen and L216 main chain carbonyl (Fig. 4d, right panel, Supplementary Fig. 4e). Our mutagenesis data show that aromatic residues with hydrogen bonding capacity (H and W) stabilize this active state, whereas other residues, including F, reduce SHP2 activity (Supplementary Fig. 4f and Supplementary Data 2).

The C-SH2/PTP interface is more dynamic in the closed conformation simulations, and motions around R111 and H114 suggest a plausible sequence of events for the initiation of SHP2 activation (Fig. 4c, d). In contrast to the open conformation, R111 points away from the PTP domain and does not participate in significant interactions in the closed conformation starting structure, while H114 engages T218 through hydrophobic interactions with its side chain (Fig. 4d, left vs right panels). During the simulations, R111 rotates into the C-SH2/PTP cleft and intermittently forms ion pairs with E249 or E250, despite the lack of large inter-domain rearrangements (Fig. 4d, middle panel). When R111 interacts with E249/E250, this movement disrupts the interaction between H114 and T218 (Fig. 4d, middle panel, Fig. 4e). Without this stabilizing interaction, the C-SH2/PTP linker becomes dynamic, enabling T218 to adopt multiple conformations (Fig. 4f). This allows the C-SH2 domain to rotate around the C-SH2/PTP interface and sample the tighter stabilized conformation in the SHP2 open structures (Fig. 4d, right panel, Fig. 4g). Our mutational scanning data support the notion that T218-H114 packing must be disrupted for activation: substitution of T218 with smaller residues weakens the interaction with H114 and increases SHP2 activity, while bulkier substitutions enhance the H114 interaction and decrease SHP2 activity (Supplementary Fig. 4g). Full inter-domain rearrangement cannot be captured on our simulation timescale, but the extensive reorganization of interactions at the C-SH2/PTP interface between the auto-inhibited and active states, and the mutational effects of residues on the interface in our data suggest that the interface is intimately involved in SHP2 activation.

Based on our analysis, loss-of-activity mutations at this interface operate by stabilizing the closed conformation of SHP2. These mutations would make SHP2 less capable of binding upstream phosphoproteins, as demonstrated previously for E249⁴⁷. Consequently, no disease-associated PTP mutations appear in region III (Fig. 3g). On the other hand, the T218S and T218A mutations are observed in developmental disorders (Supplementary Data 3). The functional effects of these mutations were previously unknown, but our results show that disruption of the T218-H114 interaction in the auto-inhibited conformation promotes SHP2 activation (Supplementary Fig. 4b). We also note that the R111M and H114Y mutations have been observed in skin cancer, suggesting that there could be a role for loss-of-function mutations at the C-SH2/PTP interface in human diseases.

Activating SHP2_{PTP} disease mutants regulate SHP2-specific WPD loop motions

Gain-of-activity mutants in the SHP2_{PTP} construct reside within region IV, where their effects on the SHP2_{FL} construct are neutral or activating. While mildly activating SHP2_{PTP} mutants are widely distributed, the most activating mutants cluster around the WPD loop (Fig. 5a and

Supplementary Fig. 5a). WPD loop closure positions a key catalytic residue (D425) for catalysis. Thus, mutations in this cluster likely alter WPD loop dynamics and favor a WPD-closed conformation^{50,51}. In SHP2_{FL}, WPD loop closure is blocked in the auto-inhibited state by the steric hinderance from the N-SH2 domain (Supplementary Fig. 5b). Consequently, the activating effects of some mutants in region IV are less significant in the SHP2_{FL} context when compared to those disrupting auto-inhibition (region I).

In the MD trajectories of the active conformations of SHP2, we captured several instances of the WPD open-to-closed transition. These events reveal coordinated movements between key residues in this region and provide insights into how SHP2 may differ from other phosphatases, such as PTP1B (Fig. 5b–d). When SHP2 adopts the WPD open conformation, W423, P424, and P429 pack together, and this cluster is buttressed by F469 (Fig. 5b). Notably, F469 adopts multiple rotameric states in crystal structures and in our MD trajectories (Fig. 5d, e)⁵². In the transition to the WPD-closed conformation, F469 rotates away from the WPD loop, and W423 moves away from the loop center into the cleft that is otherwise occupied by F469 (Fig. 5c, d). This transition releases P424 from its packing interactions, allowing it to access more diverse conformations. We quantified these movements by measuring the distance between W423 C α 3 and P429 C γ , the χ_1 angle of F469, and the distance between P424 C γ and the V428 carbonyl. In our simulations, we observed instances of concerted W423/F469/P424 movements that ultimately culminated in WPD loop closure, as well as instances where these residues sampled on-pathway states and then reverted back to their inactive conformation, thereby aborting WPD loop closure. Our analysis shows that the F469 rotation acts as a gatekeeper for W423 movement and WPD loop closure (Fig. 5d).

Although a similar WPD loop motion has been observed before in the well-studied phosphatase PTP1B, the phenylalanine in PTP1B (F225) corresponding to SHP2 F469 only adopts one rotameric state that permits facile WPD loop opening and closing⁵³. F225 is locked in this orientation in by the adjacent L110, while in SHP2 this leucine is substituted by a less bulky threonine (T356), allowing for the gatekeeping function of F469 (Fig. 5e). Indeed, the T356L mutation in SHP2 is activating in our mutational scans and in biochemical assays, suggesting that F469 is held away from the WPD loop in this mutant (Supplementary Fig. 5c and Supplementary Data 2). Notably, in human classical PTPs, only SHP2 and MEG2/PTPN9 have this Phe-Thr combination, making the gatekeeping feature unique to these two phosphatases (Supplementary Fig. 5d).

Disease-associated mutations in the vicinity of the WPD loop can have a range of outcomes on SHP2 function. Mutations in this area that activate SHP2_{PTP} should promote WPD loop closure. For mutations just outside of the WPD loop that allosterically modulate loop dynamics, this activating effect is overridden by N-SH2 binding to the PTP domain. Potentially pathogenic mutations at F420, P432, G433, V435, and the gatekeeper F469 fall into this category (Supplementary Fig. 5a). F469S, found in colon cancer, should favor WPD loop closure, and indeed this mutation activates SHP2_{PTP} in our mutational scan (Supplementary Fig. 5e). Such mutations would only unleash their effects on signaling when bound to an activator phosphoprotein that displaces the N-SH2 domain. On the other hand, strongly activating mutations directly on the WPD loop can enhance PTP domain activity and also disrupt auto-inhibition. For example, all P424 mutants are activating in both SHP2_{PTP} and SHP2_{FL} screens, including brain cancer and Noonan Syndrome mutants P424S and P424L. In PTP1B, mutations at this conserved proline stabilize a WPD-closed conformation⁵¹, which would displace the N-SH2 domain from the auto-inhibitory interface in SHP2_{FL} (Supplementary Fig. 5b). Indeed, basal activity and melting curve measurements confirm that P424L activates SHP2 and destabilizes its auto-inhibited state (Supplementary Data 2 and Supplementary Fig. 3d). This dual-activating effect makes clinically observed mutants that are activating in both screens very likely to be

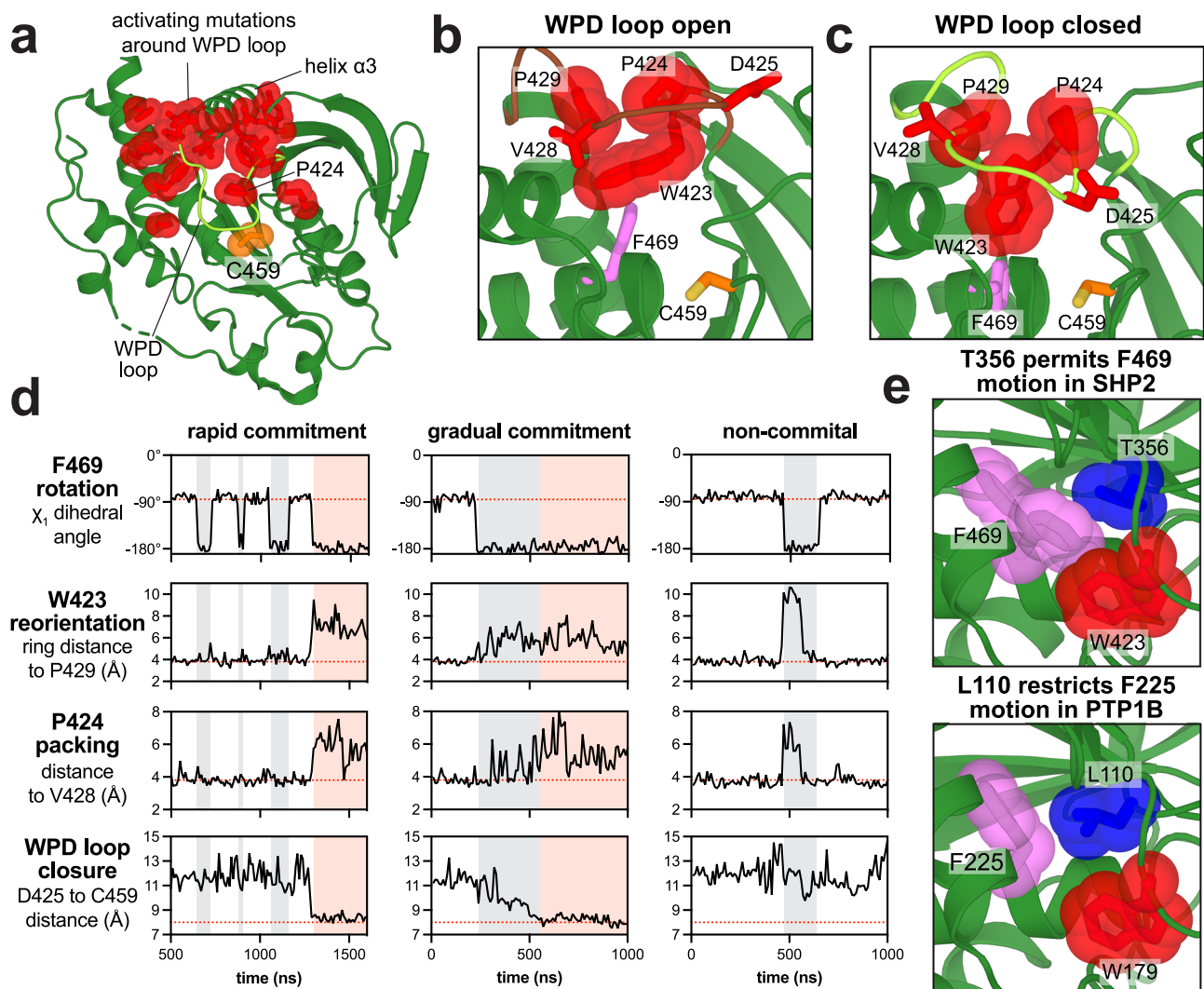


Fig. 5 | Mutational effects that alter WPD loop conformational dynamics.

a Residues with consistent gain-of-function mutational effects around the WPD loop (PDB code 3ZMO). **b** Frame from MD trajectory of SHP2 starting from PDB 6CRF, showing packing of W423, P424, P429, and F469 in WPD loop open conformation. F469 points towards the WPD loop in this structure. **c** Frame from an MD trajectory starting from PDB 6CRF, showing that, in the WPD-closed conformation, W423 displaces F469, and P424 is released from its intimate packing in the turn of the loop. **d** Three MD trajectories starting from the open conformation, showing coordinated movements between F469, W423, and P424. The first trajectory is an example of these movements culminating in WPD closure. The second trajectory also shows WPD loop closure, but more gradually. The third trajectory shows

coupled F469, W423, and P424 motions that revert back to their starting state, resulting in no WPD loop closure. F469 rotation is quantified by the χ_1 angle of this residue. W423 reorientation is quantified by the distance between W423 C ζ_3 and P429 C γ . P424 packing is quantified by the distance between P424 C γ and the V428 carbonyl. WPD loop closure is quantified by the distance between the D425 and C459 C α atoms. Gray shaded segments denote F469 in a permissive state for WPD loop closure. Pink shaded segments denote WPD loop closure. **e** Small residue T356 (blue) permits F469 rotation in SHP2, restricting W423 rotation and WPD loop closure (PDB code 3ZMO, top). Multiple F469 rotamers are modeled in this crystal structure. Bulky residue L110 (blue) constrains F225 rotation in PTP1B, permitting facile W179 rotation and WPD loop closure (PDB code 5K9V, bottom).

pathogenic. Intriguingly, the WPD loop cancer mutation V428M lies in region I instead of region IV (Supplementary Data 3), indicating that this mutation alters the WPD loop conformation in a way that decreases SHP2_{PTP} activity while still disrupting auto-inhibition in the SHP2_{FL} construct.

Pathogenic N-SH2 core mutants balance domain destabilization and phosphoprotein binding

Outside of the PTP domain, the N-SH2 domain is well-known to host disease mutations, most of which disrupt the auto-inhibitory interface. In addition to these interface mutations, the SHP2_{FL} mutational scan revealed a large cluster of activating mutations in the hydrophobic core of the N-SH2 domain (W6, F7, I11, A16, L20, F29, L30, L43 and V45) (Figs. 2d, 6a). We hypothesized that mutations at these sites disrupt hydrophobic packing, thereby destabilizing the N-SH2 domain and

weakening its auto-inhibitory binding to the PTP domain. Consistent with this hypothesis, polar mutations at these sites are more activating than nonpolar mutations, as are proline substitutions, which likely destabilize the central N-SH2 β -sheet. (Fig. 6b). Surprisingly, N-SH2 core mutants are rare in human diseases. Moreover, of the core mutants that can be accessed from wild-type SHP2 by a single nucleotide substitution, the few that are clinically observed tend to be modestly activating (Fig. 6c). For example, one of the few documented disease mutations in the N-SH2 core, L43F, only increases basal SHP2 activity three-fold (Supplementary Data 2). By contrast, the strongly activating L43P and L43R mutations have not been documented in clinical samples (Fig. 6c).

We hypothesized that hyper-destabilizing N-SH2 core mutations such as L43R activate SHP2 at the expense of phosphoprotein binding capability, which may be expendable in our yeast selection assay but

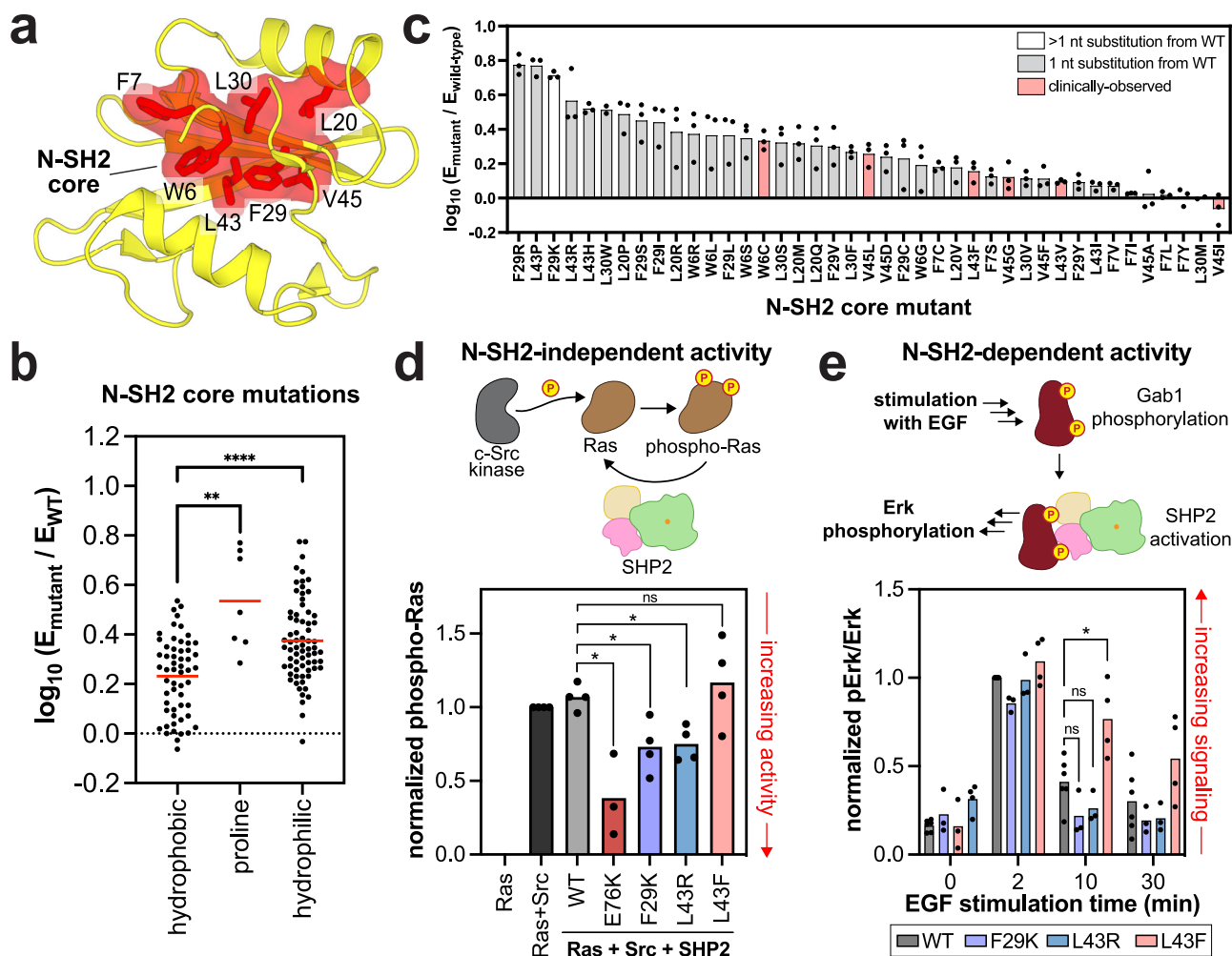


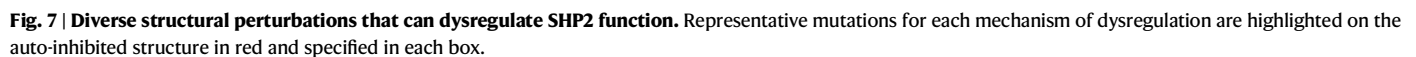
Fig. 6 | Functional effects and pathogenicity of mutations in the N-SH2 domain core. **a** Hydrophobic core residues in the N-SH2 domain (PDB code 1AYB). **b** Distribution of mutational effects at the residues shown in panel (a) in the SHP2_{FL} + v-Src_{FL} selection. Hydrophilic amino acids include C, D, E, H, K, N, Q, R, S, and T; hydrophobic amino acids include A, F, G, I, L, M, V, W, Y. The red line indicates the mean enrichment for all mutations in each group. Significance values are derived from a Welch's two-tailed unpaired *t*-test (hydrophobic vs proline *p* value = 0.0063, hydrophobic vs hydrophilic *p* value = 0.000002). **c** Functional effects of select N-SH2 core mutations in the SHP2_{FL} + v-Src_{FL} selection. Mutations observed in clinical databases (pink bars) are generally not the most activating, although many highly activating mutations are genetically accessible (1 nucleotide substitution away, gray bars). The SHP2_{FL} + v-Src_{FL} selection was used for panels b and c because it has a greater dynamic range over activating mutations than the

SHP2_{FL} + c-Src_{KD} selection. **d** N-SH2-independent catalytic activity of SHP2 N-SH2 core mutants and E76K in a Ras dephosphorylation assay in HEK 293 cells. Phospho-Ras levels are normalized to total Ras levels and to phospho-Ras in the Ras+Src sample (*n* = 4 replicate total and phospho-Ras measurements, *n* = 3 for E76K). Significance values are derived from a Welch's two-tailed unpaired *t*-test (wild-type vs E76K *p* value = 0.0427, wild-type vs F29K *p* value = 0.0246, wild-type vs L43R *p* value = 0.0062). **e** Downstream signaling activity of SHP2 N-SH2 core mutants, dependent on N-SH2 binding to Gab1, in HEK 293 cells stimulated with epidermal growth factor (EGF). Phospho-Erk levels are shown, normalized to total Erk levels and to the 2 min time point for the wild-type SHP2 sample (*n* = 6 for wild-type, *n* = 3 F29K and L43R, *n* = 4 for L43F). Significance values are derived from a Welch's two-tailed unpaired *t*-test (wild-type vs L43F, 10 min, *p* value = 0.0329).

not in the context of pathogenic human signaling pathways. We were unable to purify SHP2_{FL} proteins or isolate N-SH2 domains containing polar substitutions in the hydrophobic core, likely because these mutations are severely destabilizing. Thus, to test our hypothesis, we expressed and assayed the hyperactive F29K and L43R mutants alongside the mildly activating L43F mutant in two cellular assays for SHP2 activity, one that depends on N-SH2 function and one that does not. First, we examined the ability of the SHP2 mutants to dephosphorylate N-Ras that has been phosphorylated by co-expressed Src kinase⁵⁴. This activity is not thought to depend on SH2 interactions (Fig. 6d). Both F29K and L43R dephosphorylated N-Ras more efficiently than wild-type SHP2, whereas L43F was comparable to wild-type SHP2, consistent with the mutational scanning data (Fig. 6d and Supplementary Fig. 6a).

Next, we tested the signaling capabilities of these mutants in a context where N-SH2 binding functions are operative. SHP2 was co-

expressed with Gab1, a known binding partner, cells were stimulated with the epidermal growth factor, and Erk phosphorylation was monitored as a downstream marker of SHP2-mediated activation of this pathway²². We previously showed that canonical hyperactivating mutations in SHP2, like E76K, dramatically enhance Erk phosphorylation in this assay²². By contrast, the F29K and L43R mutants, which also activate SHP2, did not enhance Erk phosphorylation over wild-type SHP2, despite their increased catalytic activities (Fig. 6e and Supplementary Fig. 6b). The L43F disease mutant, on the other hand, modestly enhanced Erk phosphorylation. This mutant is destabilized but can be purified (Supplementary Fig. 6c, d), and it is still competent to bind phosphoproteins²². These results suggest that while the more disruptive N-SH2 core mutants display higher basal activity, their N-SH2 domains are too unstable to facilitate interactions with binding partners. Consequently, these mutants are unable to drive pathogenic signaling in human diseases. This balance between the disruption of



In this study, we used a yeast selection assay to compare the catalytic activities of nearly all possible point mutants in the human tyrosine phosphatase SHP2. We conducted deep mutational scans of both full-length SHP2 (593 residues, ~12,000 variants) and its isolated catalytic domain (289 residues, ~6000 variants). Thus far, our understanding of SHP2 regulation has largely been driven by the characterization of 10–20 disease-associated mutations, most of which cluster at the N-SH2/PTP interface^{23,24,45}. For some of these well-studied disease mutants that are located on the phosphatase domain, comparative characterization of the full-length SHP2 constructs and the isolated PTP domains have allowed for rigorous dissection of the contributions of auto-inhibition and intrinsic phosphatase activity to overall enzyme dysregulation^{24,42,45}. Our mutational scanning approach adds new depth to these analyses by enabling comparative mechanistic

dissection across all PTP domain residues. More broadly, our datasets provide functional and mechanistic information on roughly 600 clinically observed SHP2 mutants, some within and some outside of the PTP domain, the vast majority of which were previously uncharacterized.

We focused our analysis on sets of mutations with significant but divergent activating or deactivating effects in either full-length SHP2 or the isolated PTP domain (Fig. 2g). While our experiments do not directly assess pathogenicity, we conclude that mutants in each set dysregulate SHP2 via distinct mechanisms and are likely to affect SHP2-mediated signaling pathways. It is noteworthy that some of the new mechanisms of dysregulation proposed in our study involve rare clinical variants, which is likely why these mechanisms have not been described to date. Further studies of the phenotypic effects of these mutations in mammalian cells and animal models will provide more valuable insight into their potential pathogenicity. On the other hand, clinical variants with a neutral effect on basal catalytic activity, which our method primarily measures, are not necessarily benign and could also warrant further investigation. For example, we previously reported that the pathogenic N-SH2 mutant T42A alters phosphoprotein binding specificity, thereby sensitizing SHP2 to activation by specific proteins²². Moreover, our mutational scans captured the activating effects of pathogenic mutants that lack clear mechanistic explanations, such as the N-SH2/C-SH2 linker mutation D106A, and C-SH2 mutations R138Q and E139D (Fig. 2 and Supplementary Fig. 2). Alternative analytical methods will be required to dissect their mechanisms of dysregulation.

Juxtaposition of our mutational scanning datasets with MD simulations allowed us to gain insights into the dynamics of two critical regions of SHP2: the C-SH2/PTP interface and the WPD loop. These analyses shed light on how these regions contribute to SHP2 regulation and pathogenicity. They also demonstrate how SHP2 diverges in its regulation from other PTP family members. For example, the role of T356 in governing WPD loop movement only exists in SHP2 and potentially PTPN9 (Supplementary Fig. 5d), and the C-SH2/PTP interface of SHP2, which governs activation and active state stability, diverges from that of its paralog SHP1 (Supplementary Fig. 7)⁵⁵. These variations in regulation and dynamics likely drive the broad range of catalytic activity and signaling functions seen across the PTP family. Our data could also be used to guide therapeutic strategies for specific SHP2 mutants. Inhibition of SHP2 by small molecule binding to a “tunnel site” between the C-SH2 and PTP domains is currently the most broadly applicable strategy for SHP2 inhibition⁵⁶. However, mutations at this binding site and mutations that strongly disrupt auto-inhibition can confer resistance to tunnel site binders²⁰. The mechanistic distinctions gleaned from our study could motivate efforts to target alternative sites or pursue dual allosteric inhibition^{57,58}.

Finally, we note that our selection platform reports solely on the phosphatase activities of SHP2 variants, which do not fully reflect their signaling roles in a native context or take into account potential post-translational modifications. Selection assays with more signaling-relevant readouts could potentially be developed in a mammalian cell system to evaluate these properties of SHP2⁵⁹. On the other hand, the direct activity readout of our yeast selection assay should make it readily adaptable to other protein tyrosine phosphatases with minimal optimization^{37,38}. We envision that this platform, particularly when coupled with MD simulations, experimental biochemistry/biophysics, and cell signaling studies, will become an invaluable tool to map sequence-structure-function relationships across this important enzyme family.

Methods

General note about library construction and sequencing

Information about boundaries for the mutational scanning tiles, the sequences of all variants in the oligonucleotide pools used to make the

libraries, and the sequences of all primers used to construct, amplify, and deep sequence the tiles can be found in Supplementary Data 1.

SHP2 mutagenesis library preparation

Saturation mutagenesis libraries of SHP2 were prepared with the Mutagenesis by Integrated Tiles (MITE) method³⁹. The 1782 bp (593AA + stop codon) full-length SHP2 gene we used was first optimized to yeast-favorable codons. Then, the full-length sequence was divided into 15 separate tiles, each spanning around 40 amino acids. Two saturation mutagenesis oligo pools of alternating tiles were designed and acquired from Twist Bioscience. Each single amino acid substitution was encoded by a single oligo, thus there is no degeneracy in the library. The oligo sequences include invariant overhang sequences on each end designed for Gibson assembly to replace the wild-type sequence of each tile of a pET28 plasmid containing the yeast-optimized human SHP2_{FL} coding sequence. With PCR primers annealing to their overhang sequences, individual tiles were amplified from their oligo pool. Backbone amplification primers complementary to the tile amplification primers were used to amplify Gibson backbone DNA for each tile from the wild-type pET28 yeast-optimized SHP2_{FL} plasmid. Tile mutagenesis library inserts were then cloned onto their corresponding backbones using Gibson assembly to generate 15 separate plasmid libraries, each containing all single mutants of one tile region. For each tile, the SHP2 full-length construct was PCR amplified with primers annealing right outside of the SHP2 gene on the plasmid, carrying overhangs for homologous recombination onto the yeast expression plasmid PWJ1781. For tiles 7–13 in the PTP domain, isolated PTP domain constructs spanning residues 235–539 were also amplified with homologous recombination overhangs.

Yeast expression plasmids construction and transformation

For expression of the kinase and phosphatase libraries in yeast, we used a galactose-induced expression plasmid PWJ1781⁶⁰. For the purposes of double transformation in yeast, we switched the LEU2 marker to a URA3 marker to generate PWJ1781-URA. v-Src_{FL} and c-Src_{KD} DNA sequences were then integrated into PWJ1781-URA by Gibson assembly to make kinase expression plasmids PWJ1781-URA-c-Src_{KD} and PWJ1781-URA-v-Src_{FL}. SHP2 expression plasmids were constructed with homologous recombination in yeast for both the SHP2_{FL} construct and SHP2_{PTP}. In both cases, inserts bearing each mutagenized tile were integrated into PWJ1781 separately. PWJ1781 was first digested with HpaI to yield linear backbone DNA. Then, a molar ratio of 1:2 of digested backbone and amplified insert was co-transformed into yeast strain YPH499 (ATCC) with the LiAc/PEG/ssDNA method⁶¹. Specifically, 2 µg of the two DNA pieces, in total, were transformed into 3×10^8 YPH499 cells. The transformed cells were then grown in 500 mL synthetic complete media without leucine supplemented with 4% glucose at 30 °C with shaking for ~40 h to reach a high concentration (OD_{600} ~6–10, $6\text{--}10 \times 10^7$ cells/mL). The cells from each transformation were collected as stock for one tile. Cells from each stock were directly subjected to kinase transformation. In the kinase plasmid transformation, 2 µg of PWJ1781-URA-c-Src_{KD} and PWJ1781-URA-v-Src_{FL} was transformed into 6×10^7 YPH499 cells bearing phosphatase plasmid with the LiAc/PEG/ssDNA method. The doubly transformed cells were grown in 100 mL of synthetic complete media without leucine and uracil, supplemented with 4% glucose at 30 °C with shaking for ~28 h to reach a high OD (OD_{600} ~4, 4×10^7 cells/mL). Each transformation yielded a cell stock transformed with one of the v-Src_{FL} or c-Src_{KD} expression plasmids and one of the SHP2_{FL} or SHP2_{PTP} plasmids bearing one mutagenized tile. Each stock was then subjected to out-growth and selection individually.

Tile library selection with the yeast growth assay

Each cell stock bearing one kinase expression plasmid and a SHP2 library with one variable tile was inoculated into synthetic complete

media without uracil and leucine, supplemented with glycerol lactate (2% lactic acid, 3% glycerol, 0.05% glucose in media) at a starting $OD_{600} = 0.1$. The culture was grown at 30 °C with shaking for 16 h and then used to inoculate synthetic complete media without uracil and leucine, supplemented with 4% galactose for co-expression and selection. The rest of the cells grown in glycerol lactate media were harvested as unselected samples. For each outgrowth culture, two parallel selection cultures were made with a starting $OD_{600} = 0.05$. The cultures were allowed to grow at 30 °C with shaking for 24 h, and the cells were then harvested as post-select samples.

Deep sequencing

For cell stocks with each tile before and after selection, plasmid DNA was extracted with the Zymoprep Yeast Plasmid Miniprep II kit from $\sim 1 \times 10^8$ cells. Each mutagenized tile DNA library was PCR amplified from its corresponding SHP2 plasmid with tile-specific primers bearing overhangs for the addition of Illumina sequencing adapters. The PCR mix was directly followed by another round of PCR amplification with Illumina barcoding primers to append Illumina sequencing adapters and 5' and 3' indices (D700 and D500 series primers). The PCR products were gel-purified and quantified with QuantiFluor® dsDNA System (Promega). Then, samples were pooled at ratios that all mutants are theoretically equally represented, and the pooled libraries were sequenced on a MiSeq using V2 300-cycle reagent kits. Each sequencing run contained no more than 24 samples to ensure good read counts.

Sequencing data analysis

Paired-end sequencing reads were first merged with FLASH⁶², followed by trimming with Cutadapt to remove constant sequences outside of the mutagenized libraries⁶³. Then, read counts for each mutant in the libraries were calculated using in-house Python scripts (<https://doi.org/10.5281/zenodo.15304851>). For each sequenced library, frequencies of the mutants (f_{mut}) were first calculated by taking the ratio of the mutants' read counts (n_{mut}) over total reads (n_{total}) in the library (equation 1). Then, enrichment (E_{mut}) were calculated through dividing after selection ($f_{selected}$) by frequencies before selection ($f_{unselected}$) (Eq. 2). The enrichment scores data that we are presenting on the heat maps are \log_{10} -transformed enrichment normalized to WT (Eq. 3). The distribution of enrichment scores from selection in the SHP2_{FL} context was fit to a Gaussian distribution, and datapoints outside of $\mu \pm 2\sigma$ were defined as significantly gain- or loss-of-activity. The distribution of enrichment scores from selections in the SHP2_{TP} context followed a bimodal distribution and was fit to the sum of two Gaussians. Datapoints outside of $\mu \pm 2\sigma$ for the distribution around 0 were defined as significantly gain- or loss-of-activity.

$$f_{mut} = \frac{n_{mut}}{n_{total}} \quad (1)$$

$$E_{mut} = \frac{f_{selected}}{f_{unselected}} \quad (2)$$

$$Score_{mut} = \log_{10} E_{mut} - \log_{10} E_{WT} \quad (3)$$

Purification of full-length SHP2 proteins

pET28-His-TEV plasmid encoding the human SHP2_{FL} sequence (human cDNA) was used for QuikChange mutagenesis to generate SHP2_{FL} mutants²². For purification of wild-type and mutant SHP2_{FL} constructs, plasmids were first transformed into chemically competent BL21(DE3) cells and grown on LB agar plates supplemented with 50 µg/mL kanamycin. Then, colonies scraped off the plates were inoculated into 100 mL LB with kanamycin and grown at 37 °C to an $OD_{600} = 1$. About 50 mL of this culture was used to inoculate 1 L LB cultures with

kanamycin at a starting $OD_{600} = 0.05$, and the 1 L cultures were incubated at 37 °C until their OD_{600} reached 0.5. About 0.5 mM IPTG was supplemented to induce expression, and the expression cultures were grown overnight at 18 °C. The cultures were then spun down at $4000 \times g$ for 30 min, and resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol, and freshly added 2 mM β -mercaptoethanol). Before lysis, the cell suspensions were supplemented with protease inhibitors (200 µM AEBSF, 20 µM Leupeptin, 1 µM Pepstatin A). The cell suspensions were lysed using sonication (Fisherbrand Sonic Dismembrator), and spun down at $14,000 \times g$ for 45 min. The His-tagged SHP2_{FL} constructs were extracted from the supernatant with a 5 mL Ni-NTA column (Cytiva). The column was subsequently washed with 50 mL lysis buffer and 50 mL wash buffer (50 mM Tris pH 8.5, 50 mM NaCl, 20 mM imidazole, 10% glycerol, and freshly added 2 mM β -mercaptoethanol), followed by elution of the tagged SHP2 with a mixture of 25 mL wash buffer + 25 mL elution buffer (50 mM Tris pH 8.5, 50 mM NaCl, 500 mM imidazole, 10% glycerol) directly onto a 5 mL HiTrap Q Anion exchange column (Cytiva). The Q column was washed once with 40 mL anion exchange buffer A (50 mM Tris pH 8.5, 50 mM NaCl, 1 mM TCEP), and the protein was eluted off with a salt gradient between Anion A buffer and Anion B buffer (50 mM Tris pH 8.5, 1 M NaCl, 1 mM TCEP). The eluted protein fractions were collected and cleaved with 0.10 mg/mL His₆-tagged TEV protease at 4 °C overnight to remove the His tag. The cleavage mixture was applied through 2 mL of Ni-NTA gravity column (Thermo Fisher) to remove uncleaved protein and TEV protease, and the flow through was concentrated to less than 1 mL. Finally, the concentrated protein solution was loaded onto a Superdex 200 10/300 gel filtration column (Cytiva) equilibrated with SEC buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP, and 10% glycerol) for size exclusion purification. Pure fractions were pooled, concentrated, and flash frozen in liquid N₂ for storage at -80 °C.

Purification of SHP2 SH2 domains

pET28-His6-SUMO-N-SH2-Avi and pET28-His6-SUMO-C-SH2-Avi plasmids in our lab were used as templates for wild-type and mutant SH2 domain purification²². The Avi tags were first removed in one cloning step, and the resulting pET28-His6-SUMO-N/C-SH2 plasmids were applied to QuikChange mutagenesis to generate the desired SH2 mutants. For purification of wild-type and mutant SH2 constructs, plasmids were first transformed into chemically competent BL21(DE3) cells and grown on LB agar plates supplemented with 50 µg/mL kanamycin. Then, colonies scraped off the plates were inoculated into 100 mL LB with kanamycin and grown at 37 °C to an $OD_{600} = 1$. About 50 mL of this culture was used to inoculate 1 L LB cultures with kanamycin at a starting $OD_{600} = 0.05$, and the 1 L cultures were incubated at 37 °C until their OD reached 0.5. About 0.5 mM IPTG was supplemented to induce expression, and the expression cultures were grown overnight at 18 °C. The cultures were then spun down at $4000 \times g$ for 30 min, and resuspended in lysis buffer (50 mM Tris pH 7.5, 300 mM NaCl, 20 mM imidazole, 10% glycerol, and freshly added 2 mM β -mercaptoethanol). Before lysis, the cell suspensions were supplemented with protease inhibitors (200 µM AEBSF, 20 µM Leupeptin, 1 µM Pepstatin A). The cell suspensions were lysed using sonication (Fisherbrand Sonic Dismembrator), and spun down at $14,000 \times g$ for 45 min. The His-tagged SH2 constructs were extracted from the supernatant with a 5 mL Ni-NTA column (Cytiva). The column was subsequently washed with 50 mL lysis buffer and 50 mL wash buffer (50 mM Tris pH 7.5, 50 mM NaCl, 20 mM imidazole, 10% glycerol, and freshly added 2 mM β -mercaptoethanol), followed by elution of the tagged SH2 with a mixture of 25 mL wash buffer + 25 mL elution buffer (50 mM Tris pH 7.5, 50 mM NaCl, 500 mM imidazole, 10% glycerol) directly onto a 5 mL HiTrap Q Anion exchange column (Cytiva). The Q column was washed once with 40 mL anion exchange buffer A (50 mM Tris pH 7.5, 50 mM NaCl, 1 mM TCEP), and the protein was eluted off

with a salt gradient between Anion A buffer and Anion B buffer (50 mM Tris pH 7.5, 1 M NaCl, 1 mM TCEP). The eluted protein fractions were collected and cleaved with 0.05 mg/mL His₆-tagged Ulp1 protease at 4 °C overnight to remove the His tag. The cleavage mixture was applied through 2 mL of Ni-NTA gravity column (Thermo Fisher) to remove uncleaved protein and TEV protease, and the flow through was concentrated to less than 1 mL. Finally, the concentrated protein solution was loaded onto a Superdex 75 16/600 gel filtration column (Cytiva) equilibrated with SH2-SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP, and 10% glycerol) for size exclusion purification. Pure fractions were pooled and concentrated, and flash frozen in liquid N₂ for long-term storage at −80 °C.

SHP2 basal activity measurements

Basal activities of wild-type and mutant SHP2 were measured against the fluorogenic substrate 6,8-difluoro-4-methylumbelliferyl phosphate (DiFMUP). Initial DiFMUP dephosphorylation rates by SHP2 variants were measured at 37 °C. Reactions were done in black polystyrene flat-bottom half-area 96-well plates at a working volume of 50 µL. Initial rates of each protein were measured in a set of three replicates. In each replicate, the reaction mix contains a fixed concentration of protein (see below) and a DiFUP concentration series of 4000, 2000, 1000, 500, 250, 125, 62.5, and 31.25 µM. With each plate, the absorbance of the dephosphorylation product DiFMU at a concentration series of 200, 100, 50, 25, 12.5, 6.25, 3.125, and 0 µM was measured as a standard curve to convert absorbance values to product concentrations. Reactions were started by the addition of the protein, and emitted fluorescence at 455 nm was measured every 25 s within 50 min with a BioTek Synergy Neo2 multi-mode reader. Fluorescence values were converted into DiFMU concentrations, and initial rates were determined by the slope of the first 5 min on the reaction curves. All the initial rates were fitted onto Michaelis–Menten curves using GraphPad Prism to determine k_{cat} and K_{M} values (Supplementary Data 2)⁶⁴.

Melting temperature measurements via differential scanning fluorimetry (DSF)

Melting temperature measurements were conducted in DSF buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 0.4% DMSO) on MicroAmp Fast Optical 96-well Reaction plates (Applied Biosystems, #4346906) at working volumes of 20 µL. The mixtures contained 10 µM protein and 25x SYPRO Orange Protein Gel Stain (Thermo Fisher, catalog no. S-6650). Melting curves were measured on an Applied Biosystems Step-One Plus RT-PCR thermocycler between 25 and 95 °C with a gradient of +0.5 °C per minute (excitation: 472 nm; emission: 570 nm). Fluorescence reads under each temperature were analyzed using DSFworld and melting temperatures were calculated with dRFU⁶⁵.

Description of mammalian expression vectors

The full-length construct of SHP2 was amplified from the pGEX-4TI SHP2 WT plasmid (Addgene plasmid #8322)⁶⁶ and cloned into the mammalian pEF expression vector with an N-terminal Myc-tag. The Gab1 gene was amplified from FRB-GFP-Gab1(Y628F/Y660F) (Addgene #188658)⁶⁷ and cloned into a pEF vector with an N-terminal 3x-FLAG tag. The native phosphorylation sites (Y628 and Y660) were reverted from phenylalanine to tyrosine by site-directed mutagenesis. For c-Src, we cloned a constitutively active version starting from the pCMV5-mouse-src plasmid (Addgene #13663) by deleting the auto-inhibitory tail (Δ528–535). HA-tagged N-Ras was expressed using the pCGN HA-linker-N-Ras wt plasmid (Addgene #14723)⁶⁸.

General cell culture

HEK 293 cells (ATCC) were grown at 37 °C with 5% CO₂ in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS).

EGF stimulation experiments

About 2.2×10^6 HEK 293 cells were seeded in a 10 cm plate. The next day, cells were transfected overnight with 5 µg Gab1 and 5 µg SHP2 using 30 µg of polyethylenimine in 1 mL DMEM. The transfection media was refreshed, and cells were serum-starved for 24 h in DMEM. Cells were harvested by scraping, washed three times in 1 mL PBS at room temperature. Prior to stimulation, an aliquot of cells was taken as an unstimulated control ($t = 0$). Cells were then resuspended in 25 ng/mL EGF in preheated PBS and placed in a 37 °C heat block. Aliquots were taken at 2, 10, and 30 min; placed on ice and spun down in a 4 °C tabletop centrifuge at $1000 \times g$ for 5 min. Supernatant was aspirated and cells were lysed in 75 µL lysis buffer (20 mM Tris pH 8.0, 137 mM NaCl, 2 mM EDTA, 10% glycerol, 0.5% NP-40, with freshly added phosphatase- and protease inhibitors) for 25 min on ice. Lysates were spun down for 15 min at $17,000 \times g$ in a 4 °C tabletop centrifuge. Supernatant was used in a bicinchoninic acid (BCA) assay to determine protein concentration. About 15 µg of total protein was loaded on a 12% acrylamide gel and transferred to a 0.45-µm nitrocellulose membrane using the StandardSD protocol on the Bio-Rad Trans-Blot Turbo. Membranes were blocked for 1 h at room temperature using 5% bovine serum albumin (BSA) in TBS. Primary antibodies were stained overnight at 4 °C (Erk 1:1000, p-Erk 1:2000, Vinculin 1:1000, Myc 1:5000, FLAG 1:5000) in 5% BSA in TBST. Membranes were washed three times in 5 mL TBST for 5 min each. Secondary antibodies were incubated in 5% BSA in TBST for 1 h at room temperature (1:10,000). Membranes were imaged on a LiCor Odyssey, and bands were quantified using ImageStudio.

Ras dephosphorylation assay (Rassay)

About 0.8×10^6 HEK 293 cells were seeded in a 6-cm plate. The next day, cells were transfected with Ras, Ras and Src, or Ras, Src and SHP2; to a total of 3 µg (a pEF vector with no open reading frame was used to make up the difference between conditions) in 300 µL DMEM with 9 µg of polyethylenimine. The transfection medium was refreshed the next morning and replaced with warm DMEM with 10% FBS. 36 h after transfection, cells were harvested by scraping and washed three times in 1 mL cold PBS. Cells were lysed in 150 µL lysis buffer (20 mM Tris pH 8.0, 137 mM NaCl, 2 mM EDTA, 10% glycerol, 0.5% NP-40, with freshly added phosphatase- and protease inhibitors) for 25 min on ice. Lysates were spun down for 15 min at $17,000 \times g$ in a 4 °C tabletop centrifuge. Supernatant was used in a BCA assay to determine protein concentration. About 80 µg of protein was used in an immuno-precipitation (IP) with 5 µg of packed Pierce anti-HA magnetic beads (Fisher, #88836) in a total volume of 350 µL. Samples were incubated at 4 °C overnight while rotating. The next morning, beads were washed three times on a magnetic rack using 1 mL of lysis buffer, and finally resuspended in 65 µL of 1x Laemmli buffer. All samples were boiled at 100 °C for 8 min, and 15 µg of total protein (total cell lysate) or 15 µL of each sample (IP) was loaded onto a 12% acrylamide gel. Proteins were transferred to a 0.45-µm nitrocellulose membrane using the Standard SD protocol on the Bio-Rad Trans-Blot Turbo. Membranes were blocked for 1 h at room temperature using 5% BSA in TBS. Primary antibodies were stained for 2 h at room temperature (TCL: Src 1:1000, β-actin 1:5000, Myc 1:5000, HA 1:1000; IP: HA 1:1000, pTyr 1:2000) in 5% BSA in TBST. Membranes were washed three times in 5 mL TBST for 5 min each. Secondary antibodies were incubated in 5% BSA in TBST for 1 h at room temperature (1:10,000). Membranes were imaged on a LiCor Odyssey, and bands were quantified using ImageStudio. The specific antibodies used for this study are listed in Table 1.

Analysis of COSMIC, TCGA, and ClinVar databases

We extracted clinically observed SHP2 mutant data from COSMIC¹², TCGA¹³, and ClinVar²⁵ (Supplementary Data 3). COSMIC and ClinVar data were downloaded in September 2023. TCGA data was

downloaded in October 2024. All disease mutations extracted from the three databases combined are annotated as “clinically observed” in our analysis. Cancer mutations in TCGA and COSMIC are classified as “all cancer mutations”, and cancer mutations with more than five reports are classified as “high-frequency cancer mutations”. Pathogenic mutations and variants of uncertain significance annotated in ClinVar are also separately classified, and enrichment scores of mutations in all categories were plotted to visualize their activity distributions (Fig. 3b, e). Mutants in COSMIC are also classified based on tissue type, and enrichment scores of variants in each cancer type were plotted as well (Fig. 3c, f).

Molecular dynamics simulations

Preparation of structural models for simulations. We built and simulated near full-length SHP2 (residues 1–528) in the closed, auto-inhibited state, and in two active conformational states. For all three states, we considered the wild-type protein, as well as the protein with Glu 76 mutated to Lys. For the simulations of the protein in the auto-inhibited state, we used the crystal structure 4DGP as the starting structure²⁴. Missing residues were modeled as follows. Residues 1–3 at the N-terminal end were built using PyMOL⁶⁹. Residues 235–245 were taken from a model of the full-length SHP2 that was predicted by AlphaFold2⁷⁰, since this region is missing in all crystal structures. Residues 293–303 were taken from the crystal structure 6CRF²⁰ and residues 314–232 were taken from the crystal structure 4RDD⁷¹. For simulations of the protein with the E76K mutation, Glu 76 was mutated to Lys using PyMOL. Crystalline waters were retained in the starting structure. For the protein in the first open, active state, we used the crystal structure 6CRF as the starting structure²⁰. In this structure, Glu 76 is mutated to Lys, so for the wild-type simulations, this residue was mutated to Glu using PyMOL. Missing residues 89–93, 140–145, 154–166, and 203–209 were taken from 4DGP²⁴. Residues 237–244 were taken from a model generated by AlphaFold2 (Uniprot ID Q06124)⁷⁰. Residues 313–324 were taken from 4RDD⁷¹, and missing

C-terminal residues 526–528 were built in using PyMOL. The alternative active state model was generated by AlphaFold2, by inputting the wild-type SHP2 sequence (Uniprot ID Q06124, residues 1–529) into ColabFold with the default settings⁷², then residue 529 was manually removed for consistency with the other starting structures. The N-SH2 domains are displaced in the AlphaFold2 structure when compared to the crystal structure (PDB code 6CRF), but they are otherwise similar. Crystalline waters are present in 6CRF. These waters were retained in the simulations starting from the crystal structure and were added to the AlphaFold2 structures. Cys 459 is deprotonated in all systems. Both N- and C-termini were capped with acetyl and amide groups, respectively, in all systems. Details of the simulated conditions are given in Table 2.

Simulation protocol. Each system was solvated with TIP3P water⁷³, and ions were added such that the final ionic strength of the system was 100 mM using the tleap package in AmberTools22⁷⁴. The energy of each system was minimized first for 5000 steps while holding the protein atoms and crystalline waters fixed, followed by minimization for 5000 steps while allowing all the atoms to move. Following minimization, three individual trajectories were generated for each system, with distinct initial velocities for each. The temperature of each system was raised in two stages—first to 100 K over 0.5 ns and then to 300 K over 0.5 ns. The protein atoms and crystalline waters were held fixed during the heating stage. Each system was then equilibrated for 2 ns, followed by production runs. Three production trajectories, each 2.5-μs long, were generated for each system. All equilibration runs and production runs were performed at constant temperature (300 K) and pressure (1 bar). The simulations were carried out with the Amber package⁷⁵ using the ff14SB force field for proteins⁷⁶ using an integration timestep of 2 fs. The particle mesh Ewald approximation was used to calculate long-range electrostatic energies⁷⁷. All hydrogens bonded to heavy atoms were constrained with the SHAKE algorithm⁷⁸. The Langevin thermostat was used to control the temperature with a collision frequency of 1 ps⁻¹. Pressure was controlled while maintaining periodic boundary conditions.

Analyses. MD trajectories were compiled from the raw data using the CPPTRAJ module of AmberTools22⁷⁴. Structures were extracted from the trajectories in both 1 and 10 ns increments for analysis and visualization. All measurements and calculations were done using the PDB module in Biopython using in-house scripts (<https://doi.org/10.5281/zenodo.15304851>)⁷⁹. For most calculations reported in the main text, trajectories were sampled every 10 ns. The measurements from all three replicates of each system were combined to determine the reported distributions. In cases where distance calculations involved a redundant atom (e.g., distances between two possible nitrogens and two possible oxygens in a Glu/Arg ion pair), all combinations of distance measurements were calculated, then the shortest distance at each frame was determined and used for the distribution plots. For visualization, trajectories were sampled every 10 ns. All structure visualization and rendering in this study was done using PyMOL⁶⁹.

Table 1 | Antibodies used for western blot analysis in this study

Description	Vendor	Catalog #
Vinculin	CST	13901S
FLAG (Clone 2B3C4)	Fisher	50-173-6451
Myc (Clone 9E10)	Invitrogen	R95025
p44/42 (Erk1/2) (L34F12)	CST	4696S
p-p44/42 (p-Erk1/2) (197G2)	CST	13214S
c-Src	CST	2123S
B-actin	Sigma	A5441
HA (TCL)	Sigma	SAB5600116
Phosphotyrosine (P-Tyr-1000) MultiMab	CST	8954S
HA (IP)	Sigma	11867423001
IRDye® 680RD Goat anti-Rabbit IgG	LiCor	926-68071
IRDye® 800CW Goat anti-Mouse IgG	LiCor	926-32210
IRDye 800CW Goat anti-Rat IgG	LiCor	926-32219

Table 2 | System and box sizes for the six different simulated conditions

Starting structure	Simulation box dimensions (Å)	Total number of atoms	Total number of water molecules	Salt concentration (mol/L)
4DGP _{WT}	120.5 × 120.6 × 120.6	163,815	51,714	0.1
4DGP _{E76K}	120.5 × 120.6 × 120.6	163,832	51,718	0.1
6CRF _{WT}	120.7 × 120.8 × 120.7	164,436	51,921	0.1
6CRF _{E76K}	120.8 × 120.7 × 120.7	164,684	52,002	0.1
AlphaFold2 _{WT}	116.1 × 116.1 × 115.9	146,273	45,876	0.1
AlphaFold2 _{E76K}	116.1 × 116.1 × 115.9	146,164	45,838	0.1

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw sequencing reads from this study are accessible via the NCBI Sequence Read Archive under accession number [PRJNA1232161](https://doi.org/10.5061/dryad.83bk3jb18). Deep sequencing data, molecular dynamics starting structures and trajectories, and all files relating to the AlphaFold2 model for the alternate open state of SHP2 can be found in a Dryad repository [<https://doi.org/10.5061/dryad.83bk3jb18>]. The following structures from the Protein Data Bank were used for structural analysis and production of figures: 4DGP, 6CRF, 3ZM0, 5K9V, 1AYB, 2B3O, and 3PS5. Unless otherwise stated, all data supporting the results of this study can be found in the article, supplementary, and source data files. Source Data are provided with this paper. There are no restrictions to access and use the data associated with this publication. Source data are provided with this paper.

Code availability

Scripts to process the deep sequencing data and calculate distances and angles in molecular dynamics trajectories can be found via Zenodo [<https://doi.org/10.5281/zenodo.15304851>]. There are no restrictions to access and use the code associated with this publication.

References

- Yu, H. et al. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933–945 (1994).
- Zhou, S. et al. SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**, 767–778 (1993).
- Newton, A. C. Protein kinase C: structure, function, and regulation. *J. Biol. Chem.* **270**, 28495–28498 (1995).
- Schmedt, C. et al. Functional characterization of the RNA-binding domain and motif of the double-stranded RNA-dependent protein kinase DAI (PKR). *J. Mol. Biol.* **249**, 29–44 (1995).
- Lim, W. A. The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr. Opin. Struct. Biol.* **12**, 61–68 (2002).
- Huang, Q., Li, M., Lai, L. & Liu, Z. Allostery of multidomain proteins with disordered linkers. *Curr. Opin. Struct. Biol.* **62**, 175–182 (2020).
- Papaleo, E. et al. The role of protein loops and linkers in conformational dynamics and allostery. *Chem. Rev.* **116**, 6391–6423 (2016).
- Nussinov, R., Tsai, C.-J. & Jang, H. Autoinhibition can identify rare driver mutations and advise pharmacology. *FASEB J.* **34**, 16–29 (2020).
- Barford, D. & Neel, B. G. Revealing mechanisms for SH2 domain mediated regulation of the protein tyrosine phosphatase SHP-2. *Structure* **6**, 249–254 (1998).
- Grossmann, K. S., Rosário, M., Birchmeier, C. & Birchmeier, W. in *Advances in Cancer Research* (eds Vande Woude, G. F. & Klein, G.) Ch. 2 (Academic Press, 2010).
- Mohi, M. G. & Neel, B. G. The role of Shp2 (PTPN11) in cancer. *Curr. Opin. Genet. Dev.* **17**, 23–30 (2007).
- Sondka, Z. et al. COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res.* **52**, D1210–D1217 (2024).
- The Cancer Genome Atlas Research Network. et al. The cancer genome atlas Pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
- Ruess, D. A. et al. Mutant KRAS-driven cancers depend on PTPN11/SHP2 phosphatase. *Nat. Med.* **24**, 954–960 (2018).
- Bentires-Alj, M. et al. A role for the scaffolding adapter GAB2 in breast cancer. *Nat. Med.* **12**, 114–121 (2006).
- Aceto, N. et al. Tyrosine phosphatase SHP2 promotes breast cancer progression and maintains tumor-initiating cells via activation of key transcription factors and a positive feedback signaling loop. *Nat. Med.* **18**, 529–537 (2012).
- Bard-Chapeau, E. A. et al. Ptpn11/Shp2 acts as a tumor suppressor in hepatocellular carcinogenesis. *Cancer Cell* **19**, 629–639 (2011).
- Chong, P. S. Y. et al. IL6 promotes a STAT3-PRL3 feedforward loop via SHP2 repression in multiple myeloma. *Cancer Res.* **79**, 4679–4688 (2019).
- Cai, P. et al. Expression and clinical significance of tyrosine phosphatase SHP-2 in colon cancer. *Biomed. Pharmacother.* **68**, 285–290 (2014).
- LaRochelle, J. R. et al. Structural reorganization of SHP2 by oncogenic mutations and implications for oncoprotein resistance to allosteric inhibition. *Nat. Commun.* **9**, 4508 (2018).
- Pádua, R. A. P. et al. Mechanism of activating mutations and allosteric drug inhibition of the phosphatase SHP2. *Nat. Commun.* **9**, 4507 (2018).
- Van Vlimmeren, A. E. et al. The pathogenic T42A mutation in SHP2 rewires the interaction specificity of its N-terminal regulatory domain. *Proc. Natl Acad. Sci. USA* **121**, e2407159121 (2024).
- Keilhack, H., David, F. S., McGregor, M., Cantley, L. C. & Neel, B. G. Diverse biochemical properties of Shp2 mutants. *J. Biol. Chem.* **280**, 30984–30993 (2005).
- Yu, Z.-H. et al. Structural and mechanistic insights into LEOPARD syndrome-associated SHP2 mutations. *J. Biol. Chem.* **288**, 10472–10482 (2013).
- Landrum, M. J. et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–D985 (2014).
- Fowler, D. M., Stephany, J. J. & Fields, S. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat. Protoc.* **9**, 2267–2284 (2014).
- Meier, G. et al. Deep mutational scan of a drug efflux pump reveals its structure–function landscape. *Nat. Chem. Biol.* **19**, 440–450 (2023).
- Newberry, R. W., Leong, J. T., Chow, E. D., Kampmann, M. & DeGrado, W. F. Deep mutational scanning reveals the structural basis for α -synuclein activity. *Nat. Chem. Biol.* **16**, 653–659 (2020).
- Rollins, N. J. et al. Inferring protein 3D structure from deep mutation scans. *Nat. Genet.* **51**, 1170–1176 (2019).
- Persky, N. S. et al. Defining the landscape of ATP-competitive inhibitor resistance residues in protein kinases. *Nat. Struct. Mol. Biol.* **27**, 92–104 (2020).
- Tsuboyama, K. et al. Mega-scale experimental analysis of protein folding stability in biology and design. *Nature* **620**, 434–444 (2023).
- Matreyek, K. A. et al. Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* **50**, 874–882 (2018).
- Grønbaek-Thygesen, M. et al. Deep mutational scanning reveals a correlation between degradation and toxicity of thousands of aspartoacylase variants. *Nat. Commun.* **15**, 4026 (2024).
- Starr, T. N. et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310.e20 (2020).
- Mighell, T. L., Thacker, S., Fombonne, E., Eng, C. & O’Roak, B. J. An integrated deep-mutational-scanning approach provides clinical insights on PTEN genotype-phenotype relationships. *Am. J. Hum. Genet.* **106**, 818–829 (2020).
- Brugge, J. S. et al. Expression of Rous sarcoma virus transforming protein pp60v-src in *Saccharomyces cerevisiae* cells. *Mol. Cell. Biol.* **7**, 2180–2187 (1987).
- Montalibet, J. et al. Residues distant from the active site influence protein-tyrosine phosphatase 1B inhibitor binding. *J. Biol. Chem.* **281**, 5258–5266 (2006).

38. Harris, L. K., Frumm, S. M. & Bishop, A. C. A general assay for monitoring the activities of protein tyrosine phosphatases in living eukaryotic cells. *Anal. Biochem.* **435**, 99–105 (2013).
39. Melnikov, A., Rogov, P., Wang, L., Gnirke, A. & Mikkelsen, T. S. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res.* **42**, e112 (2014).
40. Kontaridis, M. I., Swanson, K. D., David, F. S., Barford, D. & Neel, B. G. PTPN11 (Shp2) mutations in LEOPARD syndrome have dominant negative, not activating, effects. *J. Biol. Chem.* **281**, 6785–6792 (2006).
41. Mohi, M. G. et al. Prognostic, therapeutic, and mechanistic implications of a mouse model of leukemia evoked by Shp2 (PTPN11) mutations. *Cancer Cell* **7**, 179–191 (2005).
42. Martinelli, S. et al. Diverse driving forces underlie the invariant occurrence of the T42A, E139D, I282V and T468M SHP2 amino acid substitutions causing Noonan and LEOPARD syndromes. *Hum. Mol. Genet.* **17**, 2018–2029 (2008).
43. Edouard, T. et al. How do Shp2 mutations that oppositely influence its biochemical activity result in syndromes with overlapping symptoms? *Cell. Mol. Life Sci.* **64**, 1585–1590 (2007).
44. Zhang, R.-Y. et al. Mechanistic insights explain the transforming potential of the T507K substitution in the protein-tyrosine phosphatase SHP2. *J. Biol. Chem.* **295**, 6187–6201 (2020).
45. Yu, Z.-H. et al. Molecular basis of gain-of-function LEOPARD syndrome-associated SHP2 mutations. *Biochemistry* **53**, 4136–4151 (2014).
46. Serbina, A. & Bishop, A. C. Quantitation of autoinhibitory defects in pathogenic SHP2 mutants by differential scanning fluorimetry. *Anal. Biochem.* **680**, 115300 (2023).
47. Kim, S. H., Bulos, M. L., Adams, J. A., Yun, B. K. & Bishop, A. C. Single ion pair is essential for stabilizing SHP2's open conformation. *Biochemistry* **63**, 273–281 (2024).
48. Tao, Y. et al. A novel partially open state of SHP2 points to a 'multiple gear' regulation mechanism. *J. Biol. Chem.* **296**, 100538 (2021).
49. Anselmi, M. & Hub, J. S. Atomistic ensemble of active SHP2 phosphatase. *Commun. Biol.* **6**, 1–11 (2023).
50. Choy, M. S. et al. Conformational rigidity and protein dynamics at distinct timescales regulate PTP1B activity and allostery. *Mol. Cell* **65**, 644–658.e5 (2017).
51. Cui, D. S., Lipchock, J. M., Brookner, D. & Loria, J. P. Uncovering the molecular interactions in the catalytic loop that modulate the conformational dynamics in protein tyrosine phosphatase 1B. *J. Am. Chem. Soc.* **141**, 12634–12647 (2019).
52. Grosskopf, S. et al. Selective inhibitors of the protein tyrosine phosphatase SHP2 block cellular motility and growth of cancer cells in vitro and in vivo. *ChemMedChem* **10**, 815–826 (2015).
53. Keedy, D. A. et al. An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *eLife* **7**, e36307 (2018).
54. Bunda, S. et al. Inhibition of SHP2-mediated dephosphorylation of Ras suppresses oncogenesis. *Nat. Commun.* **6**, 8859 (2015).
55. Yang, J. et al. Crystal structure of human protein-tyrosine phosphatase SHP-1. *J. Biol. Chem.* **278**, 6516–6520 (2003).
56. Chen, Y.-N. P. et al. Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature* **535**, 148–152 (2016).
57. Day, J. E. H. et al. Fragment-based discovery of allosteric inhibitors of SH2 domain-containing protein tyrosine phosphatase-2 (SHP2). *J. Med. Chem.* **67**, 4655–4675 (2024).
58. Fodor, M. et al. Dual allosteric inhibition of SHP2 phosphatase. *ACS Chem. Biol.* **13**, 647–656 (2018).
59. Hidalgo, F. et al. A saturation-mutagenesis analysis of the interplay between stability and activation in Ras. *eLife* **11**, e76595 (2022).
60. Reid, R. J. D. et al. A synthetic dosage lethal genetic interaction between CKS1B and PLK1 is conserved in yeast and human cancer cells. *Genetics* **204**, 807–819 (2016).
61. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).
62. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
63. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10 (2011).
64. Radushev, D. GraphPad Prism. (2022).
65. Wu T, Gale-Day Z. J. & Gestwicki, J. E. DSFworld: A flexible and precise tool to analyze differential scanning fluorimetry data. *Protein Science*. **33**, e5022 (2024).
66. O'Reilly, A. M., Pluskey, S., Shoelson, S. E. & Neel, B. G. Activated mutants of SHP-2 preferentially induce elongation of *Xenopus* animal caps. *Mol. Cell. Biol.* **20**, 299–311 (2000).
67. Fauser, J. et al. Dissecting protein tyrosine phosphatase signaling by engineered chemogenetic control of its activity. *J. Cell Biol.* **221**, e202111066 (2022).
68. Fiordalisi, J. J., Johnson, R. L., Ülkü, A. S., Der, C. J. & Cox, A. D. Mammalian expression vectors for Ras family proteins: generation and use of expression constructs to analyze Ras family function. *Methods Enzymol.* **332**, 3–36 (2001).
69. Schrödinger, L. & Delano, W. PyMOL (2020).
70. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
71. He, R. et al. Exploring the existing drug space for novel pTyr mimetic and SHP2 inhibitors. *ACS Med. Chem. Lett.* **6**, 782–786 (2015).
72. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
73. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
74. Case, D. A. et al. Amber 2022. (2022).
75. Case, D. A. et al. The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
76. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
77. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089 (1993).
78. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
79. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).

Acknowledgements

We would like to thank the members of the Shah Lab for their scientific insights and helpful discussions throughout the development of this project. We thank So Jung Lee from the Rothstein Lab for guidance with yeast work. We thank Jeanine Amacher for constructive feedback on this manuscript. This research was funded by NIH grant R35 GM138014 and American Cancer Society grant RSG-23-1038049-01-TBE to NHS. This work used the Expanse GPU cluster at the San Diego Supercomputer Center through allocation BIO220139 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS)

program, which is supported by NSF grants 2138259, 2138286, 2138307, 2137603, and 2138296.

Author contributions

Z.J. and N.H.S. conceived the project. Z.J., A.E.v.V., and N.H.S. designed the experiments. Z.J., A.E.v.V., and A.S. conducted the experiments. D.K. conducted the molecular dynamics simulations. Z.J., A.E.v.V., D.K., and N.H.S. analyzed and interpreted the data. Z.J. and N.H.S. wrote the manuscript. Z.J., A.E.v.V., D.K., and N.H.S. edited 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/s41467-025-60641-4>.

Correspondence and requests for materials should be addressed to Neel H. Shah.

Peer review information *Nature Communications* thanks Ulrich Stelzl and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025