Special Issue
"Progress of Theoretical and Computational Biophysics"

Vol. 16, pp. 452–465 (2019)
doi: 10.2142/biophysico.16.0_452

Biophysics and Physicobiology

https://www.jstage.jst.go.jp/browse/biophysico/

Regular Article

Mutational effects of Cys113 on structural dynamics of Pin1

Teikichi Ikura[1], Yasushige Yonezawa[2] and Nobutoshi Ito[1]

[1]*Department of Structural Biology, Medical Research Institute, Tokyo Medical and Dental University (TMDU), Bunkyo-ku, Tokyo 113-8510, Japan*
[2]*High Pressure Protein Research Center, Institute of Advanced Technology, Kindai University, Kinokawa, Wakayama 649-6493, Japan*

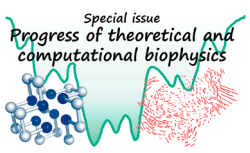
Received July 1, 2019; accepted September 17, 2019

Pin1 is a peptidyl-prolyl isomerase (PPIase) which catalyzes *cis/trans* isomerization of pS/pT-P bond. Its activity is related to various cellular functions including suppression of Alzheimer's disease. A cysteine residue C113 is known to be important for its PPIase activity; a mutation C113A reduced the activity by 130-fold. According to various nuclear magnetic resonance experiments for mutants of C113 and molecular dynamics (MD) simulation of wild-type Pin1, the protonation sate of Sγ of C113 regulates the hydrogen-bonding network of the dual-histidine motif (H59, H157) whose dynamics may affect substrate binding ability. However, it was still unclear why such local dynamic changes altered the PPIase activity of Pin1. In this study, we performed 500 ns of MD simulations of full-length wild-type Pin1 and C113A mutant in order to elucidate why the mutation C113A drastically reduced the PPIase activity of Pin1. The principal component analysis for both MD trajectories clearly elucidated that the mutation C113A suppressed the dynamics of Pin1 because it stabilized a hydrogen-bond between Nε of H59 and Oγ of S115. In the dynamics of wild-type protein, the phosphate binding loop (K63-S71) as well as the interdomain hinge showed the closed-open dynamics which correlated with the change of the hydrogen-bonding network of the dual-histidine motif. In contrast, in the dynamics of C113A mutant, the phosphate binding loop took only the closed conformation together with the interdomain hinge. Such closed-open dynamics must be essential for the PPIase activity of Pin1.

Key words: peptidyl-prolyl isomerase, molecular dynamics simulation, principle component analysis, dual-histidine motif, hydrogen-bonding network

Corresponding authors: Teikichi Ikura, Department of Structural Biology, Medical Research Institute, Tokyo Medical and Dental University (TMDU), 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan. e-mail: ikura.str@tmd.ac.jp; Yasushige Yonezawa, High Pressure Protein Research Center, Institute of Advanced Technology, Kindai University, 930 Nishimitani, Kinokawa, Wakayama 649-6493, Japan. e-mail: yonezawa-wk@waka.kindai.ac.jp

Peptidyl-prolyl isomerase (PPIase) catalyzes *cis/trans* isomerization of a peptide bond preceding a proline residue. Peptidyl-prolyl isomerase NIMA-interacting 1 (Pin1) is a member of the parvulin family of PPIases which specifically catalyze a sequence motif including a phosphoserine or a phosphothreonine followed by a proline (pS/pT-P) [1,2]. Pin1 functions as a molecular switch in various cellular processes including regulation of mitosis [1,3], cell growth [4], maintenance of germinal centers [5], and as a suppressor of neurofibrillary tangles in Alzheimer's disease [6]. Pin1 consists of two domains, the N-terminal WW domain and the



◀ ***Significance*** ▶

Five-hundred ns of MD simulations of full-length wild-type Pin1 and C113A mutant were performed in order to elucidate why a mutation C113A drastically reduced the PPIase activity of Pin1. In the dynamics of wild-type protein, the phosphate binding loop (K63-S71) as well as the interdomain hinge showed the closed-open dynamics which correlated with the change of the hydrogen-bonding network of the dual-histidine motif. In contrast, in the dynamics of C113A mutant, the phosphate binding loop took only the closed conformation together with the interdomain hinge. Such closed-open dynamics must be essential for the PPIase activity of Pin1.

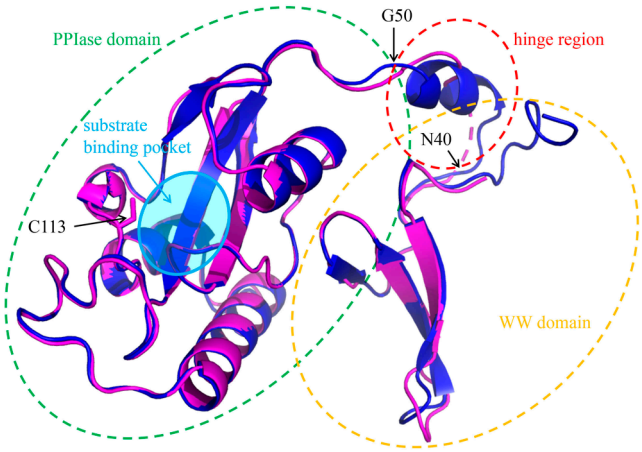


Figure 1 Initial structure for the MD simulation of wild-type Pin1. To obtain the full-length structure of the protein, the hinge region was mutated as G44E/G45N/K46L/N47Y/G48F. The whole residues of this protein (blue) were clearly determined by X-ray crystallography. Its structure was quite similar to the known wild-type structure (PDB ID: 1PIN; magenta) where M1-E5 and N40-G44 were missing: the RMSD between both structures was only 0.71 Å for 153 Cα atoms of the corresponding amino acid residues (K6-G39 and G45-E163). The WW domain (M1-G39), hinge region (N40-Q49), and PPIase domain (G50-E163) are indicated by dashed circles. The substrate binding pocket (L60, L61, M130, Q131, I156 and I159) are shown as a cyan circle.

C-terminal PPIase domain. Both domains bind to the common sequence motif, but their isomer-specificities are different from each other; the WW domain only binds to the *trans* isomer, while the PPIase domain preferentially binds the *cis* isomer, but can bind both the *cis* and *trans* isomers. Many experimental and theoretical analyses have elucidated the catalytic aspects of Pin1. In early studies, the complex structure formed between Pin 1 and a peptide A-P solved by X-ray crystallography suggested a nucleophile attack catalysis by C113 [3]. However, this was later denied by mutational studies on C113; the point mutations C113A, C113D and C113S reduced the PPIase activity of Pin1 by more than 130-fold but they failed to abolish the activity [3,7]. Thereafter, the catalytic site was precisely estimated with a pS-P substrate analogue by nuclear magnetic resonance (NMR) experiments. The catalytic site consists of the substrate binding pocket (L60, L61, M130, Q131, I156, and I159) and the catalytic loop (K63-K80) including the phosphate binding loop (K63-S71) [8] (Fig. 1). However, the importance of C113 was clarified by molecular dynamics (MD) simulations and NMR experiments; the protonation state of Sγ of C113 altered the hydrogen-bonding network of the dual-histidine motif (H59, H157) whose dynamics may affect substrate binding ability [9–11]. The mutations C113A, C113D and C113S also affected the hydrogen-bonding network of the motif [11]. However, it is still unclear why such changes in local dynamics alter the PPIase activity of Pin1 so drastically.

Therefore, in the present study, we performed long MD simulations of full-length wild-type Pin1 and the C113A mutant to elucidate why the substitution of Ala for C113 drastically reduced the PPIase activity of Pin1. The principal component analysis (PCA) of the MD trajectories showed that only wild-type protein had closed-open dynamics of the phosphate binding loop and the interdomain hinge which correlated with the change of the hydrogen-bonding network of the dual-histidine motif. In contrast, in the dynamics of the C113A mutant, the hydrogen-bonding networks of the motifs were fixed in a pattern where the motifs were stabilized in the closed state. Such closed-open dynamics must be essential for the PPIase activity of Pin1.

Materials and Methods

MD simulations

The structure of full-length Pin1 was determined by X-ray crystallography, after introducing five mutations G44E/G45N/K46L/N47Y/G48F in the linker region between the two domains, and expressing and purifying this protein [12]. The full-length C113A mutant was obtained by substituting a hydrogen for the sulfur Sγ of C113. Hydrogen atoms were created at pH 7. Proteins were immersed in a pre-equilibrated TIP3P water rectangular box [13], and overlapping water molecules were discarded. Minimum distances between proteins and box boundaries were set to 18 Å. Adequate numbers of sodium and chloride ions were replaced by water molecules to establish an electrically neutral system presenting an ionic strength of 150 mM. Simulations were performed under periodic boundary conditions using the Amber03 force field [14] for the protein. Each water molecule was treated as a rigid body using the SETTLE algorithm [15]. Protein chemical bonds were considered rigid using the LINCS algorithm [16]. The system was equilibrated using a constant number of atoms–constant volume (NVT) simulation followed by a constant number of atoms–constant pressure (NPT) simulation by applying restraints to the heavy atoms at initial reference positions. The initial velocity was randomly generated using a Maxwell–Boltzmann distribution. NVT simulations were conducted using the weak-coupling thermostat [17] at 310 K, and NPT simulations were performed by the Parrinello–Rahman method [18] at 1 bar and 310 K. Equations of motion were integrated by the leapfrog method. The time step length was set to 2 fs. Electrostatic interactions were calculated by the particle mesh Ewald method [19] with a grid spacing of 1.2 Å and a real space cutoff of 8 Å. The cutoff length was set to 8 Å for van der Waals interactions. After the equilibration was completed, 500 ns of simulations were conducted without any restraints. The last 470 ns of each trajectory was used for analysis. Trajectories were recorded every 10 ps for analysis. All MD simulations were conducted using the Gromacs 4.5.5 package [20].

Analysis of MD trajectories

PCA

PCA is a linear dimensionality reduction technique that can be utilized for extracting information from a high-dimensional space by projecting it into a lower-dimensional sub-space. In the present study, we applied PCA to 47,000 snapshots for each of the wild-type and C113A trajectories. We also applied PCA to 94,000 snapshots, which is the sum of the wild-type and C113A trajectory snapshots, and directly compared both trajectories. We used Mathematica 11.3 (Wolfram Research) first, and then a newly developed program using the Biopython library [21].

Criteria for grouping

The distribution of snapshots on the projection map of PCA was classified into groups using the following procedure:

1. After drawing the density map (the grid size is 0.1 nm×0.1 nm), the grids with low density were excluded from grouping. In this study, grids with less than 4 snapshots were excluded.
2. Some grids with the highest local density were selected on the density map (the grid size was 0.01 nm×0.01 nm). The number of grids equals the number of groups, with each grid representing the group. In this study, four grids were selected from each density map. Therefore, the number of groups was 4.
3. A single snapshot with the lowest RMSD for the average structure of each grid was selected as the representative structure of a group.
4. The boundary between two groups was located at the lowest density grid on the line connecting the two snapshots representing the groups, and also located on the line perpendicular to that line (see Fig. 4c and 6c).

Composite RMSD map

The composite RMSD map is a square map that shows which parts of a sequence of two snapshots taken from an MD trajectory change or maintain the conformation. The map size corresponds to the number of residues in the protein. In this study, Pin1 had 163 residues, so the map size is 163×163. The upper-left triangle of the square map gives the information about the backbone of the protein (Cα atoms), while the lower-right triangle of the map gives information about the whole residue (all heavy atoms). The coordinates on the map correspond to the amino acid residues, that is, the coordinates (i, j) mean the continuous part between the i^{th} residue and the j^{th} residue. The color of the coordinates (i, j) corresponds to the root-mean-square-deviation (RMSD) value for the continuous part between the i^{th} residue and the j^{th} residue. The lower the RMSD, the bluer the color, and the higher the RMSD, the more red. In this study, the RMSDs were calculated for 13366 ($=_{163}C_2+_{163}C_1=163\times162/2+163$) different parts of the sequence.

Judgements of the closed and open conformations

We judged whether the interdomain hinge and the phosphate binding loop of a snapshot exhibited the closed or open conformation by the following criteria. The distance between Cα atoms of S18 and G99 (Supplementary Fig. S1a) was measured to determine whether the interdomain hinge conformation was closed or open. When calculating the distance of the entire trajectory, the distribution was clearly divided into two. The average value for long distances was 39.4 Å, but the average value for short distances was 29.7 Å. Therefore, when the distance was less than 36.0 Å, it was judged as closed conformation, and when it was not, it was judged as an open conformation. On the other hand, the distance between Cα atoms of R69 and C113 (Supplementary Fig. S1b) was used to determine whether the phosphate binding loop's conformation was closed or open. When calculating the distance of the entire trajectory, the distribution was clearly divided into two. The average value for long distances was 15.2 Å, but the average value for short distances was 11.1 Å. Therefore, when the distance was less than 14.2 Å, it was judged as closed conformation, and when it was not, it was judged as open conformation. To exclude a simple "fluctuation", the averaged distance was applied to the actual calculations. An example of the actual judgement procedure was shown in Supplementary Figure S2.

Correlation coefficients between the flip-flop of H59 and the closed-open dynamics

We expressed the states of flip-flop of H59 and the states of the closed-open dynamics of the interdomain hinge and phosphate binding loop using three variables, h, d, p, respectively. If a hydrogen-bond between Nε of H59 and Oγ of S115 was observed in the i^{th} snapshot, $h(i)$ was set as −1, otherwise as 1. If the interdomain hinge of the i^{th} snapshot exhibited a closed conformation, $d(i)$ was set as −1, otherwise as 1. If the phosphate binding loop of the i^{th} snapshot exhibited a closed conformation, $p(i)$ was set as −1, otherwise as 1. Then, we calculated the correlation coefficient between the flip-flop of H59 and the closed-open dynamics of the interdomain hinge ($CC(h, d)$), the correlation coefficient between the flip-flop of H59 and the closed-open dynamics of the phosphate binding loop ($CC(h, p)$), and the correlation coefficient between the closed-open dynamics of the interdomain hinge and the phosphate binding loop ($CC(d, p)$).

$$CC(h, d) = \frac{1}{n}\sum_{i=1}^{n} h(i)\times d(i),$$

$$CC(h, p) = \frac{1}{n}\sum_{i=1}^{n} h(i)\times p(i),$$

$$CC(d, p) = \frac{1}{n}\sum_{i=1}^{n} d(i)\times p(i),$$

where n is the number of snapshots.

Molecular graphics and graph plots

The molecular structure was drawn using the open source PyMOL molecular graphics system. The graphs were plotted using KaleidaGraph 4.5.2 (Synergy Software).

Results and Discussions

Initial structures

Although various MD simulations have previously been performed for Pin1, these studied only for the PPIase domain or were not applied for mutants of C113 [9,22,23]. There is no information to understand how the mutation C113A affects the overall molecular dynamics of Pin1. Therefore, we decided to use two structures for MD simulation: full-length Pin1 and its C113A mutant. However, when searching for the full-length Pin1 structure in the Protein Data Bank (PDB), no intact structure was found. Forty-one structures of full-length Pin1 have been registered in PDB, but several amino acid residues in the N-terminal region and the hinge region between the WW and PPIase domains, typically M1-E5 and N40-S44 (PDB ID: 1PIN) and sometimes M1-K6 and G39-G50 (PDB ID: 2Q5A), are missing for all the structures. In order to avoid an artifact, a full-length protein is required. To obtain the full-length structure of the protein, we stabilized the linker region by mutations of G44E/G45N/K46L/N47Y/G48F and then determined its crystal structure at 0.99 Å resolutions (Fig. 1). In the final model, the coordinates of all heavy atoms for the whole residues of Pin1 were determined, showing that the structure was quite similar to the known wild-type structure (PDB ID: 1PIN): the RMSD between both structures was only 0.71 Å for 153 Cα atoms of the corresponding amino acid residues (6^{th}-39^{th} and 45^{th}-163^{rd}) (Fig. 1). In additional experiments, the functions of this mutant were shown to be equivalent to those of wild-type Pin1 [12]. Therefore, we dealt with this mutant protein as the wild-type protein in this study, and then obtained the C113A mutant by substituting a hydrogen for the sulfur Sγ of C113.

Dynamical aspects of each protein

Molecular dynamics simulations for wild-type Pin1 and the C113A mutant were performed at 310 K in 150 mM NaCl for 500 ns. The last 470 ns of the trajectories ware used for analysis. Thus, 47,000 snapshots were collected for each trajectory, as the snapshots were sampled every 10 ps. At 30,010 ps (the first of the collected structures), the RMSDs from the crystal structure for the wild-type and C113A proteins were 1.63 and 1.78 Å for all Cα atoms, respectively, while the RMSD between the wild-type and C113A structures was 2.02 Å. These values suggested that the two proteins changed their structures to some extent from the original structure. Figure 2 shows the RMSD of each snapshot from the first structure; the averaged RMSDs for all Cα atoms were 3.86 and 2.93 Å for the wild-type and C113A trajectories, respectively. These RMSDs suggest that the wild-type protein fluctuated more than the C113A mutant.

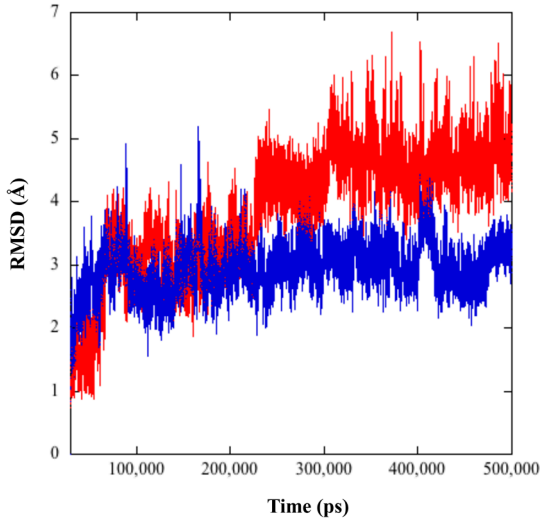


Figure 2 RMSD for all Cα atoms of each snapshot from the first structure. The averaged RMSDs for all Cα atoms were 3.86 and 2.93 Å for the wild-type (red) and C113A (blue) trajectories, respectively. These RMSDs suggest that wild-type protein was more fluctuating than the C113A mutant.

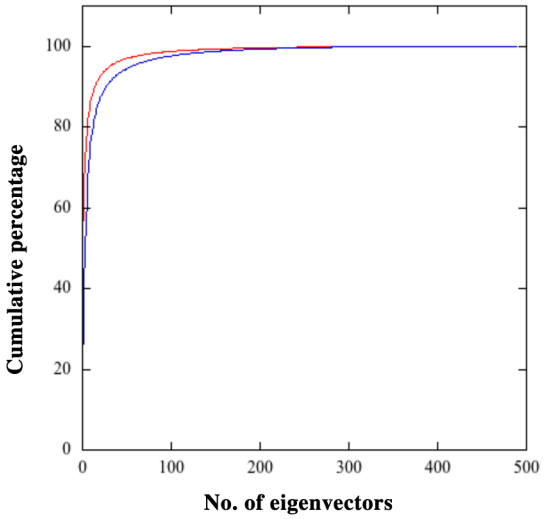


Figure 3 Cumulative percentage of the PCA modes for the wild-type (red) and C113A (blue). The first eigenvector extremely characterizes the dynamics of wild-type Pin1. On the other hand, the mutant C113A requires more eigenvectors than wild-type Pin1 to characterize its dynamics.

To characterize the dynamics of the two proteins over the simulation time, we applied PCA to the trajectories. In the PCA for the trajectory of wild-type protein, the first, second and third modes contributed to the dynamics by 56.7, 8.9 and 5.8%, while each of the eleventh and subsequent modes only contributed only by less than 1.0% (Fig. 3). The contribution of the first mode is 6.4 and 9.8 times larger than those of the second and third modes, respectively, suggesting that

the first eigenvector extremely characterizes the dynamics of the wild-type protein. Figure 4a shows the projection of the wild-type trajectory onto the first 2 eigenvectors. Porcupine plots of the first and second eigenvectors of PCA were shown in Supplementary Figure S3a and b. We also calculated the density at each grid of the projection map to clearly recognize the distribution of snapshots (Fig. 4b). In Figure 4b, we found that the distribution of wild-type's snapshots was not uniform on this plane, but instead formed at least 4 locally densest regions suggesting local minima in the conformational free energy. Therefore, we selected one structure from each of the 4 densest grid as the representative structure of the local minimum; the snapshots of 49,380 ps, 148,880 ps, 255,100 ps, and 441,670 ps were selected for the first, second, third, and fourth minimum, respectively (Fig. 4c & e). Then, we divided the distribution of snapshots into 4 groups as shown in Figure 4c. Figure 4d shows the transitions between the groups during 470 ns. The protein frequently jumped from one group to the other. The protein jumped mainly between the first and second groups up to 200 ns, and then jumped primarily between the third and fourth groups. This indicates that the free energy barrier between the first and second groups or between the third and fourth groups is very low. On the other hand, the free energy barrier between the first (and/or second) and the third (and/or fourth) groups is relatively large. In further analysis, four representative structures were compared with each other. First of all, the composite RMSD maps were applied to determine what parts of the sequence change or maintain the conformation when the protein jumped between the groups. Figure 5 clearly shows that the first and second groups had three common main-chain structures corresponding to segments M1-S42, S43-G50 and E51-E163. The orientations of the three segments differed between the groups although the differences were less than 2.00 Å of RMSDs for Cα atoms. Furthermore, the side-chain structure of R121 had a significantly different orientation. The first and third groups had four common main-chain structures corresponding to segments A2-S42, S43-G50, E51-Q66, and S71-E163. The orientations of the four segments differed between the groups, and the largest difference was observed between the regions of E51-Q66 and S71-E163; the averaged RMSD was more than 3.80 Å for Cα atoms. The structural similarity between the first and fourth groups was almost identical to that between the first and third groups. The second and third groups had three common main-chain structures corresponding to segments A2-G50, E51-Q66, and P70-E163. The orientations of the three segments differed between the groups, and the largest difference was observed between the regions of E51-Q66 and P70-E163; the averaged RMSD was more than 4.00 Å for Cα atoms. The structural similarity between the second and fourth groups was almost identical to that between the second and third groups. The structural difference between the third and fourth groups was very small, as expected from the above analyses. According to a more detailed analysis, the two groups had three common main-chain structures corresponding to segments A2-S67, R68-F125, and S126-E163. The orientations of the three segments differed between the groups although the differences were less than 2.00 Å of RMSDs for Cα atoms. Such small differences between the third and fourth groups presumably allowed the protein to move back and forth frequently between these groups.

In the PCA for the mutant C113A, the first, second and third modes contributed to the dynamics by 26.2, 14.2 and 9.2%, while each of the fifteenth and subsequent modes contributed by less than 1.0% (Fig. 3). The contribution of the first mode was 1.8 and 2.8 times larger than those of the second and third modes, respectively, suggesting that the mutant C113A requires more eigenvectors to characterize its dynamics than wild-type Pin1. Figure 6a shows the projection of the C113A trajectory onto the first 2 eigenvectors. Porcupine plots of the first and second eigenvectors of PCA were shown in Supplementary Figure S3c and d. It was found that the distribution of the mutant snapshots was more isotropic on this plane than the distribution of wild-type protein as expected from its eigenvalues. We also calculated the density at each grid of the projection map to clearly recognize the distribution of snapshots (Fig. 6b), suggesting that this distribution formed 4 locally densest regions. Therefore, we selected one structure from each of the 4 densest grid as the representative structure of the local minimum; the snapshots of 47,670 ps, 449,530 ps, 351,880 ps, and 343,660 ps were selected for the first, second, third, and fourth minimum, respectively (Fig. 6c & e). Then, we divided the distribution of snapshots into 4 groups as shown in Figure 6c. Figure 6d shows the transitions between the groups during 470 ns. The protein stayed in the first group up to 58,340 ps, and then jumped to the second group. Subsequently, it frequently jumped from one group to the other. However, after 58,340 ps, it returned to the first group only 114 times. This indicates that the free energy barrier between the second, third and fourth groups is much lower than the free energy barrier between the three groups and the first group. In further analysis, we compared four representative structures with each other by calculating the above-mentioned composite RMSD maps (Fig. 7). The first and second groups had two common main-chain structures corresponding to segments M1-S41 and G50-e163. The two segments, M1-S41 and G50-E163, correspond to the WW domain and PPIase domain, respectively, suggesting that each functional unit had a stable structure. The orientations of the two domains differed between the groups although the differences were less than 2.50 Å of RMSD for Cα atoms. The first and third groups had four common main-chain structures corresponding to segments A2-S42, S43-G50, E51-Q66, and S71-E163. The orientations of the four segments differed between the groups, and the largest difference was observed between the regions of E51-Q66 and S71-E163; the averaged RMSD was more than 3.80 Å for Cα atoms. The structural similarity

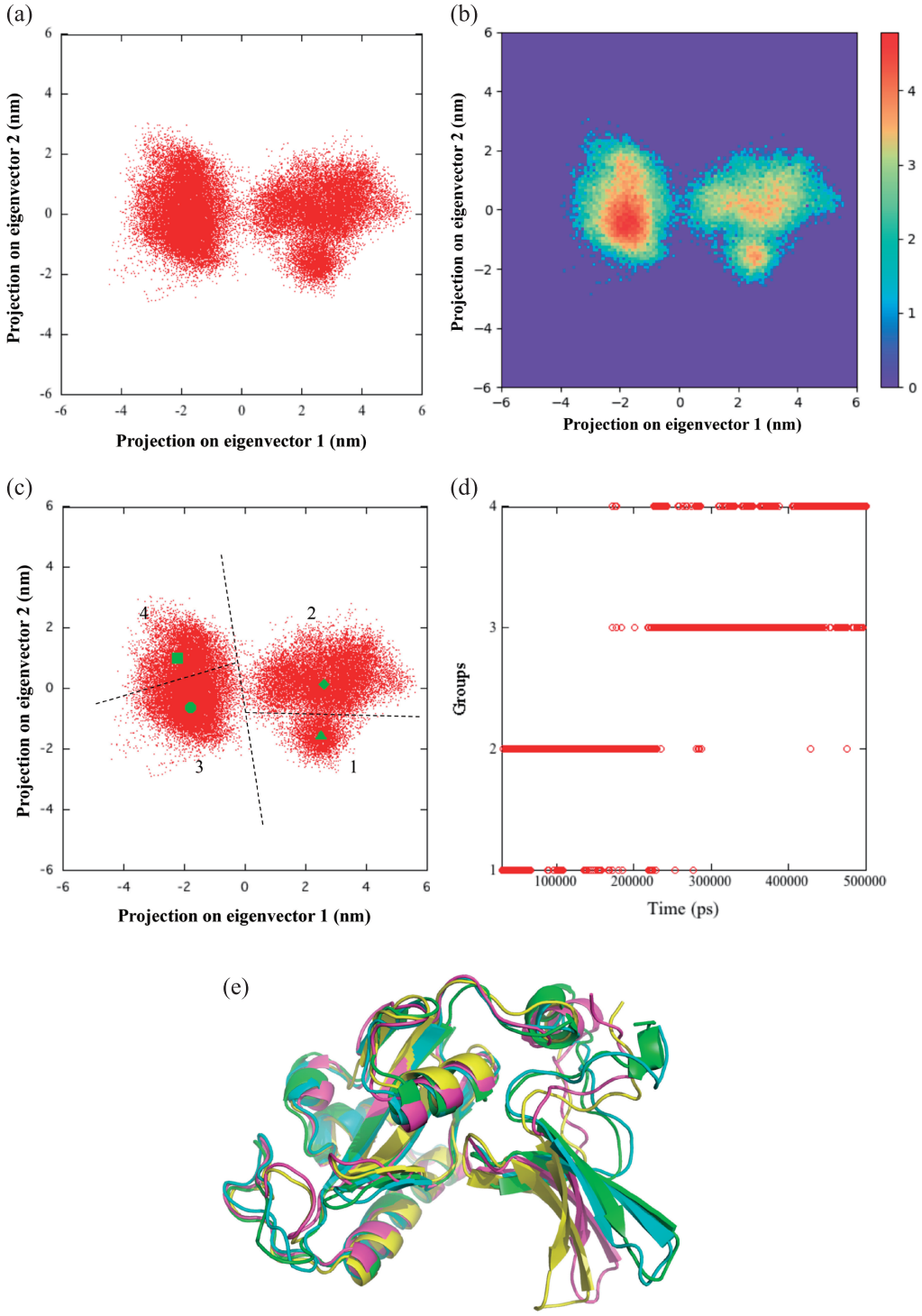


Figure 4 PCA of the wild-type trajectory. (a) Projection of the wild-type trajectory onto the first 2 eigenvectors. (b) The density map in which logarithmic number of the snapshots at each grid (0.1 nm×0.1 nm) in the projection map was represented by color; the smaller the number, the bluer the color. (c) The distribution of snapshots which was divided into 4 groups. The representative structure was chosen for each group; the snapshots of 49,380 ps, 148,880 ps, 255,100 ps, and 441,670 ps were chosen for the first (solid triangle), second (solid diamond), third (solid circle), and fourth (solid square) groups, respectively. (d) Transition between 4 groups during the simulation. (e) Superposition of the representative structures of the groups. Magenta, yellow, green and cyan stand for the first, second, third and fourth group, respectively.

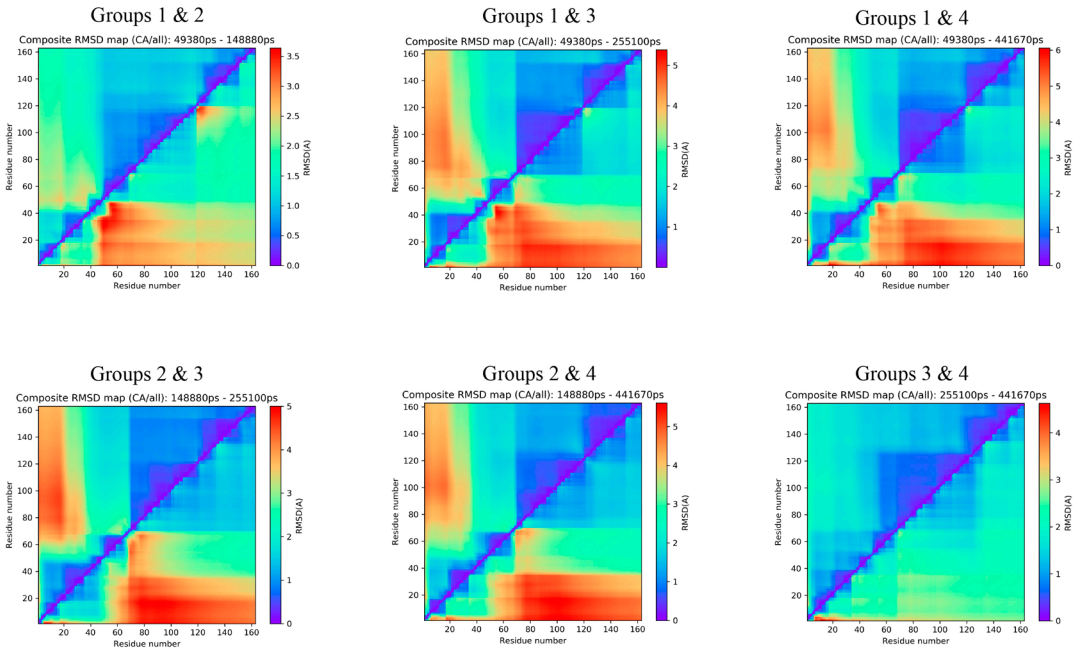


Figure 5 Composite RMSD maps of the wild-type groups, which show RMSDs for Cα atoms (upper left triangle) and all heavy atoms (lower right triangle) for 13,366 segments with all combinations of starting and ending points. Blue squares or triangles represent regions with structural segments common to each of the two groups represented by the snapshot. The third and fourth groups are similar, but very different from both of the first and second groups, while the first and second groups are slightly different from each other.

between the first and fourth groups was almost identical to that between the first and third groups. On the other hand, the second, third and fourth groups were very similar. The averaged RMSDs for all Cα atoms were 1.46 Å between the three groups. Frequent jumps of the protein between the groups are presumed to be because of this structural similarity.

Dynamic differences between wild-type and C113A proteins

To compare the structures of the wild-type and C113A trajectories, we calculated the composite RMSD maps between the representative structures of each trajectory (Fig. 8). Each of the four groups of the wild-type trajectory showed maps with similar patterns for any of the four groups of the C113A trajectory. The structure of the first wild-type group was similar to the structures of the four C113A groups in the regions M1-S41 and G50-E163, suggesting that each of WW and PPIase domains has a common structure between both proteins although the orientations of the domains were different. The similarities between the structures of the second wild-type group and the four C113A groups were almost identical to the similarities between the first wild-type group and the four C113A groups. The structure of the second wild-type group was especially similar to the structure of the third C113A group.

In contrast, although each of the third and fourth wild-type groups very much differed from any of the four C113A groups, three common segments, A2-S42, E51-Q66, and S71-E163 were detected between the groups. The region between the segments A2-S42 and E51-Q66 is located at the hinge between the WW and PPIase domains, while the region between the segments E51-Q66 and S71-E163 overlaps with the phosphate binding loop (K63-S71). The most different position was located in the hinge region between the WW and PPIase domains. The difference in the phosphate binding loop was relatively small but not negligible. Even including the mutational residue C113A, the region S71-E163 had a common structure.

In summary, the composite RMSD maps between wild-type and C113A groups suggest that, (1) structural differences between the C113A groups were less than 2.00 Å of RMSD, (2) the first and second groups of wild-type were relatively similar to the C113A groups (less than 2.00 Å of RMSD), and (3) the third and fourth groups of wild-type were very different from the C113A groups (more than 4.00 Å of RMSD).

The scale of the distribution of snapshots of the wild-type and C113A trajectories was very different. Therefore, we directly compared the wild-type and C113A trajectories by applying PCA to 94,000 snapshots totaled from both the wild-type and C113A proteins. Figure 9a shows the projection of the total snapshots onto the first 2 eigenvectors. Porcupine plots of the first and second eigenvectors of PCA were shown in Supplementary Figure S3e and f. Although the distributions of each of the wild-type and C113A snapshots became less scattered than each distribution shown in Figure 4 and 6, it clearly showed that the distribution of

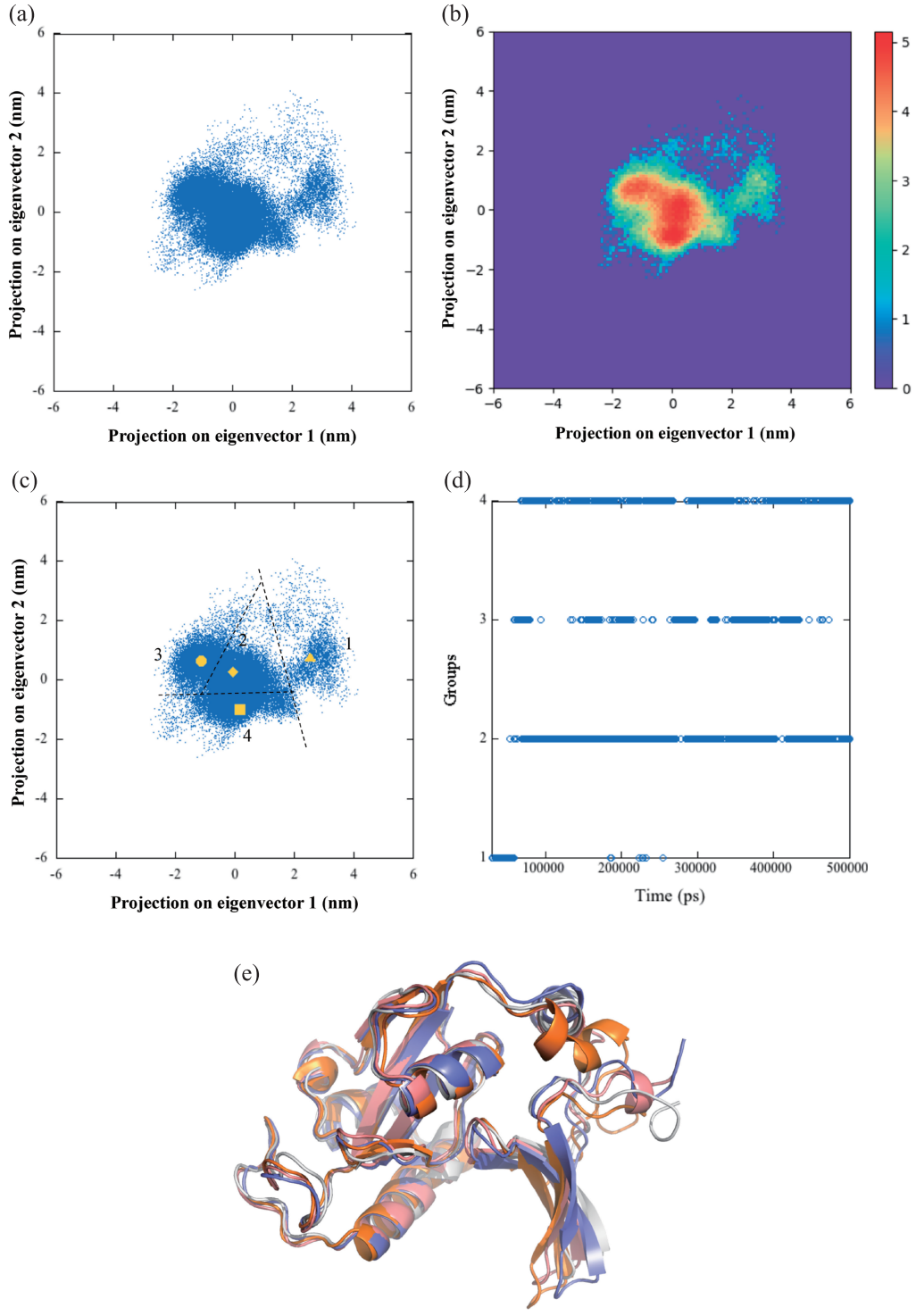


Figure 6 PCA of the C113A trajectory. (a) Projection of the C113A trajectory onto the first 2 eigenvectors. (b) The density map in which logarithmic number of the snapshots at each grid (0.1 nm×0.1 nm) in the projection map was represented by color; the smaller the number, the bluer the color. (c) The distribution of snapshots which was divided into 4 groups. The representative structure was chosen for each group; the snapshots of 47,670 ps, 449,530 ps, 351,880 ps, and 343,660 ps were chosen for the first (solid triangle), second (solid diamond), third (solid circle), and fourth (solid square) groups, respectively. (d) Transition between 4 groups during the simulation. (e) Superposition of the representative structures of the groups. Orange, white, slate blue and wheat stand for the first, second, third and fourth group, respectively.

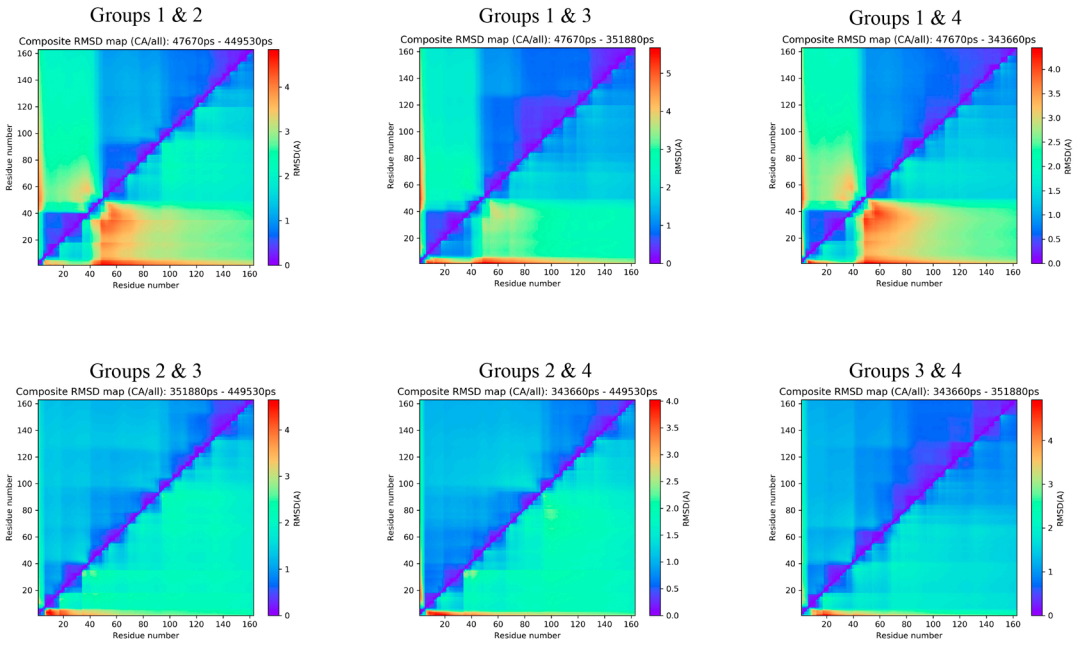


Figure 7 Composite RMSD maps of the C113A groups, which calculated RMSDs of Cα atoms (upper left triangle) and all heavy atoms (lower right triangle) for 13,366 segments with all combinations of starting and ending points. Blue squares or triangles represent regions with structural segments common to each of the two groups represented by the snapshot. The second, third and fourth groups are very similar, but slightly different from the first group.

C113A snapshots overlaps with one of the two distributions of wild-type snapshots. When placing the representative snapshots of the groups determined in Figure 4c and 6c on this projection map, we confirmed that all the groups of C113A overlapped with the first and second groups of the wild-type. This indicates that the mutation C113A restricted the dynamics of the protein to the space corresponding to the first and second groups of the wild-type. Furthermore, in order to investigate the orientations between the common segments A2-S42, E51-Q66, and S71-E163, we applied PCA to the sequences M1-Q66 and E51-E163 (Fig. 9b and 9c). As well as Figure 9a, the first and second groups of the wild-type made one group together with all the four groups of C113A, while the third and fourth groups of the wild-type made another group. This suggests that the dynamics of the interdomain hinge and phosphate binding loop are strongly correlated.

Figure 10a show a superposition of the eight segments corresponding to the sequences M1-Q66 of the representative structures of wild-type and C113A groups. The β-strand 4 (R56-K63) exhibited two orientations for the WW domain. Here, we call the conformation corresponding to the third and fourth groups of wild-type the open conformation, and another conformation as the closed conformation. Each of the conformations was stabilized by a specific hydrogen-bonding network. In the closed conformation, hydrogen bonding between T29(O)-G148(N), T152(Oγ)-H157(Nδ), H157(Nε)-H59(Nδ) and H59(Nε)-S115(Oγ) stabilized the orientation of WW domain (Fig. 10b). The two residues G148 and T152 are both located on β-strand 6 connected via hydrogen bonds to β-strand 4 including H59. Therefore, the WW domain is tightly bound to the PPIase domain in the closed conformation. On the other hand, for the open conformation, hydrogen bonds of I28(O)-G148(N), H157(Nδ)-H59(Nε) and H59(Nδ)-C113(Sγ) stabilized the orientation of WW domain (Fig. 10c). In this conformation, however, T152 did not form a hydrogen bond with H157 and was slightly away from H157 because of steric hindrance (Fig. 10c). Therefore, the WW domain is positioned slightly away from the PPIase domain in the open conformation.

Figure 11a show a superposition of the eight segments corresponding to the sequences E51-E163 of the representative structures of wild-type and C113A groups. The phosphate binding loop (K63-S71) exhibited two conformations. Furthermore, a region of β-strand 6, turn and β-strand 7 (P149-H157) also revealed two conformations positively correlated with the phosphate binding loop. Here we found the closed and open conformation again. As similar to the region M1-Q66, the conformation corresponding to the third and fourth groups of wild-type was open, and another conformation was closed. We detected the identical hydrogen-bonding networks that caused a transition between the closed and open conformations. For the closed conformation, hydrogen bonds between G155(N)-T152(O), G155(O)-T152(N), T152(Oγ)-H157(Nδ), H157(Nε)-H59(Nδ) and H59(Nε)-S115(Oγ) stabilized the conformation of the sequence P149-H157 (Fig. 11b). The hydrogen bond T152(Oγ)-H157(Nδ) was especially important in keeping

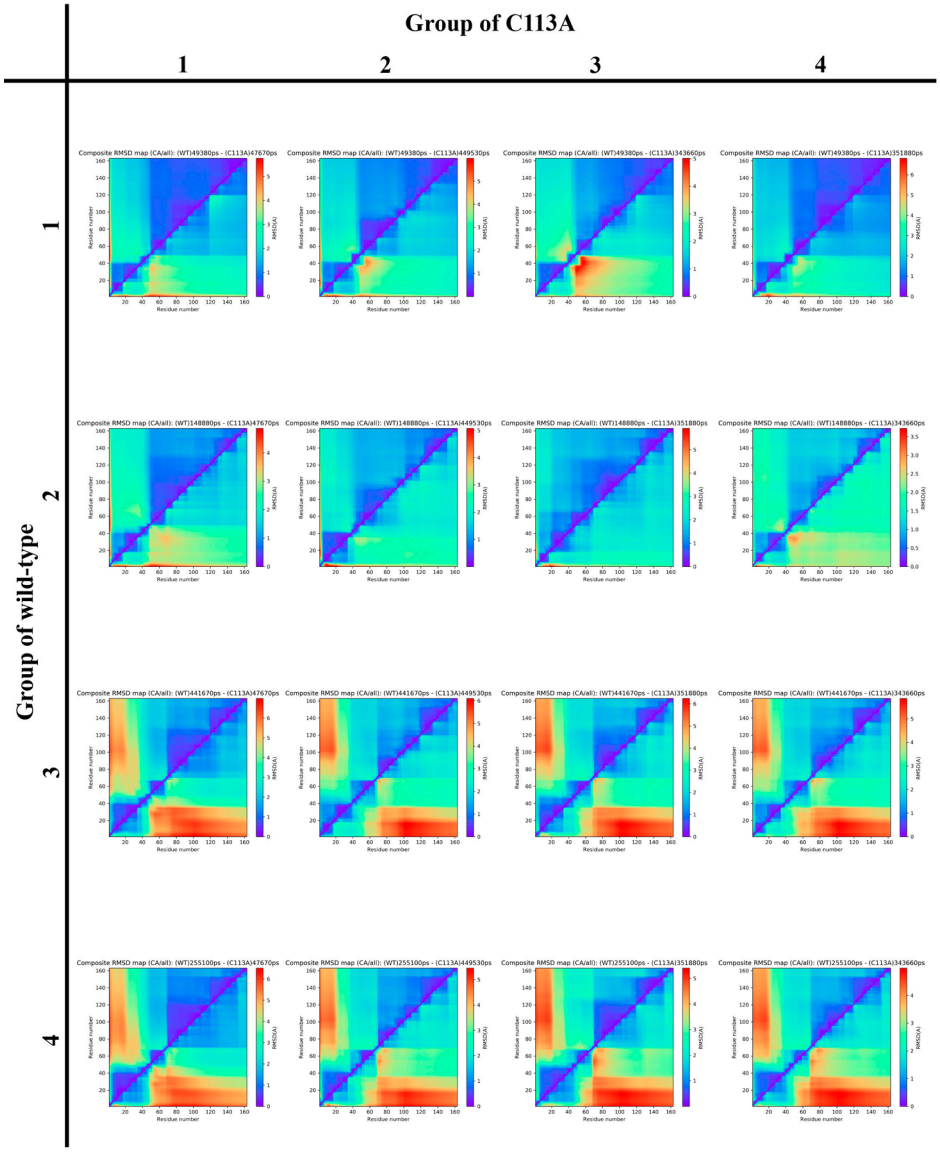


Figure 8 Composite RMSD maps between wild-type and C113A trajectories, which calculated RMSDs of Cα atoms and all heavy atoms for 13,366 segments with all combinations of starting and ending points. Blue squares or triangles represent regions with structural segments common to each of the two groups represented by the snapshot. All the groups of C113A are similar to the first and second groups of wild-type, but very different from the third and fourth groups of wild-type.

the T152-G155 in the vicinity of the substrate binding pocket. Then, the phosphate binding loop may approach the pocket to bind the substrate. In contrast, for the open conformation, hydrogen bonding between G155(N)-T152(O), G155 (O)-T152(N), T152(Oγ)-H155(O), H157(Nδ)-H59(Nε) and H59(Nδ)-C113(Sγ), and a steric hindrance between T152 and H157 kept the conformation of the P149-H157 away from the substrate binding pocket (Fig. 11c). As a result, the phosphate binding loop can leave from the pocket to release the substrate.

The transition between the closed and open conformations was caused by changes in the hydrogen-bonding network which began with the exchange of two hydrogen bonds between H59(Nε)-S115(Oγ) and H59(Nδ)-C113(Sγ). Furthermore, the exchange of the hydrogen bonds occurred simultaneously with the flip-flops of the imidazole ring of H59 (Fig. 12a). Such flip-flops of H59 have been observed since 2014 [9]. Barman and Hamelberg reported that the protonation state of C113 mediated a dynamic hydrogen-bonding network in the active site of Pin1, involving the two adjacent histidines (H59 and H157) and several other residues (S115 and T152) [9]. They calculated the differences in free energy of the four states of the double histidine motif due to the change in protonation of Sγ of C113. They also showed that the flip-flops can occur even during the substrate-binding state. However, they did not investigate

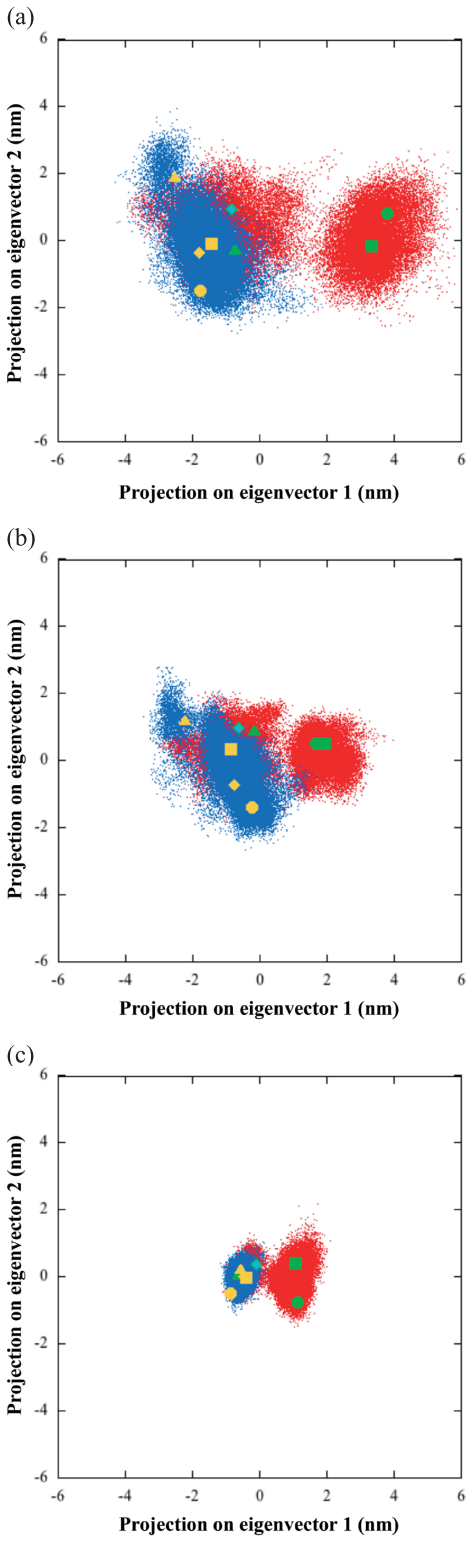


Figure 9 PCA of the whole snapshots consisting of wild-type and C113A trajectories. Projection of the wild-type (red) and C113A (blue) trajectory onto the first 2 eigenvectors. PCA was applied for the region (a) 1–163, (b) 1–66, or (c) 51–163. The representative structures for the first (solid triangles), second (solid diamonds), third (solid circles), and fourth (solid squares) groups, which were shown Figure 4c (green) and Figure 6c (yellow), were plotted on the projections.

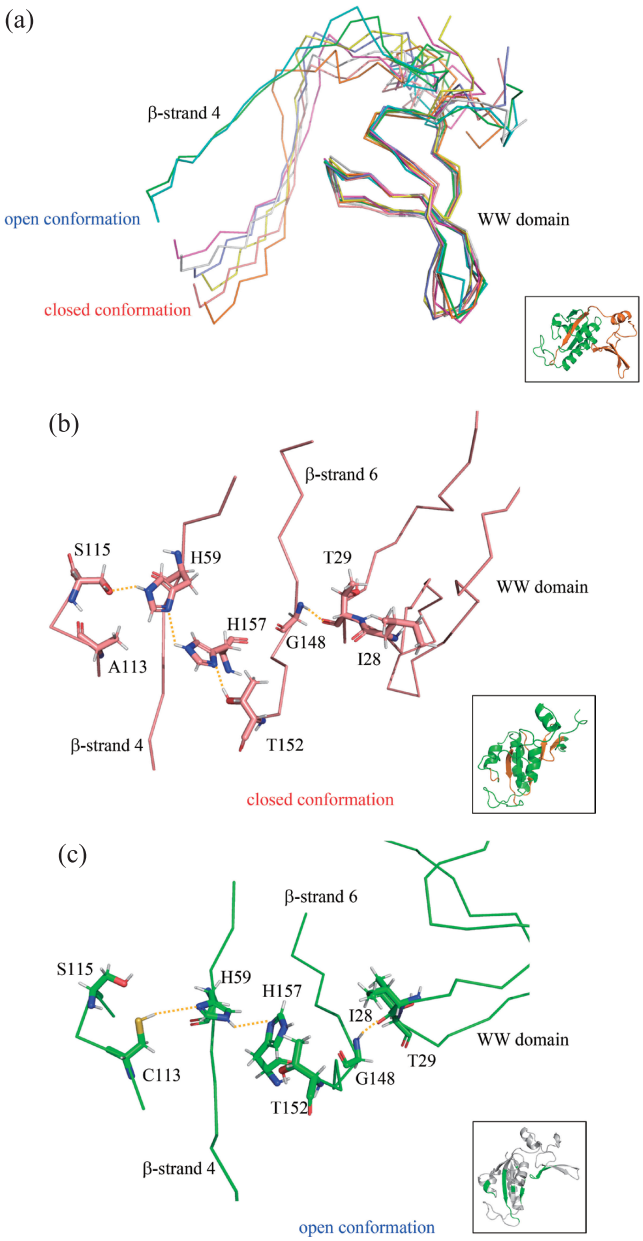


Figure 10 Conformations of the segment corresponding to the sequence M1-Q66. (a) Superposition of the eight segments aligned from M1 to S42. Magenta, yellow, green and cyan stand for the first, second, third and fourth group of wild-type, respectively. Orange, white, slate blue and wheat stand for the first, second, third and fourth group, respectively. (b) Hydrogen-bonding network in the closed conformation of the fourth group of C113A. Hydrogen bonds of T29(O)-G148(N), T152(Oγ)-H157(Nδ), H157(Nε)-H59(Nδ) and H59(Nε)-S115(Oγ) stabilized the orientation of WW domain. (c) Hydrogen-bonding network in the open conformation of the third group of wild-type. Hydrogen bonds of I28(O)-G148(N), H157(Nδ)-H59(Nε) and H59(Nδ)-C113(Sγ) stabilized the orientation of WW domain. Inset shows the view angle.

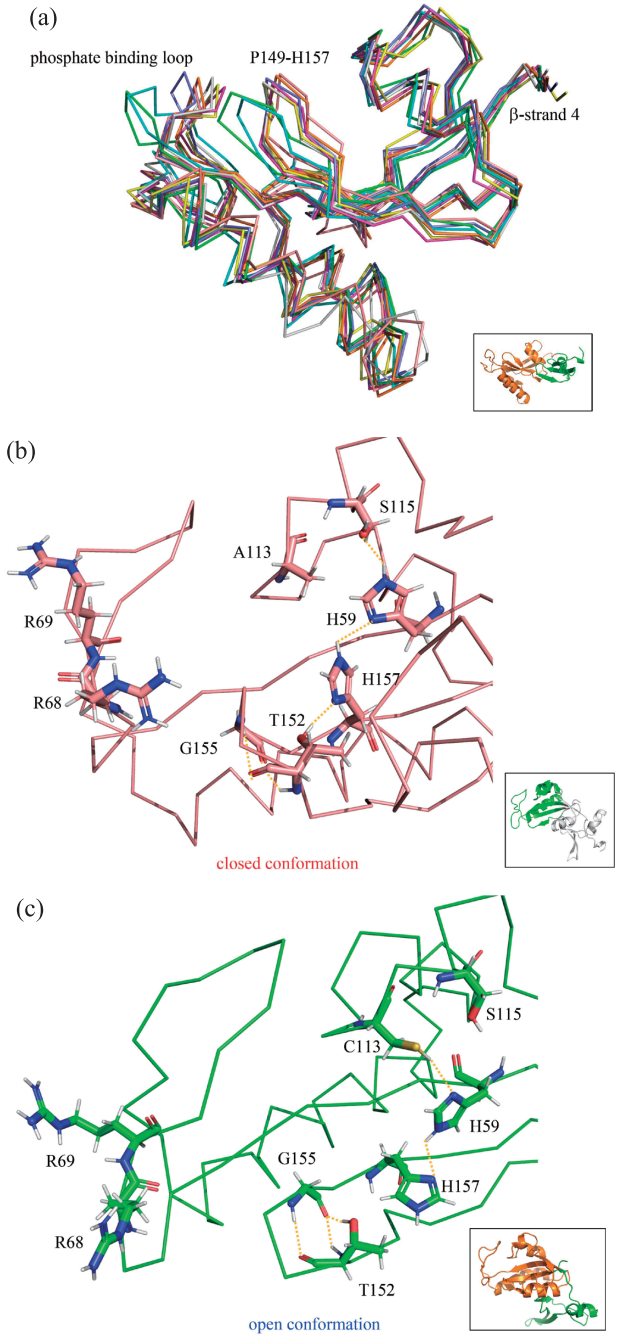


Figure 11 Conformations of the segment corresponding to the sequence E51-E163. (a) Superposition of the eight segments aligned from E51 to Q66. Magenta, yellow, green and cyan stand for the first, second, third and fourth group of wild-type, respectively. Orange, white, slate blue and wheat stand for the first, second, third and fourth group, respectively. (b) Hydrogen-bonding network in the closed conformation of the fourth group of C113A. Hydrogen bonds of G155(N)-T152(O), G155(O)-T152(N), T152(Oγ)-H157(Nδ), H157(Nε)-H59(Nδ) and H59(Nε)-S115(Oγ) stabilized the conformation of the sequence P149-H157. (c) Hydrogen-bonding network in the open conformation of the third group of wild-type. Hydrogen bonds of G155(N)-T152(O), G155 (O)-T152(N), T152(Oγ)-H155(O), H157(Nδ)-H59(Nε) and H59(Nδ)-C113(Sγ), and a steric hindrance between T152 and H157 kept the conformation of the sequence P149-H157 away from the substrate binding pocket. Inset shows the view angle.

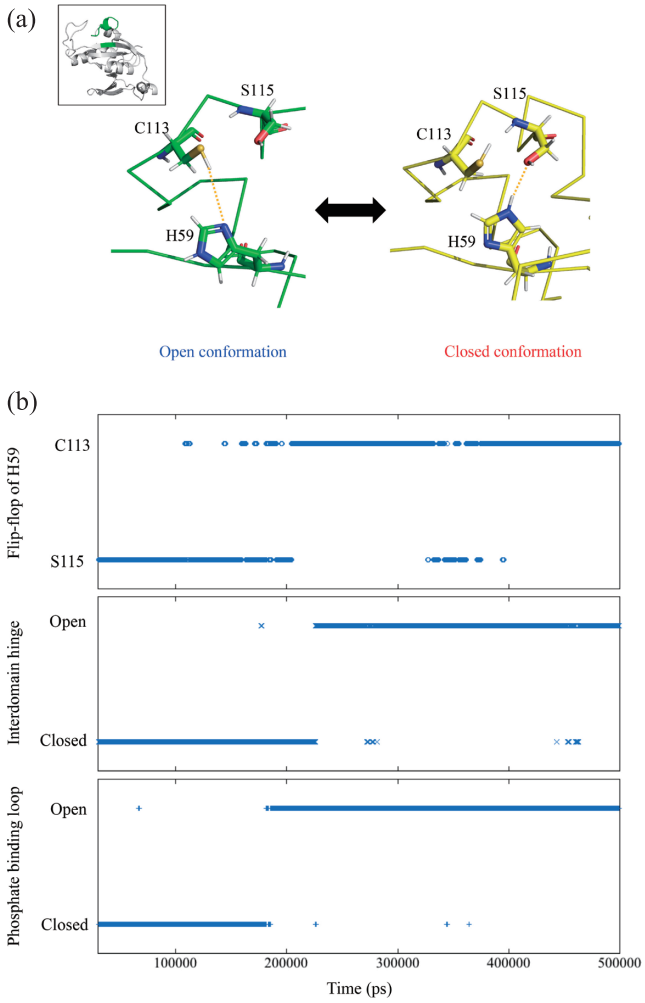


Figure 12 Flip-flop of H59 by exchanging of hydrogen bonds. (a) The transition between the closed and open conformations was caused by changing the hydrogen-bonding network which began with the exchange of two hydrogen bonds between H59(Nε)-S115(Oγ) and H59(Nδ)-C113(Sγ). The left and right figures were extracted from the open conformation of the third group of wild-type and the closed conformation of the second group of wild-type, respectively. Inset shows the view angle. (b) The transition between the closed and open conformations of the interdomain hinge and the phosphate binding loop, and the flip-flop of H59 during the simulation. The labels "C113" and "S115" in the panel "Flip-flop of H59" stand the hydrogen bond formation of H59(Nδ)-C113(Sγ) and H59(Nε)-S115(Oγ), respectively.

how the effect of the dynamic hydrogen bonding network around C113 extended outside of the substrate binding pocket. In the present study, even though the protonation of C113 did not change, the flip-flops occurred, suggesting that changes in protonation of C113 are not essential.

The flip-flop of H59 occurred in all groups but occurred more frequently in the first and second groups than in the third and fourth groups (Fig. 12b). In total, it occurred 153 times in the simulation. This simulation suggests that the hydrogen bond between H59 and S115 is more stabilized in the closed conformation, and the hydrogen bond between H59 and C113 is more stabilized in the open conformation.

The correlation between the flip-flop and the closed-open dynamic was calculated. The correlation coefficient between the flip-flop and the dynamics of phosphate binding loop was 0.78, whereas the coefficient between the flip-flop and the interdomain hinge dynamics was 0.69. The correlation coefficient between the dynamics of the phosphate binding loop and the interdomain hinge was also calculated to be 0.78. Although we cannot conclude that the flip-flop directly triggers closed-open dynamics, we can at least think that the flip-flop is correlated with closed-open dynamics. We consider that the flip-flop triggers closed-open dynamics indirectly via the hydrogen-bonding-network.

The flip-flop of H59 was transmitted via the structural change of T152 in two directions; one was the WW domain, the other was the phosphate binding loop. Since the loop is located at the active site, the mechanism by which conformational changes of H59 and T152 move the phosphate binding loop is relatively easy to understand as an allosteric mechanism. In contrast, an allosteric mechanism in the interdomain hinge between WW domain and PPIase domain was unexpected. In the previous studies, the correlation between the flip-flop of H59 and the interdomain hinge's dynamics was not analyzed because this interdomain hinge was thought to fluctuate independently. However, our correlation analysis clearly shows that the interdomain hinge motion is not independent. This result may give a new interpretation to an experimental report. The mutational effects of I28 and W34 on the PPIase activity were investigated by NMR; the reduced interdomain contact by a mutation I28A allosterically enhanced the PPIase activity, while the increased interdomain contact by a mutation W34A decreased the activity [24]. Although these effects were interpreted as interdomain perturbations, they may be transmitted to the active site via structural changes of T152.

In the mutant C113A, the exchange of the hydrogen bonds cannot occur because of loss of C113 which is essential to H59(Nδ)-C113(Sγ). As a result, the hydrogen bond H59(Nε)-S115(Oγ) stabilizes the conformation of H59 in the closed conformation. This could explain why the mutant C113A took only the closed conformation. According to Xu, N., *et al.* (2014), the mutation C113D stabilized the hydrogen bond between H59 and the 113th amino acid residue (D113) more than that in wild-type Pin1 and the backbone structure of its phosphate binding loop changed to the position away from the substrate binding pocket [10]. This report is consistent with the results of our MD simulations, suggesting that the open conformation requires the hydrogen bond H59(Nδ)-C113(Sγ). As the mutant C113D reduced the PPIase activity of Pin1, however, the balance between the closed and open conformations may be important for enzyme activity.

Is the flip-flop of H59 and the closed-open dynamics the intrinsic behavior of Pin1 or an accidental rare event? As mentioned above, the flip-flop of H59 occurred 153 times. One of the orientations lasted 104,700 ps at maximum as shown in Figure 12b. On the other hand, the exchange between the closed and open conformations occurred at least 17 and 13 times for the interdomain hinge and the phosphate binding loop, respectively (Fig. 12b). One of the conformations lasted 161,510 ps and 135,050 ps at maximum for the interdomain hinge and the phosphate binding loop, respectively. The duration of our MD simulation is only 500 ns. This time is much shorter than Pin1's functional time scale. Therefore, the events observed here must occur much frequently in the real world. Judging from these results, the observed behavior is presumed to be intrinsic to Pin1.

The open conformation of Pin1 is not registered in PDB. One of the reasons may be the problem of crystal contact. Looking at the symmetric molecules in the crystal structure, it was found that the phosphate binding loop was in contact with the symmetric molecules in the closed conformation (data not shown). It is likely that changes in crystal contact would be necessary to allow the loop to adopt its open conformation in the crystal.

In this study, we dealt with a mutant Pin1 as the wild-type protein because only this mutant showed an intact full-length structure. The present results are at least true for this mutant, even if not for wild-type and other mutant proteins. In this study, no direct or indirect interaction between C/A113 and the mutational residues was observed, suggesting that the mutations did not affect the present results. Therefore, we consider that the results obtained with our mutant are true for the real wild-type Pin1.

Conclusions

Our long MD simulations clearly elucidated why the mutation C113A dramatically reduced PPIase activity. We also revealed that the closed-open dynamics of the phosphate binding loop and the interdomain hinge are essential for activity. However, it is still unclear how Pin1 isomerizes the substrate. Further investigation is needed to unravel this fundamental problem.

Acknowledgements

The authors are very grateful to Professor Nobuhiro Go for his constant encouragement through his attitudes toward research. This work was supported by Grants-in-Aid for Scientific Research (C) (JSPS KAKENHI Grant Number JP17K07359) (T. I.).

Conflicts of Interest

The authors declare that they have no conflict of interest.

Author Contributions

T. I. and Y. Y. planed and designed this research. Y. Y. performed MD simulations. T. I. analyzed the MD trajectories. T. I., Y. Y. and N. I. wrote the manuscript.

References

[1] Lu, P. K., Hanes, S. D. & Hunter, T. A human peptidyl–prolyl isomerase essential for regulation of mitosis. *Nature* **380**, 544–547 (1996).

[2] Yaffe, M. B., Schutkowski, M., Shen, M., Zhou, X. Z., Stukenberg, P. T., Rahfeld, J.-U., *et al.* Sequence-Specific and Phosphorylation-Dependent Proline Isomerization: A Potential Mitotic Regulatory Mechanism. *Science* **278**, 1957–1960 (1997).

[3] Ranganathan, R., Lu, K. P., Hunter, T. & Noel, J. P. Structural and Functional Analysis of the Mitotic Rotamase Pin1 Suggests Substrate Recognition Is Phosphorylation Dependent. *Cell* **89**, 875–886 (1997).

[4] Uchida, T., Takamiya, M., Takahashi, M., Miyashita, H., Ikeda, H., Terada, T., *et al.* Pin1 and Par14 Peptidyl Prolyl Isomerase Inhibitors Block Cell Proliferation. *Chem. Biol.* **10**, 15–24 (2003).

[5] Phan, R. T., Saito, M., Kitagawa, Y., Means, A. & Dalla-Favera, R. Genotoxic stress regulates expression of the proto-oncogene Bcl6 in germinal center B cells. *Nat. Immunol.* **8**, 1132–1139 (2007).

[6] Liou, Y. C., Sun, A., Ryo, A., Zhou, X. Z., Yu, Z. X., Huang, H. K., *et al.* Role of the prolyl isomerase Pin1 in protecting against age-dependent neurodegeneration. *Nature* **424**, 556–561 (2003).

[7] Behrsin, C. D., Bailey, M. L., Bateman, K. S., Hamilton, K. S., Wahl, L. M., Brandl, C. J., *et al.* Functionally important residues in the peptidyl-prolyl isomerase Pin1 revealed by unigenic evolution. *J. Mol. Biol.* **365**, 1143–1162 (2007).

[8] Mercedes-Camacho, A. Y., Mullins, A. B., Mason, M. D., Xu, G. G., Mahoney, B. J., Wang, X., *et al.* Kinetic isotope effects support the twisted amide mechanism of Pin1 peptidyl-prolyl isomerase. *Biochemistry* **52**, 7707–7713 (2013).

[9] Barman, A. & Hamelberg, D. Cysteine mediated dynamic hydrogen-bonding network in the active site of Pin1. *Biochemistry* **53**, 3839–3850 (2014).

[10] Xu, N., Tochio, N., Wang, J., Tamari, Y., Uewaki, J., Utsunomiya-Tate, N., *et al.* The C113D mutation in human Pin1 causes allosteric structural changes in the phosphate binding pocket of the PPIase domain through the tug of war in the dual-histidine motif. *Biochemistry* **53**, 5568–5578 (2014).

[11] Wang, J., Tochio, N., Kawasaki, R., Tamari, Y., Xu, N., Uewaki, J., *et al.* Allosteric Breakage of the Hydrogen Bond within the Dual-Histidine Motif in the Active Site of Human Pin1 PPIase. *Biochemistry* **54**, 5242–5253 (2015).

[12] Ikura, T., Tochio, N., Kawasaki, R., Matsuzaki, M., Narita, A., Kikumoto, M., *et al.* The trans isomer of Tau peptide is prone to aggregate, and the WW domain of Pin1 drastically decreases its aggregation. *FEBS Lett.* **592**, 3082–3091 (2018).

[13] 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**, 926–935 (1983).

[14] Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012 (2003).

[15] Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).

[16] Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

[17] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

[18] Parrinello, M. & Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **45**, 1196–1199 (1980).

[19] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

[20] Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).

[21] Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).

[22] Mahoney, B. J., Zhang, M., Zintsmaster, J. S. & Peng, J. W. Extended Impact of Pin1 Catalytic Loop Phosphorylation Revealed by S71E Phosphomimetic. *J. Mol. Biol.* **430**, 710–721 (2018).

[23] Guo, J., Pang, X. & Zhou, H.-X. Two Pathways Mediate Interdomain Allosteric Regulation in Pin1. *Structure* **23**, 237–247 (2015).

[24] Wang, X., Mahoney, B. J., Zhang, M., Zintsmaster, J. S. & Peng, J. W. Negative Regulation of Peptidyl-Prolyl Isomerase Activity by Interdomain Contact in Human Pin1. *Structure* **23**, 2224–2233 (2015).