PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d5cp00015g

Mutation-induced rigidity in the Fyn SH2 domain enhances pY-binding affinity at the cost of peptide specificity†

Li Deng,[a] Yang Zou,[a] Junbao Zhu,[b] Lei Li*[a] and Yanting Wang*[cd]

Received 2nd January 2025,
Accepted 27th May 2025

DOI: 10.1039/d5cp00015g

rsc.li/pccp

Interactions between SH2 domains and tyrosine-phosphorylated (pY) peptides are essential for cellular signaling. While structural studies have revealed how triple-point Fyn SH2 mutants achieve ultra-high pY-peptide affinity, the dynamic consequences of these mutations remain unexplored. In this study, we performed extensive all-atom molecular dynamics simulations on the isolated wild-type Fyn SH2 domain, its mutant, and their complexes with the pY-peptide (EPQpYEEIPIYL). Comparative analyses of these simulations provided dynamic insights into how mutations within the pY-binding pocket alter the interaction between Fyn SH2 domain and the pY-peptide. Our results demonstrate that the mutations significantly influence the dynamic stability of unstructured regions within the SH2 domain and the domain-peptide interface. Specifically, the mutations enhance the rigidity and stability of the pY-binding pocket, as well as the overall structural stability of the domain, including the central β-sheet and terminal regions. This increased rigidity in the mutant enhances interactions between the pY-binding pocket and pY but weakens the interaction with the peptide residue at the +3 position relative to pY, thereby compromising the specificity of the domain-peptide interaction. These findings highlight that the interaction between SH2 domains and pY-peptides is governed not only by the structural properties of the pY-binding pocket but also by the dynamic stability of the domain itself. This insight could guide the experimental design of SH2 domains engineered to recognize post-translational modifications with diverse characteristics.

1 Introduction

Protein phosphorylation is one of the most fundamental post-translational modifications regulating essential cellular processes such as transcription, signal transduction, immune responses, and metabolism.[1–4] Among various phosphorylation events in cell, the recognition of phosphotyrosine (pY)-containing sequences by Src homology 2 (SH2) domains is crucial for initiating specific protein–protein interactions.[5] SH2 domains play important roles in both physiological and pathological processes, including cancer, Noonan syndrome, and immunodeficiencies.[6–8] The SH2 domain consists of approximately 100 amino acids, forming a highly conserved three-dimensional structure, as shown in Fig. 1. The domain comprises a central β-sheet flanked by two α-helices, forming two distinct interaction pockets: the pY-binding pocket and the specificity pocket, with the latter often recognizing the position +1 to +3 residues of the pY site.[5,9,10] Although over one hundred SH2 domain-containing proteins have been identified in the human proteome, the natural binding affinity of SH2 domains for pY-peptides generally remains in the micromolar range,[11,12] which is insufficient for sensitive detection of phosphorylation events. Therefore, developing SH2 domains with both high-affinity and precise-specificity is the key for understanding and modulating phosphorylation-mediated signaling.

To enhance the binding affinity of protein domains, the directed evolution approach (*i.e.*, phage display) has been used to generate so-called "superbinders" with nanomolar affinities for peptides with various post-transcriptional modifications.[13,14] The first discovered SH2 domain superbinder, identified by Kaneko *et al.* through systematic screening of residues within the pY-binding pocket, is a mutant containing three mutations.[15] Beyond engineered superbinder SH2 domains generated through directed evolution, a large and diverse family of naturally

[a] School of Health and Life Sciences, University of Health and Rehabilitation Sciences, 369 Dengyun Road, Qingdao 266071, China. E-mail: lilei@uhrs.edu.cn
[b] College of Life Sciences, Qingdao University, 308 Ningxia Road, Qingdao 266071, China
[c] Institute of Theoretical Physics, Chinese Academy of Sciences, 55 East Zhongguancun Road, P. O. Box 2735, Beijing 100190, China. E-mail: wangyt@itp.ac.cn
[d] School of Physical Sciences, University of Chinese Academy of Sciences, 19 Yuquan Road, Beijing 100049, China

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d5cp00015g

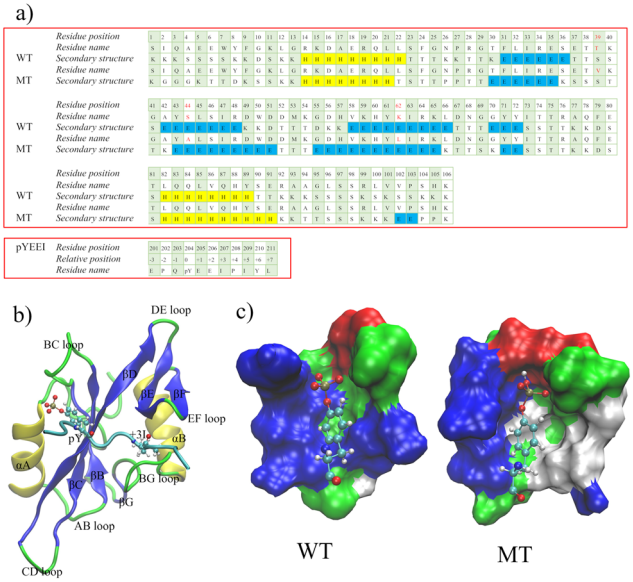

Fig. 1 (a) The sequence and secondary-structure information of wild-type and mutant Fyn SH2 as well as the sequence information of pYEEI. (b) Complex structures of the wild-type Fyn SH2 and pYEEI with different regions labelled. (c) Surfaces of the pY-binding pockets, where positive charge, negative charge, polar and nonpolar surfaces are colored in blue, red, green, and white, respectively.

occurring superbinder SH2 domains without specificity pocket are found within the bacterial kingdom.[16] Although mammalian-derived engineered superbinder SH2 domains often incorporate specific pockets for residues flanking pY, the enhanced affinity exhibited by these superbinder domains for phosphorylated peptides tends to be broadly applicable.[17] Structural analysis of the crystal structure of the complex of superbinder and pY has shown that affinity gains are driven by a newly formed hydrophobic core in the pY-binding pocket by the mutations.[15] However, due to the weak interaction between the specificity pocket and the corresponding residue, the loss of specificity cannot be fully explained by the static structures alone, leaving the mechanistic basis of altered selectivity unresolved.

Complementary to structural experiments, all-atom molecular dynamics (MD) simulations provide valuable time-resolved insights into the flexibility and interactions of biomolecular complexes.[18–20] MD simulations have become a crucial tool for elucidating the allosteric effects of SH2 domains,[21–24] particularly in understanding the activation mechanism of the SHP2 phosphatase.[25,26] These studies have shown that the binding of phosphopeptides to the N-terminal SH2 domain (N-SH2) of SHP2 is not merely a ligand binding event but rather remotely regulates SHP2 activity through intricate allosteric networks. Combined with nuclear magnetic resonance experiment, MD simulations have revealed that pY binding can induce the unzipping of the central β-sheet of N-SH2 domain, consequently affecting the conformation of the active site of SHP2 and thus coupling ligand recognition and enzyme activation.[27,28] There are some studies focused on structural determinants of the pY-binding pocket and the recognition mechanisms of specific phosphorylated motifs such as ITIM/ITSM,[29] but few studies have systematically analysed how mutations in specific residues within the pY-binding pocket directly affect the interaction network within that pocket. Although the structural influence of oncogenic mutations like N642H in the STAT5B SH2 domain has been investigated, the study focuses on the impact of mutations on the overall structure and binding affinity,[30] lacking a detailed analysis of the changes in the interface interactions between two pockets within the SH2 domain. Thus, it remains unclear how mutations of residues within the pY-binding pocket of SH2 superbinder cause the domain to exhibit enhanced affinity but loss of specificity.

In this work, we performed extensive all-atom MD simulations to investigate the dynamic consequences of three point mutations (T39V, S44A, K62L) in the superbinder of Fyn SH2 domain,[15] both in isolated form and in complex with the pY-peptide EPQpYEEIPIYL (pYEEI). Our results show that these mutations, while strengthening the pY interaction, increase the structural rigidity of the entire SH2 domain, and particularly stabilize the central β-sheet and terminal regions. This enhanced rigidity suppresses the dynamic adaptability of the specificity pocket and weakens its interaction with the +3 peptide residue. These findings provide new insights into the recognition of phosphorylation signals by SH2-containing proteins and may inform the rational design of engineered SH2 domains with both high affinity and target-specific selectivity.

2 Methods

2.1 System modelling

The experimental structure of Fyn SH2 and pYEEI complex (PDB Id: 1AOT)[31] was used to represent the initial structures of isolated Fyn SH2 domain and the complex. To generate the initial structure for the mutant complex, the structure of mutated Fyn SH2 predicted by Alphafold2 with ColabFold[32,33] was aligned to the structure of the wild-type complex, and then the coordinates of the mutant after alignment were merged with the coordinates of pYEEI. All the four systems were placed at the center of a dodecahedron box, with the smallest distances between proteins and the box boundary was set to be 1.5 nm to minimize self-interaction across periodic images. The systems were explicitly solvated with water molecules, and sodium or chloride ions were added for electric neutralization. The number of atoms contained in different systems varies, but was approximately around 30 000.

2.2 Molecular dynamics simulation

The GROMACS2023 software package was used to perform all MD simulation.[34] The Fyn SH2 domain and pYEEI were parameterized by using the CHARMM36 force field,[35] and the explicit solvent was represented by the TIP3P water model.[36] The periodic boundary conditions were applied in all simulations to mimic an infinite system and eliminate boundary effects. The particle-mesh Ewald method with a cutoff of 1.2 nm was adopted to evaluate the long-range electrostatic interaction.[37] The cuttoff of the van der Waals interactions was set to be 1.2 nm and the

non-bonded lists method was used to evaluate the short-range interaction. The SHAKE algorithm[38] was applied to all bonds involving hydrogen atoms. Prior to the simulation of sampling configurations, we pre-equilibrated the system by performing energy minimization and solvent relaxation steps. The steepest decent method was adopted to perform energy minimization procedures. In the solvent relaxation procedure, we firstly conducted a 100-ps restrained simulation at a constant volume (about 216 nm^3) and temperature (300 K) and then a 100-ps restrained simulation at a constant pressure (1 bar) and temperature (300 K), where all the non-hydrogen atoms in the systems were restrained with a harmonic potential. The V-rescale method[39] with a coupling time constant of 0.1 ps was adopted for temperature coupling and the Berendsen method[40] with a coupling constant of 1 ps was used for pressure coupling. After the solvent-relaxation procedure, we performed a 600-ns simulation at a constant pressure (1 bar) and temperature (300 K) to sample conformations for each system. The method of temperature coupling was the same as the one used in the solvent-relaxation procedure, but the method of pressure coupling was replaced by the Parrinello–Rahman barostat[41,42] with a relaxation time of 2.0 ps. The time steps in all simulations were set to be 2 fs. We sampled every 100 ps and performed statistical analysis on the conformations from the last 500-ns simulation. To further validate our 500-ns simulation results, we conducted additional three 200 ns simulations for each system, starting from three individual representative conformations selected out from the last 500-ns simulation trajectory. This operation extends the total simulation time for each system to 1.2 μs.

2.3 RMSD and RMSF

The rmsdist module in the GROMACS software package was used to calculate the root-mean-square deviation (RMSD) of the protein backbone conformations in the simulation trajectory to quantify the change of the protein structure. The rmsf module in the GROMACS software package was used to calculate the root-mean-square fluctuation (RMSF) of the backbone Cα atoms to characterize the thermal fluctuation intensity of the backbone residues.

2.4 Solvent accessible surface area, hydrogen bond and salt-bridge

Solvent accessible surface area (SASA), hydrogen bonds and salt-bridges are critical factors in assessing the stability of protein structures. By analysing the differences in their dynamics across various systems, we can gain a deeper understanding of the relationship between protein structures and functions. We utilized the sasa module[43] in the GROMACS software package to calculate the SASA of the isolated and complexes of the wild-type and mutant Fyn SH2. By setting the minimum threshold of the absolute charge of hydrophilic atoms to 0.2 electrons, the SASA was classified into two types: hydrophilic and hydrophobic. Hydrogen bonds and salt-bridges were analysed with the VMD software package.[44] Hydrogen bonds were evaluated using a cutoff of 0.35 nm for the donor–acceptor distance along with an angle cutoff of 30°. Salt-bridges were calculated using the criterion that the distance between any of the oxygen atoms in the acidic residues and the nitrogen atoms in the basic residues is within the cut-off distance of 0.32 nm.

2.5 Free energy landscape

The free energy landscape of a protein provides a clear characterization of the most stable conformation as well as possible pathways and energy barriers for conformational changes. The Gibbs free energy was used in the analysis of MD simulations, which is defined as $\Delta G = -k_B T(\ln P(\vec{x}) - \ln P(\vec{x}))$, where $\vec{x}$ is the coordinate vector and $P(\vec{x})$ is the sampling probability. Due to the high dimensionality of the protein's coordinates, the complete free energy landscape cannot be effectively displayed. Therefore, in our work, we performed principal component (PC) analysis on the coordinate correlation to reduce the dimensionality and chose the first two PCs as the coordinates to construct the free energy landscape. The covar, anaeig, and sham modules implemented in the GROMACS software package were utilized to carry out these analyses.

2.6 Secondary structure

The define secondary structure of proteins (DSSP) algorithm[45] is a general tool for calculating the secondary-structure content of proteins according to the input three-dimensional coordinates. This tool has also been incorporated into the GROMACS software package. The DSSP algorithm was used to analyse both a single conformation and multiple conformations in the simulation trajectories, which classifies the secondary structures of proteins into nine types,[46] as list in Table S1 (ESI†). When analysing trajectories, DSSP can generate a corresponding secondary-structure sequence for each conformation in a trajectory. For a set of secondary-structure sequences, we utilized the Weblogo algorithm[47] to statistically analyse the distribution of secondary structures at each residue and created a graphical representation. Since Weblogo can only recognize single-character labels of either bases or amino acids, we modified the secondary structure labels "B" and "~" to "D" and "K", respectively, as list in Table S1 (ESI†).

To quantify the differences for the same residue in different systems, we defined the difference in secondary-structure distributions at residue i between system N and M as

$$d^i_{N,M} = \sum_{x \in S} \left| p(x)^N_i - p(x)^M_i \right| \quad (1)$$

where S represents a set of nine different secondary-structure states. To analyse the diversity of secondary structures at each residue, we used information entropy to quantify the degree of disorder in the secondary structure of each residue as

$$s_i = \sum_{x \in S} -p(x)_i \times \log\left(p(x)_i\right) \quad \text{if } p(x)_i \neq 0 \quad (2)$$

2.7 Residue–residue contact map

The residue–residue contact map is a crucial representation of the three-dimensional structure of proteins, which can help us better understand the characteristics of the three-dimensional

structural changes under various conditions. We used the mdmat module in the GROMACS software package to calculate the average distances between different residues in the sampled configurations, which form the residue–residue contact map. The difference at the same point i in two systems N and M is quantified by

$$\Delta \mathrm{dist}^i_{N,M} = \begin{cases} \left(\mathrm{dist}^i_N - \mathrm{dist}^i_M\right)/2 & \text{if } \mathrm{dist}^i_N < 1.5 \text{ and } \mathrm{dist}^i_M < 1.5 \\ 0.75 & \text{if } \mathrm{dist}^i_N = 1.5 \text{ and } \mathrm{dist}^i_M < 1.5 \\ -0.75 & \text{if } \mathrm{dist}^i_N < 1.5 \text{ and } \mathrm{dist}^i_M = 1.5 \\ 0 & \text{if } \mathrm{dist}^i_N = 1.5 \text{ and } \mathrm{dist}^i_M = 1.5 \end{cases} \quad (3)$$

2.8 Protein–peptide interface

The interface between a peptide and a protein is crucial for understanding the interaction stability and specificity. Typically, the contact SASAs between a protein and a peptide in a complex can be used to estimate the size and distribution characteristics of their interface. A contact SASA is defined as the difference between the sum of the individual SASAs of the protein and the peptide within the complex and the SASA of the complex as a whole. SASAs were also calculated by the sasa module[44] in the GROMACS software package. We then statistically analysed the contact residues for each residue of the pY-peptide to elucidate the interface characteristics in detail. Furthermore, we calculated the proportion of water in contact with each pY-peptide residue to assess its degree of solvent exposure. In this analysis, the select module in the GROMACS package was used to fetch the contact residues for each residue. A contact residue was defined as the one that is located within 0.6 nm from the target residue. To assess the stability of protein residues at the interface, the occupancy of a contact residue is defined as the proportion of time it appears in the sampled configurations. Based on the residue occupancy, we classified the contacts into different stability levels. It is considered a stable contact when the occupancy is greater than 50%, a moderately stable contact when it is greater than 10% but less than 50%, and an unstable contact when it is less than 10%. The three-dimensional structures were visualized with VMD[44] and PyMOL,[48] and the plots were drawn with QtGrace.[49]

3 Results and discussion

3.1 Little effects of mutations on the Fyn SH2 structure

Based on the structure of the mutated Fyn SH2 predicted by Alphafold2, we can analyse the impact of the mutations on the structures from a structural perspective. As shown in Fig. 1a, the distributions of the secondary structures for both wild-type and mutant are similar, particularly in the structured regions, such as αA and βB. The names of all structural regions, their corresponding sequence intervals, and their positions in the three-dimensional space are presented in Table S2 (ESI†) and Fig. 1b, respectively. By comparing the differences in the secondary structures across various regions in Table S2 (ESI†), it can be seen that the average difference in the structured regions with α helixes and β strands is 0.12, while the average difference in the unstructured regions is 0.48. Aligning the predicted structure of the mutant with the crystal structure of the wild-type reveals an RMSD difference of 0.14 nm, indicating that the structure predicted by Alphafold2 is overall very similar to the wild-type, as shown in Fig. S1 (ESI†).

Based on the structure of the complexes, we can analyse the interface between the domain and each residue of the pY-peptide by counting the number of adjacent atoms of the residue. As shown in Fig. 1c, we present the interfaces of the pY in both the wild-type and the mutant SH2, along with the corresponding polar and non-polar interactions. The interface of the wild-type is composed entirely of polar surfaces, while that of the mutant comprises both polar and non-polar surfaces. The non-polar surface is formed by three hydrophobic mutant residues adjacent to the phenyl ring of the pY, which was observed in the experiment for Src SH2.[15] By comparing the residues at the interface of pY between the wild-type and mutant complexes in Table S3 (ESI†), we have found that the interface of the mutant does not contain the R64 residue present in the wild-type interface, but includes the residues D16, Y43, I63, and I73, which are absent in the wild-type. Similarly, we have analysed the interface of the residue I207 at the +3 position relative to pY, which corresponds to the specificity pocket. The results in Table S4 (ESI†) show that the residues on the interfaces of both the wild-type and the mutant are identical, indicating that the mutations do not alter the structure of the specificity pocket. In the subsequent sections, we will comprehensively analyse the effects of the mutations on protein–peptide interactions from a dynamic perspective.

3.2 Mutations reduce thermal fluctuations in the Fyn SH2 domain

To assess the dynamic properties of the SH2 domain, we calculated its RMSD and RMSF as an initial analysis. Fig. S2 (ESI†) displays five snapshots of the system taken at 100 ns intervals, illustrating the overall structural stability of the domain. However, due to the small molecular weight of the domain, the RMSD of molecules may be more prone to being influenced by terminal fluctuations, as shown in Fig. S3 (ESI†). To mitigate the influence of thermal fluctuations on the assessment of structural stability, we calculated the RMSD for the central region containing 88 residues, from αA to αB, as shown in Fig. 2a and Fig. S4 (ESI†). For all systems, the average

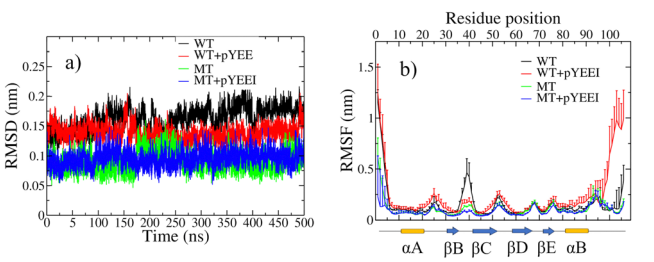

Fig. 2 RMSDs (a) and RMSFs (b) for different systems.

RMSD values are significantly below 0.25 nm, which is the conventional threshold for assessing structural similarity. These results indicate that the central region of the protein exhibits high stability. The average RMSD of the central region of the mutant is 0.09 nm, lower than 0.16 nm observed for the wild-type. In the complex, the average RMSD of the central region of the mutant remains unchanged, while that of the wild-type decreases to 0.14 nm. These results suggest that interactions with the pY-peptide facilitate the stabilization of the central region in the wild-type but has a smaller impact on the mutant. Regardless of its interaction with the pY-peptide, the central region of the wild-type exhibits a larger RMSD compared to the initial structure, indicating that the mutations can also enhance the stability of the central region of the SH2 domain.

The RMSF of each residue's Cα atom was used to assess the thermal-fluctuation intensity of each residue, as shown in Fig. 2b and Fig. S5 (ESI†). Thermal fluctuations in unstructured regions are typically greater than those in structured regions, and both regions are labelled in Table S2 (ESI†). For example, the average RMSF value is 0.09 nm in the structured regions, while the average RMSF value is 0.13 nm in the unstructured regions for the isolated wild-type. For the isolated mutant, the average RMSF values for the structured and unstructured regions are 0.067 nm and 0.138 nm, respectively, both of which are lower than those observed for the wild-type. This indicates that the mutations not only reduce the thermal fluctuations in the structured regions but also in the unstructured regions. Meanwhile, the thermal fluctuations of the residues in the mutant are lower than those in the wild-type, particularly in the terminals, and BC loop regions. This indicates that the mutations contribute to reduced flexibility in the SH2 domain. The various flexibility of proteins results in different responses during their interactions with pY-peptides. In the complex, the average RMSF values for the structured and unstructured regions of the wild-type are 0.146 nm and 0.288 nm, respectively, indicating an overall enhancement in structural dynamics, particularly reflected by increased fluctuations in the C-terminal region. For the mutant, the average RMSF values for the structured and unstructured regions are 0.063 nm and 0.11 nm, respectively, indicating a slight reduction in the overall thermal fluctuations of the domain.

3.3 Mutations enhance internal interactions in the Fyn SH2 domain

The thermal fluctuations of protein residues arise from a balance between intramolecular interactions and interactions with surrounding water molecules, and these fluctuations can be influenced by the solvent-accessible surface area (SASA), which reflects the degree of solvent exposure. By calculating the hydrophilic and hydrophobic SASAs of the isolated wild-type and mutant proteins, as shown in Fig. 3a, the wild-type shows hydrophobic and hydrophilic SASAs of 34.59 nm^2 and 39.85 nm^2, respectively, both of which are larger than the mutant's values of 32.88 nm^2 and 37.26 nm^2. At the same time, the SASA distributions of the mutant are more concentrated.

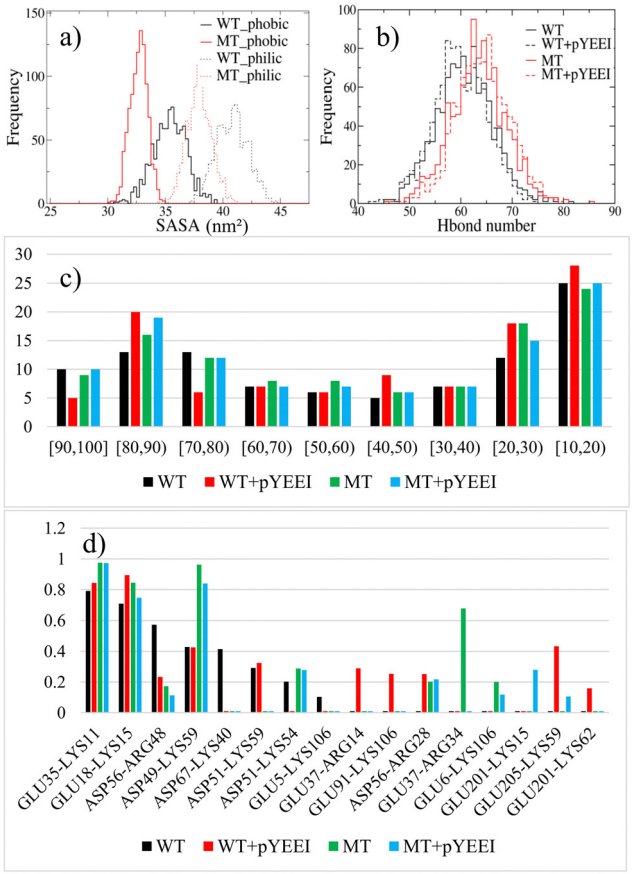

Fig. 3 (a) Distributions of hydrophobic and hydrophilic SASAs for the isolated wild-type and mutant. (b) Frequency distributions of hydrogen bonds. (c) Frequency distributions of hydrogen bonds with different occupancies. (d) Occupancy of salt-bridges.

These results imply that the structure of the mutant is more compact than that of the wild-type. In Fig. S6 (ESI†), we show the hydrophilic and hydrophobic SASA of the SH2 domain in the complex. In the complex, the distribution of the hydrophilic and hydrophobic SASAs of the mutant become more concentrated, while the overall change in the wild-type is not significant. This result suggests that the mutations induce conformational changes that alter the surface properties of the SH2 domain during its interaction with the pY-peptide.

To further illustrate the effects of mutations on protein dynamic fluctuations and interactions with pY-peptides, we analysed the statistics of hydrogen bonds and salt-bridge interactions within the protein in different systems. Hydrogen-bond interactions within the SH2 domain are predominant, with the average total number of hydrogen bonds in different systems being approximately 62, in contrast to merely 5 salt-bridges. In Fig. 3b, we present the distributions of the total number of hydrogen bonds within the SH2 domain in different systems. For the isolated domain, the number of internal hydrogen bonds in the wild-type is 61.12, less than 63.89 observed in the mutant. Compared to the isolated domain, the number of internal hydrogen bonds in the wild-type complex slightly decreases to 60.42, while that in the mutant slightly increases

to 64.76. These results indicate that mutations within the pY-binding pocket strengthen internal hydrogen-bonding networks, which may contribute to enhanced structural stability of the protein.

The occupancy of interactions is also an important metric to assess protein thermal stability, so the distribution of the number of internal hydrogen bonds with different occupancies were also evaluated. The numbers of hydrogen bond types with occupancy greater than 10% were 98, 106, 108, and 108 for the isolated wild-type, the wild-type complex, the isolated mutant, and the mutant complex. Meanwhile, the numbers of hydrogen bonds with occupancy less than 10% were 701, 759, 545 and 477, respectively. This result suggests that the mutation enhances the structural stability of the protein, as evidenced by a reduction in the number of transient hydrogen bonds. Fig. 3c shows that the distribution of the number of hydrogen bonds with occupancy is greater than 10% for different systems. Compared to the wild-type, the number of hydrogen bonds with occupancy in the ranges of [80, 90] and [20, 30] are notably increased respectively by 3 and 6 in the mutant. In the complex, the number of hydrogen bonds with occupancy in the ranges [80, 90] and [20, 30] are increased by 7 and 6, respectively; in contrast, the number of hydrogen bonds with occupancy in the ranges [90, 100] and [70, 80] are notably decreased by 5 and 7, respectively. For the complex of the mutant, only the number of hydrogen bonds with occupancy in the range [80, 90] increases by 3 with respect to the isolated mutant, while the number of hydrogen bonds with occupancy in the range [20, 30] increases by 3. These results demonstrate that the mutations similarly affect internal hydrogen-bond interactions, as binding to the pY-peptide modulates the response of these interactions in the complex state.

Similar to hydrogen-bond interactions, the mutations enhance salt-bridge interactions within the proteins and alter their salt-bridge interactions with pY-peptides. The numbers of salt-bridge types in different systems are 28, 42, 27, and 31 for the isolated wild-type, the wild-type complex, the isolated mutant, and the mutant complex, respectively. In Fig. 3d, we have plotted the occupancy of 16 types of salt-bridges for different systems, with each salt-bridge exhibiting an occupancy exceeding 10% in at least one system. For the isolated domain, the occupancies of 7 salt-bridges are enhanced by the mutations while the occupancies of the other 4 salt-bridges are weakened. In the complex, 8 types of salt-bridges are strengthened, while only 4 types are weakened in the wild-type. By contrast, interactions with the pY-peptide enhances 3 types of salt-bridges while reduces 7 others.

3.4 Mutations and pY-binding modulate conformational dynamics in the Fyn SH2 domain

To understand the dynamics of the wild-type and mutant domains in the isolated or complex state, we have analysed the conformational space occupied by the protein for the first two principal components. We have evaluated the primary contributions of the two principal components (PC1 and PC2) based on their vector compositions, as shown in Fig. S7 (ESI†). In Fig. 4, we present the free energy landscape along with three

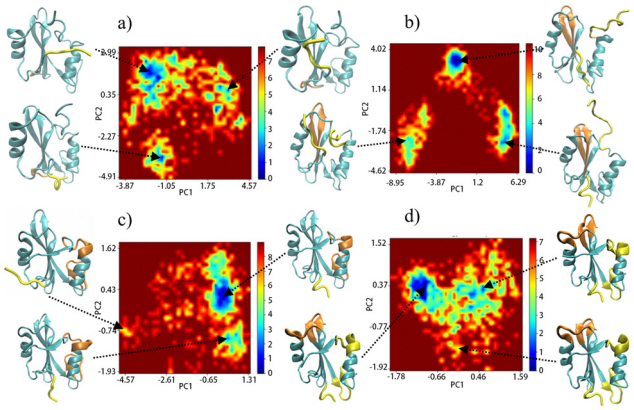

Fig. 4 Free energy landscape of the isolated wild-type Fyn SH2 (a), complex wild-type Fyn SH2 (b), isolated mutant (c), and complex mutant (d). Three conformations corresponding to the local minima of the free energy are presented, where the residues are colored in yellow and orange, respectively, according to their contributions to PC1 and PC2.

structures corresponding to the local minima of the free energy for each system. Additionally, we highlight the residues whose motions primarily contribute to PC1 and PC2 in yellow and orange, respectively, to emphasize the differences between the structures. Overall, whether in the isolated or complex state, the free energy landscape of the wild-type is more dispersed. The coordinate of PC1 for the isolated wild-type ranges from −3.87 to 4.57, while the coordinate of PC2 ranges from −4.91 to 2.99. In the complex, the coordinate of PC1 ranges from −8.95 to 6.29, while the coordinate of PC2 ranges from −4.62 to 4.02. For the isolated mutant, the coordinate of PC1 ranges from −4.57 to 1.31, while the coordinate of PC2 ranges from −1.93 to 1.62. In the complex, the coordinate of PC1 ranges from −1.78 to 1.59, while the coordinate of PC2 ranges from −1.92 to 1.52. It is also evident that the interactions with the pY-peptide increase the range of PC1 for the wild-type while decreasing it for the mutant. However, the interactions of the pY-peptide have little effect on the range of PC2 for either the wild-type or the mutant. These results suggest a reduction in the range of conformational space explored for both the mutations and pY-binding, which indicates a change in the overall dynamics of the domain.

For the isolated wild-type, the primary contribution to PC1 arises from the thermal motion of the unstructured N-terminal region, while PC2 is influenced by the thermal motions of both the unstructured N-terminal region and the CD loop region. In the 55-ns conformation with the lowest free energy shown in the upper left corner of Fig. 4a, the N-terminal is positioned close to the αB helix, but they are separated in the other two conformations. For the isolated mutant, the primary contribution of PC1 arises from the thermal motion of the unstructured N-terminal region, while the main contributions of PC2 stem from the unstructured BG loop region, the EF loop region, and the N-terminal regions. In the 155-ns conformation with the lowest free energy shown in the upper right corner of Fig. 4c, the N-terminal adopts a regular turn and is positioned near the DE loop. Meanwhile, the connections between the αB helix and

the BG loop exhibit less deformation compared to the other two conformations. In the complexes, both the dynamic patterns and structures of the wild-type and mutant domains exhibit significant changes. In the complex, the primary contribution of PC1 in the wild-type arises not only from the N-terminal but also from the C-terminal. Conversely, for PC2, the primary contribution no longer involves the thermal motion of the BC loop region; instead, it pertains to the CD loop region. In the 202.5-ns conformation with the lowest free energy, shown in the upper right corner of Fig. 4b, the N-terminal is near the αB helix, but this proximity occurs at the beginning of the αB region. Meanwhile, the C-terminal forms a bend and is no longer close to the central β-sheet region. In the mutant complex, the primary contribution of PC1 arises from the key regions that contribute to both PC1 and PC2 in the isolated state. Additionally, the primary contribution of PC2 includes the newly observed thermal motion in the AB, CD, and DE loop regions. In the 383.5-ns conformation with the lowest free energy, shown on the left side of Fig. 4d, the N-terminal is close to the DE loop, similar to the conformation in the isolated state. However, the loop between the C-terminal and the αB helix has undergone significant structural changes, leading to the formation of a reverse helical structure. Compared to the mutant, the N- and C-terminal regions of the wild-type protein exhibit greater thermal fluctuations, contributing more substantially to the overall dynamics of the SH2 domain, consistent with the RMSF results. Meanwhile, mutations in the pY-binding pocket and its interaction with pY are associated with a reduction in the thermodynamic motion of the BC loop. By contrast, the intense motion of the C-terminal in the wild-type makes the contributions of the EF and BG loops less prominent to the principal components, which are more apparent in the mutant.

3.5 Mutations strengthen the secondary-structure stability of residues around the pY binding pocket

To analyse the effects of mutations on structural stability across various systems, we first examined the statistical distribution of secondary structures at each residue using DSSP to evaluate the secondary structures of all sampled configurations. To compare the secondary structures of the dynamically stable conformations with those observed in the crystal structures, we identified the most frequently occurring secondary-structure elements at each residue position during the MD simulations, as shown in Fig. S8 (ESI†). We then conducted a statistical analysis of the structural differences between various regions, as presented in Table S5 (ESI†). We observed that, for both the wild-type and mutant, the dynamically stable secondary structures exhibit certain differences from those of the initial structures, with the numbers of differing residues being 36 and 19, respectively. The number of differing residues in the dynamically stable secondary structures between the wild-type and mutant in the isolated state is 18, significantly less than the number of 36 observed in the initial crystal structures. Furthermore, in the complex, there are 6 residues in the mutant domain where the dynamically stable secondary structures change, which is significantly fewer than the number of 25 observed in the wild-type. The results suggest that considering dynamic information in addition to structural information provides a more comprehensive understanding about the effects of mutations on the stability of the domain's secondary structure.

The secondary-structure distribution at each residue is depicted as a Weblogo in Fig. 5a for various systems. Overall, residues in the structured regions retain stable secondary-structure states across all systems, while residues in unstructured regions exhibit significant variations. Larger continuous unstructured regions are generally associated with lower secondary-structure stability. Notably, the N-terminal and the BG loop at the C-terminal are the most unstable ones. Fig. 5a clearly illustrates the secondary-structure states of the three mutations in the pY-binding pocket. For the 39th residue in the BC loop, the dominant secondary structure is hydrogen-bonded turn in the isolated wild-type, but shifts to bend in the complex. In the mutant, the 39th residue remains in the bend state, regardless of the interactions with the pY-peptide. The 44th and 64th residues, located in the βC and βD regions, remain in a stable extended strand state across all systems. These results suggest that the impact of the variant side chain on the secondary structure is influenced by the competition between sidechain and mainchain atomic interactions.

To clearly illustrate the differences in secondary structures across various systems, we quantitatively assessed the disorder and deviation degrees of the secondary structure for each residue, as shown in Fig. 5b and c shows that the disorder degrees in structured regions are low, while those in unstructured regions are high, consistent with the RMSF results (Fig. 2b). In many structured regions, the disorder degree has no significant difference across different systems. However, a significant difference is observed at the 63rd residue between the wild-type and mutant in the βD region. The mutations are associated with an enhanced extended strand conformation at

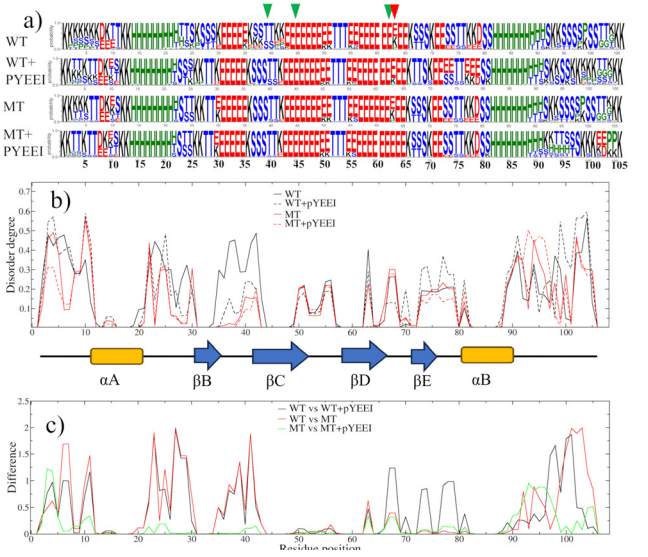

Fig. 5 Weblogos of secondary-structure sequences for all the four systems (a), as well as the degrees of disorder (b) and deviation (c) of the secondary structure for each residue.

this position, which may contribute to increased stability of the central β-sheet. For unstructured regions, the differences in the disorder degree among different systems are quite regular for small unstructured regions, but those are not regular for large unstructured regions, such as the N-terminal, C-terminal, and AB loop. The distributions of the disordered degree in the CD loop are consistent across different systems. This indicates that the mutations and interactions with the pY-peptide have little effect on the dynamics of these regions. In the BC loop region, the disorder degree decreases for both induced mutations and interactions with the pY-peptide, with mutations exerting a stronger enhancing effect than the interactions with the pY-peptide. In the EF loop region, the distribution of the disorder degree is similar for both the isolated wild-type and the mutant. However, after binding with the pY-peptide, the disorder degree decreases for the wild-type while it increases for the mutant. This result indicates that the mutations effectively enhance the stability of the local regions near the pY-binding pocket while simultaneously altering the interactions of other regions with the pY-peptide, particularly in the specificity pocket.

The secondary-structure distribution of the deviation across various systems in Fig. 5c provide further evidence for the observations mentioned above. The distributions of deviation caused by the mutations show similarity to those caused by the interactions with the pY-peptide in the unstructured regions near the pY-binding pocket, such as the BC and CD loops. By contrast, the distribution of deviation caused by mutation is observed to be lower than that resulting from the interactions of the pY-peptide in the unstructured regions near the specificity pocket, such as the DE and EF loops. Additionally, the structural deviations induced by pY-peptide binding in the mutant are smaller than those observed in the corresponding regions of the wild-type protein.

3.6 Mutations enhance interactions between the terminals and the central regions

The information regarding the secondary structure of proteins primarily highlights the spatial characteristics of backbone arrangements but does not capture the interactions between residue side-chain groups. However, the three-dimensional folding of proteins results in residues making contacts not only with nearby residues within the sequence or secondary structure but also with residues that are not in proximity. This is clearly supported by the contact curves of the mutated residues and S104 related to other residues of the protein, as shown in Fig. S9 (ESI†).

To comprehensively display the structural differences of proteins across various systems, we plot the residue–residue contact maps of the SH2 based on the average minimum distance between residues, as shown in Fig. 6. Comparing the contact maps of the wild-type and the mutant in their isolated states reveals that the mutations significantly enhance interactions between residues in the unstructured terminals and those in other regions. The difference between the two maps (as shown in Fig. S10, ESI†) indicates that the mutations are associated with an increase in contact between the C-terminal

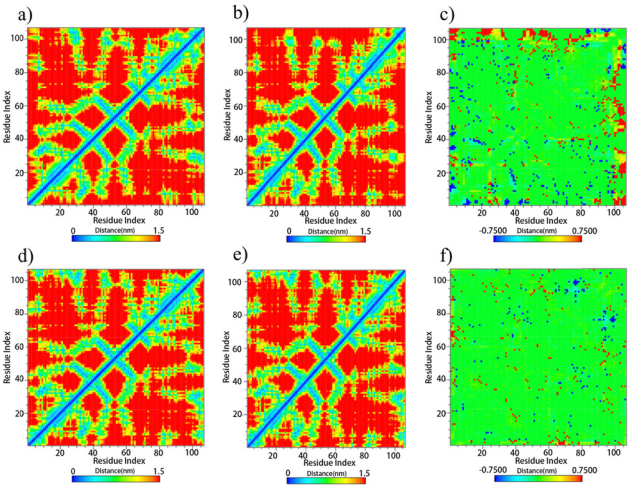

Fig. 6 Residue–residue contact map of the domain for isolated wild-type Fyn SH2 (a), complex wild-type Fyn SH2 (b), isolated mutant (d), and complex mutant (e), as well as difference in contact maps between isolated and complex wild-type Fyn SH2 (c) and that between isolated and complex mutant (f).

and αA, as well as between the N-terminal and βB, βD, βE, DE, and EF loops. In the complex, the residue–residue contacts of the wild-type exhibit more pronounced changes than that of the mutant. For the wild-type, interactions between the C-terminal and multiple regions, including the N-terminal, βB, βC, CD loop, and βD, are significantly reduced, while interactions with residues at αB are enhanced. Meanwhile, the N-terminal residues show significant increases of the interactions in multiple regions, including βB and βC, while interactions between βG and the C-terminal decrease. Compared to the wild-type, only the interaction between the BG loop and βF is enhanced in the mutant, while the interaction between the N-terminal and residues in the DE loop is weakened. The above results indicate that mutations in the pY-binding pocket enhance interactions between the terminal and other regions of the domain while reducing dynamic fluctuations at this terminal. The reduction in dynamics at these positions may affect the conformational changes of the domain during its binding to the protein, potentially resulting in alterations to the interaction map between the domain and the pY-peptide, as shown in Fig. S11 (ESI†). This figure shows that, for the wild-type, the contact between the pY-peptide and the domain is separated from the contact between the N-terminal and the protein; however, for the mutant, the two are connected.

3.7 Mutations strengthen the interface contacts of the pY-binding pocket but reducing those of the specificity pocket

Insights into interface dynamics are essential for comprehending the stability and specificity of interactions between the SH2 and its target pY-peptide. To assess the features of interfaces across different systems, we calculated the hydrogen bonds, salt-bridges, and contact SASA between the domain and the pY-peptide. As shown in Fig. S12 (ESI†), the distribution of hydrogen-bond numbers between the wild-type and the pY-peptide differs only

slightly from that of the mutant and the pY-peptide. Considering hydrogen-bond stability, the number of hydrogen bonds with occupancies greater than 60% is 3 for the wild-type and 5 for the mutant, respectively, as listed in Table S6 (ESI†), indicating a trend towards stronger hydrogen-bond interactions between the mutant and the pY-peptide than those between the wild-type and the pY-peptide. As shown in Fig. 3d, the number of salt-bridges with occupancies exceeding 10% between the mutant domain and the pY-peptide is 2, the same as that between the wild-type domain and the pY-peptide. One of these salt-bridges occurs between G205 and K59, while the other is between G201 and K15 in the mutant, and K62 in the wild-type. In Fig. S13 and S14 (ESI†), we present the distribution of contact SASAs for the wild-type and mutant complexes, as well as the average contact SASAs for each residue. Overall, both the hydrophilic and hydrophobic contact SASAs are very similar between the wild-type and mutant complexes, and the distribution of residues with significantly large contact SASAs is also the same. However, by looking at the residues with contact SASAs exceeding 0.7 nm^2 in both systems, as shown in Fig. S14 (ESI†), we find five such residues in the wild-type complex—two from the protein domain and three from the pY-peptide—whereas in the mutant, all four residues originate from the pY-peptide. This result suggests that, although the mutations do not significantly affect the overall average strength of interactions between the domain and the pY-peptide, they do alter the specific detailed interactions at the interface.

To analyse the dynamic characteristics of the interface, we evaluated the residues of SH2 and water molecules within 0.6 nm from a residue in the pY-peptide. The percentage of water molecules around each residue is illustrated in Fig. 7a. Regardless of whether it is the wild-type or mutant protein, the residues at positions pY204 and I207 exhibit a lower percentage of contact with water molecules, suggesting a tendency of a stronger interaction with the residues in the domain. This result is consistent with the experimental interface of Fyn SH2, which is in agreement with a binding preference for the residue at the +3 position relative to the pY.[50] As shown in Fig. 7b, these two residues exhibit the most stable interactions with the protein, based on the number of contact residues with occupancy values exceeding 90%. There are 13 residues in stable contact with pY204 in the wild-type complex and 14 residues in stable contact with pY204 in the mutant complex, as well as 4 residues in stable contact with I207 in both complexes.

All information regarding the contact protein residues and their occupancies is presented in Fig. S15 (ESI†). Comparing the results of the wild-type to those of the mutant, the number of contact residues for the pY-peptide residues at positions from 201 to 205 remains unchanged with the introduction of mutations, while the contact strength with high occupancy slightly increases. However, a significant change is observed in the number of protein residues contacting the pY-peptide at positions 206 to 211. For example, for Y210, the number of contact protein residues in the wild-type is 34, which is significantly greater than 24 observed in the mutant. This difference may be attributed to the more stable structure of the mutant. Moreover, the contact strength of each pY-peptide residue has changed, and the residues

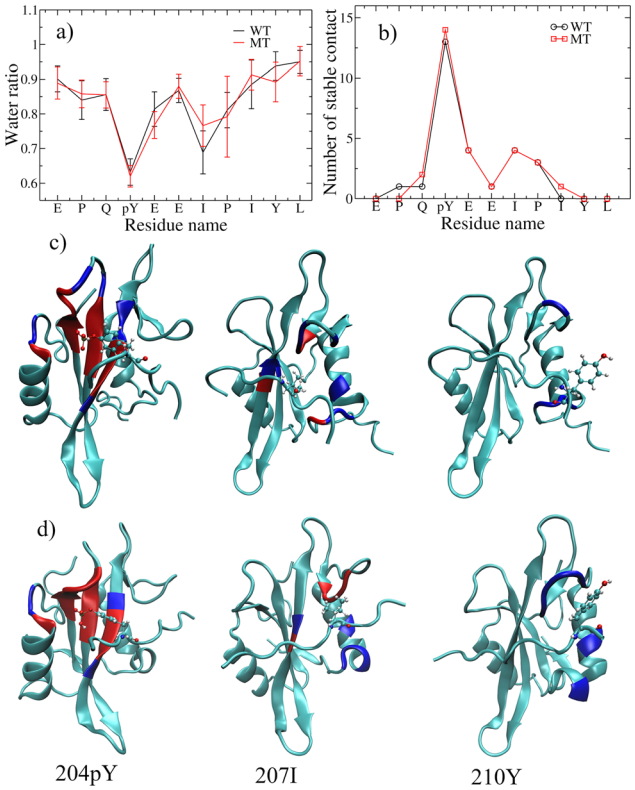

Fig. 7 Percentage of contact water molecules (a) and number of contact protein residues with occupancy exceeding 90% (b) for each residue. The contact residues of 204pY, 207I, 210Y in the three-dimensional structure for wild-type Fyn SH2 (c) and the mutant (d). The contact residues with occupancy greater than 90% are marked in red and those with occupancy greater than 50% and less than 90% are marked in blue.

with high occupancy in the mutant differ from those in the wild-type. For example, for I207, the number of contact protein residues with occupancies exceeding 50% in the wild-type is 17, which is significantly higher than 10 observed in the mutant.

In Fig. 7c and d, we visualize the three-dimensional contact information for pY204, I207, and Y210 in the conformations sampled at 202.5 ns and 383.5 ns, corresponding to free energy minimum, for both the wild-type and the mutant. To differentiate the contacts with occupancies greater than 50%, the contact residues with occupancies exceeding 90% are marked in red, while those with occupancies below 90% are marked in blue. For pY204, the spatial distributions of contact residues with occupancies greater than 50% are remarkably similar between the wild-type and mutant complexes; however, there is a notable difference in the DC loop region. In the mutant complex, the occupancy of contact residues in the DC loop exceeds 90% for all residues, whereas half of the contacts in this region have occupancy values below 90% in the wild-type complex. This finding suggests that the mutations enhance the binding of pY within the pocket region, resulting in a reduction of dynamical fluctuations in this region. For I207, the spatial distribution of contact residues with occupancies greater than 50% in the wild-type complex is more dispersed and forms a deeper specificity pocket compared to the mutant complex.

There are contact residues with occupancies less than 90% on the βC region and contact residues with occupancies greater than 90% on the BG loop. These differences suggest that the interaction between the specificity pocket of the wild-type protein and the residue at the +3 position relative to pY is stronger than that observed in the mutant. For Y210, there are no contact residues with occupancies greater than 90% in either the wild-type or mutant complexes; however, contact residues with occupancies less than 90% in the mutant are more widely distributed, particularly in the αB region. The above results imply that mutations in the pY-binding pocket can enhance its interaction with pY while decreasing the interaction of the specificity pocket with the residue at the +3 position relative to pY. To our knowledge, this is the first report highlighting correlations between two pockets within one SH2 domain.

4 Conclusions

In this study, we have conducted a comprehensive analysis to understand how the three mutations in the pY-binding pocket of the Fyn SH2 domains influence the dynamic stability of the protein domain. We further characterized the dynamic interface between the domains and pYEEI. Our results show that the mutations enhance the interaction between the pocket and pY while reduce the flexibility of SH2. This leads to a weakening of the interaction between the specificity pocket and the residue at the +3 position relative to pY.

Our results indicate that the reduction in protein flexibility can alter the domain's collective motion and free energy landscapes. Similar to pY-binding, mutations in the pY-binding pocket also decrease the domain's flexibility, as reported in the study of STAT5B SH2 domain,[30] correlated with a more compact distribution of the free energy landscape. Studies on the N-SH2 domain of the SHP2 protein have revealed that pY-binding induces changes in the SH2 domain's collective motion, which has been crucial in elucidating the molecular mechanism of SHP2 activation.[25,26] The Fyn SH2 domain displays increased rigidity when both mutations and pY-peptide binding are present, suggesting that the pY-binding pocket and the specificity pocket are dynamically coupled through collective motions.

The detailed analysis of secondary structure distributions reveals that the mutations enhance the stability of the I63 β-strand within the central β-sheet. In the N-SH2 domain of SHP2, binding of pY to the pY-binding pocket and the subsequent unzipping of the central beta-sheet is a key dynamic event in SHP2 activation.[27,28] The increased β-strand stability suggests that the mutations favor a more stable zipped conformation of the central β-sheet, which could potentially lead to a loss of function of proteins containing an SH2 domain. In addition, our three-dimensional residue–residue contact analysis indicates that this stabilization strengthens the interactions between the domain's terminals and the central β-sheet, consequently reducing disorder in the terminal regions.

Furthermore, we study in detail the dynamic characteristics of the interface between the domains and the pY-peptide. Our results indicate that the mutations do not disrupt the interface between the pY-binding pocket and the pY residue; instead, they enhance the overall strength of the interaction. However, they significantly alter the interface between the specificity pocket and the residue at the +3 position relative to pY, leading to the reduction of the contact binding strength. This is, to the best of our knowledge, the first work that directly demonstrates how mutations alter protein flexibility, and thus influences the interaction patterns between different binding pockets within the domain. Enhanced interaction between the pY-binding pocket and pY, and a corresponding decrease in the interaction between the specificity pocket and the hydrophobic residue at the +3 position, provide a reasonable explanation for the experimentally observed increase in affinity and concurrent loss of specificity for superbinder SH2 domain.[15,17] Therefore, this result not only offers valuable guidance for the design and development of high-affinity and precise-specificity domains, but also provides new insights into the complexity of protein–protein interactions.

Author contributions

LD designed this research and performed MD simulations. LL, YZ and YW promoted the research by discussion. LD, LL, and YW wrote the manuscript. LL and YZ received grant funding. All authors reviewed and edited the manuscript. All authors have read and approved the manuscript.

Data availability

The data supporting this article have been included as part of the ESI.†

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (32071430 and 32271504), the Innovation Capability Improvement Project of Science and Technology for Small and Medium-sized Enterprises in Shandong Province (2021TSGC1295), and the Qingdao Natural Science Foundation for Young Scholars Project (23-2-1-35-zyyd-jch). We are thankful for the computer facilities at the University of Health and Rehabilitation Sciences.

References

1 P. A. Levene and C. L. Alsberg, The cleavage products of vitellin, *J. Biol. Chem.*, 1906, **2**, 127–133.
2 T. M. Karve and A. K. Cheema, Small changes huge impact: the role of protein posttranslational modifications in

cellular homeostasis and disease, *J. Amino Acids*, 2011, **2011**, 207691.
3 W. Gong, D. Zhou, Y. Ren, Y. Wang, Z. Zuo, Y. Shen, F. Xiao, Q. Zhu, A. Hong, X. Zhou, X. Gao and T. Li, PepCyber:P∼PEP: a database of human protein–protein interactions mediated by phosphoprotein-binding domains, *Nucleic Acids Res.*, 2007, **36**, D679.
4 A. Forrest, D. Taylor, J. Fink, M. Gongora, C. Flegg, R. Teasdale, H. Suzuki, M. Kanamori, C. Kai, Y. Hayashizaki and S. Grimmond, PhosphoregDB: the tissue and sub-cellular distribution of mammalian protein kinases and phosphatases, *BMC Bioinf.*, 2006, **7**, 82.
5 G. Waksman, D. Kominos, S. C. Robertson, N. Pant, D. Baltimore, R. B. Birge, D. Cowburn, H. Hanafusa, B. J. Mayer, M. Overduin, M. D. Resh, C. B. Rios, L. Silverman and J. Kuriyan, Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides, *Nature*, 1992, **358**, 646–653.
6 Q. Tang, J. Fang, W. Lai, Y. Hu, C. Liu, X. Hu, C. Song, T. Cheng, R. Liu and X. Huang, Hippo pathway monomerizes STAT3 to regulate prostate cancer growth, *Cancer Sci.*, 2022, **113**, 2753–2762.
7 M. Bentires-Alj, J. Paez, F. David, H. Keilhack, B. Halmos, K. Naoki, J. Maris, A. Richardson, A. Bardelli, D. Sugarbaker, W. Richards, J. Du, L. Girard, J. Minna, M. Loh, D. Fisher, V. Velculescu, B. Vogelstein, M. Meyerson, W. Sellers and B. Néel, Activating mutations of the noonan syndrome-associated SHP2/PTPN11 gene in human solid tumors and adult acute myelogenous leukemia, *Cancer Res.*, 2004, **64**, 8816–8820.
8 F. Candotti, S. Oakes, J. Johnston, S. Giliani, R. Schumacher, P. Mella, M. Fiorini, A. Ugazio, R. Badolato, L. Notarangelo, F. Bozzi, P. Macchi, D. Strina, P. Vezzoni, R. Blaese, J. O'Shea and A. Villa, Structural and functional basis for JAK3-deficient severe combined immunodeficiency, *Blood*, 1997, **90**, 3996–4003.
9 G. Waksman, S. Shoelson, N. Pant, D. Cowburn and J. Kuriyan, Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms, *Cell*, 1993, **72**, 779–790.
10 A. Diop, D. Santorelli, F. Malagrinò, C. Nardella, V. Pennacchietti, L. Pagano, L. Marcocci, P. Pietrangeli, S. Gianni and A. Toto, SH2 domains: folding, binding and therapeutical approaches, *Int. J. Mol. Sci.*, 2022, **23**, 15944.
11 R. B. Jones, A. Gordus, J. A. Krall and G. MacBeath, A quantitative protein interaction network for the ErbB receptors using protein microarrays, *Nature*, 2006, **439**, 168–174.
12 J. E. Ladbury and S. T. Arold, Energetics of Src homology domain interactions in receptor tyrosine kinase-mediated signaling, *Methods Enzymol.*, 2011, **488**, 147–183.
13 K. Nowak, F. Rosenthal, T. Karlberg, M. Bütepage, A.-G. Thorsell, B. Dreier, J. Grossmann, J. Sobek, R. Imhof, B. Lüscher, H. Schüler, A. Plückthun, D. M. Leslie Pedrioli and M. O. Hottiger, Engineering Af1521 improves ADP-ribose binding and identification of ADP-ribosylated proteins, *Nat. Commun.*, 2020, **11**, 5199.
14 S. Li, Y. Zou, D. Zhao, Y. Yin, J. Song, N. He, H. Liu, D. Qian, L. Li and H. Huang, Revisiting the phosphotyrosine binding pocket of Fyn SH2 domain led to the identification of novel SH2 superbinders, *Protein Sci.*, 2021, **30**, 558–570.
15 T. Kaneko, H. Huang, X. Cao, X. Li, C. Li, C. Voss, S. S. Sidhu and S. S. C. Li, Superbinder SH2 domains act as antagonists of cell signaling, *Sci. Signaling*, 2012, **5**, ra68.
16 T. Kaneko, P. J. Stogios, X. Ruan, C. Voss, E. Evdokimova, T. Skarina, A. Chung, X. Liu, L. Li, A. Savchenko, A. W. Ensminger and S. S.-C. Li, Identification and characterization of a large family of superbinding bacterial SH2 domains, *Nat. Commun.*, 2018, **9**, 4549.
17 Y. Bian, L. Li, M. Dong, X. Liu, T. Kaneko, K. Cheng, H. Liu, C. Voss, X. Cao, Y. Wang, D. Litchfield, M. Ye, S. S.-C. Li and H. Zou, Ultra-deep tyrosine phosphoproteomics enabled by a phosphotyrosine superbinder, *Nat. Chem. Biol.*, 2016, **12**, 959–966.
18 H. Li, Q. Chen, C. Shan, C. Guo, X. Yang, Y. Chen, J. Zhu and Q. Ouyang, Characterizing the binding sites for GK domain of DLG1 and DLG4 *via* molecular dynamics simulation, *Front. Mol. Biosci.*, 2020, **7**, 1.
19 N. Künzel and V. Helms, How phosphorylation of peptides affects their interaction with 14-3-3η domains, *Proteins*, 2022, **90**, 351–362.
20 V. H. Man, X. He, J. Gao and J. Wang, Phosphorylation of Tau R2 repeat destabilizes its binding to microtubules: a molecular dynamics simulation study, *ACS Chem. Neurosci.*, 2023, **14**, 458–467.
21 W.-S. Liu, R.-R. Wang, W.-Y. Li, M. Rong, C.-L. Liu, Y. Ma and R.-L. Wang, Investigating the reason for loss-of-function of Src homology 2 domain-containing protein tyrosine phosphatase 2 (SHP2) caused by Y279C mutation through molecular dynamics simulation, *J. Biomol. Struct. Dyn.*, 2020, **38**, 2509–2520.
22 J. Solorza, R. Recabarren and J. Alzate-Morales, Molecular insights into the trapping effect of Ca2+ in protein kinase A: a molecular dynamics study, *J. Chem. Inf. Model.*, 2019, **60**, 898–914.
23 H. Fei, W. Li, N. Lu, Q. Liu and Y. Zhang, Molecular dynamic simulation reveals the molecular interactions of epidermal growth factor receptor with musk xylene are involved in the carcinogenicity, *RSC Adv.*, 2023, **13**, 16311–16320.
24 M. Sk, N. Jonniya, R. Roy and P. Kar, Phosphorylation-induced conformational dynamics and inhibition of Janus kinase 1 by suppressors of cytokine signaling 1, *J. Phys. Chem. B*, 2022, **126**, 3224–3239.
25 M. Anselmi and J. S. Hub, An allosteric interaction controls the activation mechanism of SHP2 tyrosine phosphatase, *Sci. Rep.*, 2020, **10**, 1–15.
26 M. Marasco, J. Kirkpatrick, V. Nanna, J. Sikorska and T. Carlomagno, Phosphotyrosine couples peptide binding and SHP2 activation via a dynamic allosteric network, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 2398–2415.
27 M. Anselmi and J. S. Hub, The loops of the N-SH2 binding cleft do not serve as allosteric switch in SHP2 activation, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2025107118.

28 M. Marasco, J. Kirkpatrick, T. Carlomagno, J. S. Hub and M. Anselmi, Phosphopeptide binding to the N-SH2 domain of tyrosine phosphatase SHP2 correlates with the unzipping of its central β-sheet, *Comput. Struct. Biotechnol. J.*, 2024, **23**, 1169–1180.
29 Y. Cheng, W. Ouyang, L. Liu, L. Tang, Z. Zhang, X. Yue, L. Liang, J. Hu and T. Luo, Molecular recognition of ITIM/ITSM domains with SHP2 and their allosteric effect, *Phys. Chem. Chem. Phys.*, 2024, **26**, 9155–9169.
30 L. Haas-Neill, D. Meneksedag-Erol, A. Chaudhry, M. Novoselova, Q. F. Ashraf, E. D. de Araujo, D. J. Wilson and S. Rauscher, The structural influence of the oncogenic driver mutation N642H in the STAT5B SH2 domain, *Protein Sci.*, 2025, **34**, e70022.
31 T. D. Mulhern, G. L. Shaw, C. J. Morton, A. J. Day and I. D. Campbell, The SH2 domain from the tyrosine kinase Fyn in complex with a phosphotyrosyl peptide reveals insights into domain stability and binding specificity, *Structure*, 1997, **5**, 1313–1323.
32 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature*, 2021, **596**, 583–589.
33 M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov and M. Steinegger, ColabFold: making protein folding accessible to all, *Nat. Methods*, 2022, **19**, 679–682.
34 S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics*, 2013, **29**, 845–854.
35 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller and A. D. MacKerell, CHARMM36m: an improved force field for folded and intrinsically disordered proteins, *Nat. Methods*, 2017, **14**, 71–73.
36 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 1983, **79**, 926–935.
37 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
38 T. R. Forester and W. Smith, SHAKE, rattle, and roll: Efficient constraint algorithms for linked rigid bodies, *J. Comput. Chem.*, 1998, **19**, 102–111.
39 G. Bussi, D. Donadio and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.*, 2007, **126**, 014101.
40 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
41 M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
42 S. Nosé and M. L. Klein, Constant pressure molecular dynamics for molecular systems, *Mol. Phys.*, 1983, **50**, 1055–1076.
43 F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander and M. Scharf, The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies, *J. Comput. Chem.*, 1995, **16**, 273–284.
44 W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graphics*, 1996, **14**, 33–38.
45 W. Kabsch and C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers*, 1983, **22**, 2577–2637.
46 S. Gorelov, A. Titov, O. Tolicheva, A. Konevega and A. Shvetsov, DSSP in GROMACS: Tool for Defining Secondary Structures of Proteins in Trajectories, *J. Chem. Inf. Model.*, 2024, **64**, 3593–3598.
47 G. Crooks, G. Hon, J. Chandonia and S. Brenner, WebLogo: a sequence logo generator, *Genome Res.*, 2004, **14**, 1188–1190.
48 W. L. Delano, The PyMOL molecular graphics system version, 2002, **https://pymol.org/**.
49 W. Andreas, QtGrace, 2018, **https://sourceforge.net/projects/qtgrace/**.
50 H. Huang, L. Li, C. Wu, D. Schibli, K. Colwill, S. Ma, C. Li, P. Roy, K. Ho, Z. Songyang, T. Pawson, Y. Gao and S. S.-C. Li, Defining the specificity space of the human SRC homology 2 domain, *Mol. Cell. Proteomics*, 2008, **7**, 768–784.

