Exploring the Binding Mechanism between Human Profilin (PFN1) and Polyproline-10 through Binding Mode Screening

Leili Zhang, David R. Bell, Binquan Luan, and Ruhong Zhou*

Computational Biology Center, IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA

Abstract

The large magnitude of protein-protein interaction (PPI) pairs within the human interactome necessitates the development of predictive models and screening tools to better understand this fundamental molecular communication. However, despite enormous efforts from various groups to develop predictive techniques in the last decade, PPI complex structures are in general still very challenging to predict due to the large number of degrees of freedom. In this study, we use the binding complex of human profilin (PFN1) and polyproline-10 (P10) as a model system to examine various approaches, with the aim of going beyond normal protein docking for PPI prediction and evaluation. The potential of mean force (PMF) was first obtained from the time-consuming umbrella sampling, which confirmed that the most stable binding structure identified by the maximal PMF difference is indeed the crystallographic binding structure. Moreover, crucial residues previously identified in experimental studies, W3, H133 and S137 of PFN1, were found to form favorable hydrogen bonds with P10, suggesting a zipping process during the binding between PFN1 and P10. We then explored both regular molecular dynamics (MD) and steered molecular dynamics (SMD) simulations, seeking for better criteria of ranking the PPI prediction. Despite valuable information obtained from conventional MD simulations, neither the commonly used interaction energy between the two binding parties nor the long-term root mean square displacement (RMSD) correlates well with the PMF results. On the other hand, with a sizable collection of trajectories, we demonstrated that the average rupture work calculated from SMD simulations correlates fairly well with the PMFs (R^2 = 0.67), making it a promising PPI screening method.

Keywords: PPI, SMD, umbrella sampling, structure screening, human profilin, polyproline

Introduction

Polyproline recognition domains were identified in large varieties of proteins involved in intracellular signaling, including WW domains, SH3 domains, EVH1 domains, GYF domains, PFN proteins, and UEV domain.[1-7] Their binding partners, polyproline or proline rich motifs, were found in proteins of vital importance for drug discoveries, including HIV-1 PTAP motif[7], p53 PXXP motif[8], Zinc finger proteins[9], and Huntingtin protein[10]. So far, the omnipresence of polyproline is still not fully understood. For example, in Huntingtin protein, it was found that the polyglutamine length and polyproline length undergo a co-evolution from primitive organisms to human beings[10], which may indicate a regulatory role of polyproline length in its biological functions. Yet the mechanism how Huntingtin proteins are regulated remains unknown. Therefore, it is critical to understand how proline enriched domains interact with polyproline recognition proteins, and more specifically, how the protein-protein interaction (PPI) interface conformations are determined.

PPI structure prediction is a very challenging problem computationally due to the complexity associated with the underlying multibody interactions.[11] There are a few promising methods developed already, mostly for protein-protein docking, including ClusPro[12], HADDOCK[13], pyDockWeb[14], and GRAMM-X[15]. These methods have proven successful in certain systems, especially when proteins are rigid and/or small; however, none of them could provide reliable criteria to determine the most trustworthy prediction. Every year the Critical Assessment of PRediction of Interactions (CAPRI) contest[16] is held multiple times to test and improve various methods and protocols. Within the CAPRI framework, the contestants can submit up to 10 structures (with preferences), and the final scores are assessed based on how many hits are predicted successfully. Generally since there is no crystal structure for comparison, in past practices people used multiple models and compared the predictions, or used other empirical scoring functions to narrow down the selections.[17-18] Unfortunately, conflicting and confusing results from different models are frequently obtained, which complicates further assessments of the PPI structures. To overcome this shortcoming, we explore the idea of applying free energy calculations, conventional molecular dynamics (MD) simulations, and steered molecular

dynamics (SMD) simulations to enhance the PPI protein docking scores (by the previously benchmarked GRAMM-X[15]) for better ranking of all predicted structures using state-of-the-art molecular modeling techniques.

Specifically, we use human profilin (PFN1) and polyproline-10 (P10) as a model system. The PPI pair PFN1-P10 is picked due to its importance and simplicity: P10 is a relatively rigid peptide adopting a polyproline II (PPII) helix[19] that binds to surface aromatic residues of PFN1. The crystal structure of PFN1-P10 has been determined by two independent experimental groups[20-21], proving the benchmark for the predictive computational methods. Meanwhile a challenging aspect of this system is that proline has unusual solution properties, identified as an anomalous residue in terms of its hydrophobicity[22-23]. Thus, PFN1-P10 provides us with two possible validation tests within one system. The first is to validate if the force field guided simulations are able to reproduce experimental findings. We conducted free energy calculations to compare the crystal structure with the best structure predicted and also to investigate its associated binding mechanism. Along the latter line, we uncovered a zipping process during the binding of P10 with PFN1, highlighting the importance of W3, H133 and S137 of PFN1. The second is to examine MD simulations, steered molecular dynamics (SMD) simulations, and free energy calculations in comparison with the protein docking score (from GRAMM-X) in order to improve the ranking of docked structures and potentially go beyond protein-protein docking. We found that by following the protocol of using SMD simulation on the GRAMM-X docked structures, we were able to rank the relative stability of the PPI docked structures with reasonably high confidence.

Methods

Molecular docking and MD simulations

The PFN1 and P10 structures were taken from the X-ray co-crystal binding complex structure deposited in protein data bank (PDB ID: 1AWI[20]). Each of the individual protein structures was uploaded separately to the previously benchmarked protein-protein docking web server GRAMM-X[15] (http://vakser.compbio.ku.edu/resources/gramm/grammx/) to obtain the docked

complex structures (referred to as binding modes hereafter). The top 10 binding modes (ranked by scores, mode1 to mode10) were chosen for the latter simulations, along with the crystal structure (named mode11).

MD simulations were performed with GROMACS 5.1.2 package[24]. The 11 PFN1-P10 binding modes were evaluated in simulations using the OPLS-AA force field[25] with virtual sites for hydrogen atoms. Following similar protocols in our previous studies[26-35], all systems were solvated in 8×8×8 nm^3 TIP3P water boxes with 150 mM NaCl. Steepest descent method was used to minimize the solvated PFN1-P10 complexes for 10000 steps. The electrostatic interactions were calculated with particle mesh Ewald (PME) method, while the van der Waals (VDW) interactions were handled with smooth cutoffs with the cutoff distance set to 1 nm. For each mode, a 50 ps of isochoric-isothermic (NVT, 310K) simulation with 1 fs timestep were then performed to equilibrate the systems. A series of 400 ns isobaric-isothermic (NPT, 310K, 1bar) simulations with 4 fs timestep were performed during production runs.

Interaction energy (IE)

We recorded the snapshots every 40 ps from the 400-ns MD simulations mentioned above. Water and salt were not included in the calculations of IE. With the same cutoff scheme and periodic boundary conditions as the MD simulations, we calculated the total energy of PFN1-P10 complexes ($E_{\mathrm{PFN1\text{-}P10}}$) and the energy of individual PFN1, P10 domains (E_{PFN1}, E_{P10}) *in vacuo*. The IE is therefore calculated as:

$$IE = E_{\mathrm{PFN1-P10}} - E_{\mathrm{PFN1}} - E_{\mathrm{P10}} \quad (1)$$

Umbrella sampling and PMF

Umbrella sampling was performed using the minimized PFN1-P10 complexes to obtain the binding free energies. Harmonic restraints between alpha carbons (with force constants set to 1000 kJ/mol/nm^2) were imposed on the individual domains (PFN1 and P10) to prevent them from unfolding. We utilized the center of mass (COM) distance between PFN1 and P10 as the collective variable. The window size of umbrella sampling was set as 0.1 nm, resulting in roughly 20 windows for any PFN1-P10 complexes. The simulations lasted 20 ns for each window. PME, VDW interactions and neighbor-list settings were the same as those used in MD

simulations. Due to the rigidity of the system, only the translational and rotational corrections are needed to calculate the binding free energy of each binding mode.[36] We omit this step because this correction is a constant offset for PFN1-P10 system, and absolute binding free energy is not the main interest in this study.

Steered molecular dynamics (SMD) and rupture work

Constant velocity SMD[37] was adopted to calculate the rupture force and rupture work of PFN1-P10 complexes. Two velocities were tested in this research (1 nm/ns and 0.1 nm/ns) along the COM distance collective variable. A total of 10 replicas were performed for each of the complexes at both velocities. Isochoric-isothermic (NVT) ensemble was used for the simulations, while all other conditions remain the same as those used in the MD simulations. The force spectra were recorded during the SMD simulations at a 0.4 ps interval. Simple maximal forces were extracted from the force spectra and the averages among 10 replicas were reported in **Table** S1. Integrations of the force spectra over the COM distances were calculated, where the maximal work in the integrated curve (see **supporting information**) was defined as the rupture work.

Results and discussions

GRAMM-X prefers hydrophobic binding interfaces for PFN1-P10

We use a previously benchmarked GRAMM-X[15] to obtain the initial PFN1-P10 binding structures to determine the binding modes. The rigid docking algorithm is suitable for PFN1-P10 based on the root mean square displacements (RMSDs) of PFN1 and P10 during the MD simulations (see below). **Figure 1** shows the top 10 docked protein-protein structures ranked by the prediction scores from mode1 to mode10 (most stable to least stable) as well as the crystal structure mode11 (PDB ID: 1AWI) added for completeness. Note that the co-crystal structure binding mode was predicted by GRAMM-X (mode9).

We then analyzed interfacial residue binding for all 11 binding modes. By the proximity of the binding interfaces, we divided them into 4 categories. Category I (mode1, mode2, mode8) consists of a mostly hydrophobic binding interface on PFN1 (here we refer to PFN1 only

because P10 is homogeneous and is therefore omitted later) with residues W3, N4, I7, D8, M11, A12, C16, Q17, S29, V30, W31, A32, A33, V34, P35, and K115 (62.5% hydrophobic). Category II (mode3, mode6, mode10) utilizes a mixed hydrophobic/hydrophilic binding interface, consisting of Y24, K25, D26, S27, P28, D41, T43, P44, A45, E46, V47, G48, V49, V51, G52, K53, D54, S57, F58, N61, G62, L63, T64, and G67 (37.5% hydrophobic). Category III (mode4, mode5, mode7) is also featured by a mixed hydrophobic/hydrophilic binding interface, but consisting of a different set of residues Y59, S61, V72, I73, R74, D86, R88, A95, P96, T97, N99, E116, G117, V118, H119, G120, G121, N124, and K125 (31.6% hydrophobic). Finally, Category IV (mode9, mode11) makes the use of the binding groove from the crystal structure between the N-terminal α helix and C-terminal α helix, mainly contributed by G2, W3, N4, Y6, D26, S27, S29, W31, H133, L134, S137, and Y139 (41.7% hydrophobic). Since both the highest scores mode1 and mode2 are in Category I, GRAMM-X seems to prefer a more hydrophobic binding interface for PFN1-P10, although it also picks out mixed binding interfaces, including the co-crystal binding interface – mode9 in Category IV.

These 11 binding modes are used as the starting points for the binding free energy calculations with umbrella sampling, as well as MD simulations and SMD simulations. In the following sections we will explore these different methods in an effort to go beyond the normal (rigid) protein-protein docking for a better ranking of the binding strengths for the 11 binding modes.

PMFs from umbrella sampling predict crystal structures to be the most stable binding modes

Umbrella sampling was adopted here to estimate the binding free energies of the 11 binding modes for the purpose of finding a practical approach of ranking binding modes from PPI docking programs and to explore the associated binding mechanism. The most rigorous way of calculating binding free energies has been discussed in multiple previous papers and is not the main focus in this study.[36, 38] Here, we used a simplified way of finding the relative binding free energy to ensure specifically that the calculated PMF differences directly correlate with the binding mode stability, by setting the two end points to be the final binding modes (shown in **Figure** 1) and their respective unbound states (defined when the minimal distance between the

two proteins, with the same conformations as in the final binding mode, is larger than 1 nm). The structures of PFN1 and P10 are restricted based on their relatively high rigidity found in the MD simulations (discussed below). This also makes the prediction of the trend of binding free energies straightforward as only a constant translational/rotational correction is needed (which is therefore omitted). Additionally, because free energy is a thermodynamic state function, the difference between the bound state and unbound state is a constant regardless of paths. With the reasons stated above, we conclude that the trend found in PMF differences between the bound state and the unbound state is the same as the trend in binding free energies.

The obtained PMF curves are shown in **Figure** 2A. The lowest binding free energies come from mode7 (Category III), mode9 and mode11 (Category IV). One important finding is that OPLS-AA force field is capable of predicting the most stable binding structure of PFN1-P10, namely the Category IV co-crystal structures (mode9, mode11). A closer look at the binding interfaces reveals that hydrophobic interfaces such as mode1 leads to a roughly 6 kcal/mol penalty compared to the mixed binding interfaces provided by mode7, mode10 and mode11 (**Figure** 2B). Such findings further solidify the argument that polyprolines are not as hydrophobic as implied in some hydrophobicity scales[39-40]. In the next section we discuss the detailed binding mechanism unveiled from umbrella-sampling simulations.

Binding mechanism of P10 in the crystal binding pocket of PFN1

To examine the binding mechanism between PFN1 and P10, we analyzed the number of hydrogen bonds between them during the umbrella sampling simulations with Category IV binding modes (mode9, mode11). The time-lapse counts of hydrogen bonds paired with the representative snapshots are shown in **Figure** 3 (mode9) and **Figure** S1 (mode11). Residues W3, Y6, H133, S137 and Y139 are found to contribute significantly to the PFN1-P10 contacts, in which W3[41], H133[41] and S137[42] were also previously suggested to be crucial experimentally. The main contributions from tryptophans and tyrosines are consistent with previous experimental findings that prolines tend to form aromatic-proline stacking[43], which is only seen in the crystal binding pocket (Category IV). These observations suggest that ideally, scoring functions should be tuned so that such interaction can be better captured.

Next we examined how the binding process evolves when P10 approaches PFN1. A stepwise view of how hydrogen bonds form and break during umbrella sampling is shown in **Figure** 3, which illustrates how P10 binds PFN1 through a "zipping" mechanism. Hydrogen bonds are first formed between the C-terminus of P10 (residue 9~10) and PFN1. Then the interactions gradually move along the chain to the N-terminal direction, up to the middle portion of the P10 (residue 5~6). The aromatic-proline stacking was also examined during the binding process. We defined an aromatic-proline stacking index (*S*) based on the minimal distances from any proline side chain on P10 to the side chains of W3, Y6, Y139 from PFN1 ($\min(d_{\mathrm{P-W3}})$, $\min(d_{\mathrm{P-Y6}})$, $\min(d_{\mathrm{P\_Y139}})$, respectively):

$$S \equiv \frac{1}{3}\left(f(\min(d_{\mathrm{P-W3}})) + f(\min(d_{\mathrm{P-Y6}})) + f\left(\min\left(d_{\mathrm{P\_Y139}}\right)\right)\right), \quad (2)$$

where

$$f(x) \equiv \begin{cases} 0, & \text{if } x \geq 1 \text{ nm} \\ \frac{1-x}{0.6}, & \text{if } 0.4 \text{ nm} \leq x < 1 \text{ nm}. \\ 1, & \text{if } x < 0.4 \text{ nm} \end{cases} \quad (3)$$

The "cutoff" switching distance of 0.4 nm was used because the average equivalent VDW radius of a residue is roughly 0.4 nm in coarse-grained models. Therefore, the close packing between two residues requires the distance to be smaller than 0.4 nm. For example, S = 0.8 indicates that the average minimal distance between any proline in P10 to W3, Y6, Y139 from PFN1 is 0.52 nm. **Figure** S2 shows that aromatic-proline stacking reaches the highest strength when P10 is close to PFN1 with a step function type of behavior. Combining **Figure** 3 and **Figure** S2, we observed a concerted formation of hydrogen bonds and aromatic-proline stacking when P10 approaches PFN1. This "zipping" mechanism is reproducible with two different binding modes for the crystal binding pocket: mode9 (**Figure** 3) and mode11 (**Figure** S1). Thus, in addition to the conventional aromatic-proline interaction dominated mechanism, we demonstrated that hydrogen bonds between PFN1 residues (serines, tyrosines and tryptophans) and prolines also contribute significantly to the interactions, thus resulting in a "zipping" mechanism. Such observation also helps to explain the "anomaly" of the hydrophobicity of prolines discussed before[22].

MD simulations disfavor hydrophobic binding interfaces

The most straightforward way of testing binding interface stability is to calculate the RMSDs of the binding modes from MD simulations. For each of the 11 initial structures prepared above, we run 400 ns of unrestrained MD simulations. We first looked at the RMSDs of individual proteins to examine their rigidity. RMSDs of PFN1 with respect to aligned initial structures are plotted in **Figure** S3A. With values always below 0.4 nm, PFN1 stays rigid throughout all simulations for all 11 binding modes, agreeing with the previous experimental findings[20]. Not surprisingly, the RMSDs of P10 with respect to the aligned initial structure (**Figure** S3B) stay below 0.3 nm. The structure of P10 stays as a polyproline II (PPII) helix throughout all of the simulations. This agrees with the widely accepted notion that polyproline should be a rigid PPII helix when the repeat length is smaller than 10.[19] However, the RMSDs of the PFN1-P10 complex with respect to the aligned initial structure (**Figure** S3C) indicate that the binding interfaces can change significantly in some of the binding modes. For example, mode3 (bright red) reaches as high as 1.2 nm RMSD in **Figure** S3C, while PFN1 and P10 remains stable individually (**Figure** S3A and **Figure** S3B). The RMSDs of P10 with respect to the initial complex structure (not aligned, **Figure** S3D) further fortify this observation, displaying a similar trend to **Figure** S3C.

To further observe what contributes to the RMSDs, we plotted the interfacial residue frequency in **Figure** S4, with the initial occurrence frequency (0-50 ns) shown in **Figure** S4A, and final frequency (200-400 ns) shown in **Figure** S4B. The binding modes that feature major interfacial residue shifts are mode2, mode3, mode6, mode7 and mode8, mainly from Category I and Category II binding modes. More specifically, mode2 and mode8 in Category I deviate from the original hydrophobic interfacial binding, with mode2 transitioning to Category III and mode8 transitioning to Category IV, respectively. Even the relatively stable binding mode in Category I (mode1) shows a slight increase in the C-terminal region, which is the signature binding site of the co-crystal structure (Category IV). The mixed hydrophobic/hydrophilic binding interface Category II is also not very stable, with mode3 moving towards a hybrid Category II and Category III binding mode, and mode6 moving towards Category IV binding mode. Overall, all binding modes from Category I (mode1, mode2, mode8) and Category II (mode3, mode6, mode10) have tendencies to shift towards Category III and Category IV, which clearly indicates that MD simulations prefer the mixed hydrophobic/hydrophilic binding interfaces for P10. This

is also phenomenologically in agreement with the umbrella sampling which predicts mode7 (Category III), mode9 and mode11 (Category IV) to be the most stable binding modes. Additionally, PFN1 C-terminus plays an important role in the final binding interfaces of mode6, mode7, mode8, mode9 and mode11. Without doubt, MD simulations (with the OPLS-AA force field) are capable to capture relevant binding features between PFN1 and P10. To further investigate if direct MD data are sufficient to warrant a reliable way of ranking the binding structures, we make several measurements discussed below.

To differentiate the short-term stability and long-term stability of the binding interfaces, we calculated the average RMSDs over 0-50 ns and 200-400 ns in **Table** S1. Interestingly, the trends of the two calculated RMSDs are vastly different. We found a positive correlation between short-term (50 ns) RMSD and PMF differences. In contrast, no correlation was found between long-term (400 ns) RMSD and PMF differences (See **Supporting Information** and **Figure** S5 for details).

Interaction energies (IEs) of PFN1-P10 are calculated with **Equation** 1 (see method for details). Similar to RMSDs, we summarized the average IEs from 0-50 ns and 200-400 ns in **Table** S1, respectively. As expected, the direct IEs are unable to recover the trend in binding free energies, regardless of short-term IEs or long-term IEs (see **Supporting Information** and **Figure** S5 for details), due to the lack of entropy contributions.

Rank binding structures with SMD results

Based on Jarzynski's inequality[44], the experimental atomic force microscopy (AFM) method has proven capable of recovering the binding free energies of complicated systems such as protein-drug complexes.[45] Concurrently, substantial efforts have been applied to adjust and optimize SMD to obtain accurate binding free energies computationally using the same inequlaity.[46-53] Note that even though the original Jarzynski's inequality does not require a working threshold of the rupture rate, it was largely accepted the slower the rate is, the more accurate the results will be.[47] In general, there are limitations in using SMD for free energy calculations due to insufficient sampling sizes; however, it is still worth comparing SMD to other PPI methods.

Without applying the optimization suggested for calculating binding free energies, here we test the applicability of using SMD as a fast method to rank the docked structures of PPI pairs such as PFN1-P10. We compare constant velocity SMD simulations with different pulling speeds (1 nm/ns and 0.1 nm/ns) to the properties acquired from MD simulations (RMSDs and IEs). The basic protocols of SMD can be found in the Methods section. The integrated work curves are shown in **Figure** S6, **Figure** S7. The maximal forces and rupture works are listed in **Table** S1.

With the simplest first order cumulant expansion approximation, we calculated the average rupture works from 10 replicas of SMD simulations for each of the PFN1-P10 binding modes. The rupture works calculated from two rupture speeds (1nm/ns and 0.1 nm/ns) both correlate well with the PMF differences (**Figure** 4 A, **Figure** 4 C), where R^2 is 0.49 for 1 nm/ns SMD, and 0.67 for 0.1 nm/ns SMD. The absolute values of rupture works from 0.1nm/ns SMD (<-15 kcal/mol) are closer to the PMF differences (see **Table** S1 for details), compared to the range (-10 ~ -30 kcal/mol) from 1 nm/ns SMD. This is in agreement with previous practice on free energy estimations from SMD simulations.[47] For our purpose, instead of following the common sense of "the slower the better" in the field (0.01 or even 0.001 nm/ns), we use a relatively fast rupture speed (1 or 0.1 nm/ns) to rank the docked structures. For example, each simulation only takes ~2.5 ns or ~25 ns for SMD with 1nm/ns or 0.1nm/ns rupture speed, respectively. This means that from 2.5 ns×10 replicas = 25 ns of simulations in total, a trend in PMFs can be predicted by SMD with high confidence, as compared to the 50 ns of MD simulation sampling a local well on the free energy landscape, or 20×20 = 400 ns of umbrella sampling for numerous binding modes from one PPI pair. Moreover, the replica numbers can be reduced to 6 to reach a comparable R^2 (see **Figure** S8 A, **Figure** S8 B) which further decreases the total simulation time to 15 ns (for 1 nm/ns SMD).

A correlation between AFM maximal forces and binding constant was reported in a cell adhesion study[54] and an antibody-antigen unbinding study[54]. Here we investigate how maximal forces from SMD (with a much faster rupture rate) correlate with the PMFs. With two rupture rates, R^2 is calculated to be 0.53 for 1 nm/ns SMD, and 0.21 for 0.1 nm/ns SMD, respectively. Dependent on the separation pathway selected on a free energy landscape, overall maximal forces are unsatisfactory for predicting the trend of binding free energies. The correlation may simply come

from the one between maximal forces and rupture works (R^2 is 0.65 for 1 nm/ns SMD and 0.51 for 0.1 nm/ns SMD) in this particular practice.

In **Figure** S9, we compare how the 5 properties (average RMSDs from MD over the first 50ns, maximal forces from two sets of SMD, and rupture works from two sets of SMD) rank the 11 binding modes as compared with PMFs. The correlation coefficients between the ranks are similar compared to the correlation coefficients between the absolute values (**Figure** S9 A-E), ruling out the "clustering effect" of the data (meaning the ranks may be interchangeable when absolute numbers are close). In **Figure** S9 F, we list the most stable binding modes predicted from the 5 metrics. Interestingly, mode7 is the most stable binding mode for 4 out of 5 metrics except for the rupture work (from 1 nm/ns SMD) where mode4 (in Category III along with mode7) is predicted to be the most stable. The ranks of mode11 are reliable except for maximal forces (from 0.1 nm/ns SMD). The ranks of mode9 are also high in the list but may be affected by its structural flexibility seen in the MD simulations. Notably, the top ranks derived from the rupture works from SMD correspond well to the most stable binding structures obtained from the umbrella sampling, indicating that this technique could provide a fast and reliable way of ranking the PPI binding modes from the protein docking programs such as GRAMM-X.

Conclusion

In this paper we systematically study the PPI binding structures of PFN1-P10 using the protein docking program GRAMM-X, regular MD simulations, free energy methods (umbrella sampling) and SMD simulations, with the aim of going beyond normal protein docking for PPI prediction and evaluation. We demonstrate that the OPLS-AA force field guided umbrella sampling is able to identify the crystal structure as the most stable binding structure, which also appears in the top 10 list from GRAMM-X. Our comparative analysis shows that aromatic-proline stacking contributes the most to the stabilization of PFN1-P10 binding, along with the formation of hydrogen bonds between serines/tyrosines/tryptophans and prolines. Although regular MD simulations provide mixed information in terms of predicting the most stable binding structure, yet we find that 50 ns RMSDs might be useful with cautions for ranking the PPI prediction (see **Supporting Information** for details). On the other hand, SMD simulations provide a fast and

reliable way of identifying the binding mode with the lowest binding free energy, as shown by a correlation coefficient (R^2) as high as 0.7 (from the rank order correlation). We argue that even though the rigidity of both PFN1 and P10 may prevent us from generalizing the conclusion, by applying some constraints, SMD can be used to quickly screen the stable binding modes found from docking programs. With a clearer understanding of the advantages and disadvantages of various PPI techniques tested, we hope to expand them with a fast binding-mode-screening method to study PPI pairs found in the human interactome network[55].

Supporting Information

Ranking binding structures with MD results. Summarized data of MD simulations (RMSDs, IEs), umbrella sampling simulations (PMFs), and SMD simulations (maximal forces, rupture works) in **Table** S1. Binding mechanism of mode11 in **Figure** S1. Aromatic-proline stacking index measurements for umbrella sampling in **Figure** S2. RMSD of PFN, P10 and the whole complex from MD simulations in **Figure** S3. Initial binding interface and final binding interface from MD simulations in **Figure** S4. Correlation between MD simulation results and PMF differences in **Figure** S5. SMD work curves of PFN-P10 in **Figure** S6, **Figure** S7. Correlation coefficient analyses versus number of trials in **Figure** S8. Rank order correlations in **Figure** S9.

Acknowledgement

We would like to thank Joseph A. Morrone, Hongsuk Kang, Frank Vazquez, Sangyun Lee, Serena Chen, Leticia Toledo-Sherman and Tien Huynh for their help with work. This work was partially funded by CHDI Foundation Inc., a charitable foundation that funds research into Huntington's disease. RZ acknowledges the support of the IBM Blue Gene Science Program (W1258591, W1464125, W1464164).

References

1. Zarrinpar, A.; Bhattacharyya, R. P.; Lim, W. A., The Structure and Function of Proline Recognition Domains. *Homo* **2003,** *332* (80), 20.
2. Mayer, B. J., Sh3 Domains: Complexity in Moderation. *Journal of cell science* **2001,** *114* (7), 1253-1263.
3. Verdecia, M. A.; Bowman, M. E.; Lu, K. P.; Hunter, T.; Noel, J. P., Structural Basis for Phosphoserine-Proline Recognition by Group Iv Ww Domains. *Nature Structural & Molecular Biology* **2000,** *7* (8), 639.
4. Tu, J. C.; Xiao, B.; Yuan, J. P.; Lanahan, A. A.; Leoffert, K.; Li, M.; Linden, D. J.; Worley, P. F., Homer Binds a Novel Proline-Rich Motif and Links Group 1 Metabotropic Glutamate Receptors with Ip3 Receptors. *Neuron* **1998,** *21* (4), 717-726.
5. Nishizawa, K.; Freund, C.; Li, J.; Wagner, G.; Reinherz, E. L., Identification of a Proline-Binding Motif Regulating Cd2-Triggered T Lymphocyte Activation. *Proceedings of the National Academy of Sciences* **1998,** *95* (25), 14897-14902.
6. Giesemann, T.; Rathke-Hartlieb, S.; Rothkegel, M.; Bartsch, J. W.; Buchmeier, S.; Jockusch, B. M.; Jockusch, H., A Role for Polyproline Motifs in the Spinal Muscular Atrophy Protein Smn Profilins Bind to and Colocalize with Smn in Nuclear Gems. *Journal of Biological Chemistry* **1999,** *274* (53), 37908-37914.
7. Pornillos, O.; Alam, S. L.; Davis, D. R.; Sundquist, W. I., Structure of the Tsg101 Uev Domain in Complex with the Ptap Motif of the Hiv-1 P6 Protein. *Nature Structural & Molecular Biology* **2002,** *9* (11), 812.
8. Baptiste, N.; Friedlander, P.; Chen, X.; Prives, C., The Proline-Rich Domain of P53 Is Required for Cooperation with Anti-Neoplastic Agents to Promote Apoptosis of Tumor Cells. *Oncogene* **2002,** *21* (1), 9.
9. Morgan, A. A.; Rubenstein, E., Proline: The Distribution, Frequency, Positioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human Proteome. *PloS one* **2013,** *8* (1), e53785.
10. Zuccato, C.; Valenza, M.; Cattaneo, E., Molecular Mechanisms and Potential Therapeutical Targets in Huntington's Disease. *Physiological reviews* **2010,** *90* (3), 905-981.
11. Smith, G. R.; Sternberg, M. J. E., Prediction of Protein–Protein Interactions by Docking Methods. *Current Opinion in Structural Biology* **2002,** *12* (1), 28-35.
12. Comeau, S. R.; Kozakov, D.; Brenke, R.; Shen, Y.; Beglov, D.; Vajda, S., Cluspro: Performance in Capri Rounds 6–11 and the New Server. *Proteins: Structure, Function, and Bioinformatics* **2007,** *69* (4), 781-785.
13. De Vries, S. J.; Van Dijk, M.; Bonvin, A. M., The Haddock Web Server for Data-Driven Biomolecular Docking. *Nature protocols* **2010,** *5* (5), 883.
14. Jiménez-García, B.; Pons, C.; Fernández-Recio, J., Pydockweb: A Web Server for Rigid-Body Protein–Protein Docking Using Electrostatics and Desolvation Scoring. *Bioinformatics* **2013,** *29* (13), 1698-1699.
15. Tovchigrechko, A.; Vakser, I. A., Gramm-X Public Web Server for Protein–Protein Docking. *Nucleic acids research* **2006,** *34* (suppl_2), W310-W314.
16. Janin, J., Assessing Predictions of Protein–Protein Interaction: The Capri Experiment. *Protein science* **2005,** *14* (2), 278-283.
17. Maheshwari, S.; Brylinski, M., Predicted Binding Site Information Improves Model Ranking in Protein Docking Using Experimental and Computer-Generated Target Structures. *BMC structural biology* **2015,** *15* (1), 23.
18. Xue, L. C.; Jordan, R. A.; El-Manzalawy, Y.; Dobbs, D.; Honavar, V. In *Ranking Docked Models of Protein-Protein Complexes Using Predicted Partner-Specific Protein-Protein Interfaces: A Preliminary Study*, Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine, ACM: 2011; pp 441-445.
19. Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A., Polyproline and the "Spectroscopic Ruler" Revisited with Single-Molecule Fluorescence. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (8), 2754-2759.
20. Mahoney, N. M.; Janmey, P. A.; Almo, S. C., Structure of the Profilin-Poly-L-Proline Complex Involved in Morphogenesis and Cytoskeletal Regulation. *Nature Structural & Molecular Biology* **1997,** *4* (11), 953-960.
21. Ferron, F.; Rebowski, G.; Lee, S. H.; Dominguez, R., Structural Basis for the Recruitment of Profilin–Actin Complexes During Filament Elongation by Ena/Vasp. *The EMBO Journal* **2007,** *26* (21), 4597-4606.
22. Schobert, B.; Tschesche, H., Unusual Solution Properties of Proline and Its Interaction with Proteins. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1978,** *541* (2), 270-277.
23. Gibbs, P.; Radzicka, A.; Wolfenden, R., The Anomalous Hydrophilic Character of Proline. *J. Am. Chem. Soc.* **1991,** *113* (12), 4714-4715.
24. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015,** *1*, 19-25.

25. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and Reparametrization of the Opls-Aa Force Field for Proteins Via Comparison with Accurate Quantum Chemical Calculations on Peptides. *The Journal of Physical Chemistry B* **2001,** *105* (28), 6474-6487.
26. Zhou, R.; Berne, B. J.; Germain, R., The Free Energy Landscape for B Hairpin Folding in Explicit Water. *Proceedings of the National Academy of Sciences* **2001,** *98* (26), 14931-14936.
27. Das, P.; Li, J.; Royyuru, A. K.; Zhou, R., Free Energy Simulations Reveal a Double Mutant Avian H5n1 Virus Hemagglutinin with Altered Receptor Binding Specificity. *Journal of computational chemistry* **2009,** *30* (11), 1654-1663.
28. Xia, Z.; Clark, P.; Huynh, T.; Loher, P.; Zhao, Y.; Chen, H.-W.; Rigoutsos, I.; Zhou, R., Molecular Dynamics Simulations of Ago Silencing Complexes Reveal a Large Repertoire of Admissible 'Seed-Less' Targets. *Scientific reports* **2012,** *2*, 569.
29. Xiu, P.; Yang, Z.; Zhou, B.; Das, P.; Fang, H.; Zhou, R., Urea-Induced Drying of Hydrophobic Nanotubes: Comparison of Different Urea Models. *The Journal of Physical Chemistry B* **2011,** *115* (12), 2988-2994.
30. Chong, Y.; Ge, C.; Yang, Z.; Garate, J. A.; Gu, Z.; Weber, J. K.; Liu, J.; Zhou, R., Reduced Cytotoxicity of Graphene Nanosheets Mediated by Blood-Protein Coating. *ACS nano* **2015,** *9* (6), 5713-5724.
31. Luan, B.; Huynh, T.; Zhao, L.; Zhou, R., Potential Toxicity of Graphene to Cell Functions Via Disrupting Protein–Protein Interactions. *ACS nano* **2014,** *9* (1), 663-669.
32. Duan, G.; Kang, S. G.; Tian, X.; Garate, J. A.; Zhao, L.; Ge, C.; Zhou, R., Protein Corona Mitigates the Cytotoxicity of Graphene Oxide by Reducing Its Physical Interaction with Cell Membrane. *Nanoscale* **2015,** *7* (37), 15214-24.
33. Li, J.; Liu, T.; Li, X.; Ye, L.; Chen, H.; Fang, H.; Wu, Z.; Zhou, R., Hydration and Dewetting near Graphite-Ch(3) and Graphite-Cooh Plates. *J. Phys. Chem. B* **2005,** *109* (28), 13639-48.
34. Zhou, R., Exploring the Protein Folding Free Energy Landscape: Coupling Replica Exchange Method with P3me/Respa Algorithm. *J. Mol. Graph. Model.* **2004,** *22* (5), 451-63.
35. Zhou, R.; Gao, H., Cytotoxicity of Graphene: Recent Advances and Future Perspective. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **2014,** *6* (5), 452-74.
36. Pohorille, A.; Jarzynski, C.; Chipot, C., Good Practices in Free-Energy Calculations. *The Journal of Physical Chemistry B* **2010,** *114* (32), 10235-10253.
37. Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K., Steered Molecular Dynamics. *Computational molecular dynamics: challenges, methods, ideas* **1999,** *4*, 39-65.
38. Wang, J.; Deng, Y.; Roux, B., Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophysical journal* **2006,** *91* (8), 2798-2814.
39. Wimley, W. C.; White, S. H., Experimentally Determined Hydrophobicity Scale for Proteins. *Nature structural biology* **1996,** *3* (10).
40. Moon, C. P.; Fleming, K. G., Side-Chain Hydrophobicity Scale Derived from Transmembrane Protein Folding into Lipid Bilayers. *Proceedings of the National Academy of Sciences* **2011,** *108* (25), 10174-10177.
41. Björkegren-Sjögren, C.; Korenbaum, E.; Nordberg, P.; Lindberg, U.; Karlsson, R., Isolation and Characterization of Two Mutants of Human Profilin I That Do Not Bind Poly (L-Proline). *FEBS letters* **1997,** *418* (3), 258-264.
42. Shao, J.; Welch, W. J.; DiProspero, N. A.; Diamond, M. I., Phosphorylation of Profilin by Rock1 Regulates Polyglutamine Aggregation. *Molecular and cellular biology* **2008,** *28* (17), 5196-5208.
43. Zondlo, N. J., Aromatic–Proline Interactions: Electronically Tunable Ch/Π Interactions. *Accounts of chemical research* **2012,** *46* (4), 1039-1049.
44. Jarzynski, C., Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters* **1997,** *78* (14), 2690.
45. Hummer, G.; Szabo, A., Free Energy Reconstruction from Nonequilibrium Single-Molecule Pulling Experiments. *Proceedings of the National Academy of Sciences* **2001,** *98* (7), 3658-3661.
46. Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K., Free Energy Calculation from Steered Molecular Dynamics Simulations Using Jarzynski's Equality. *The Journal of chemical physics* **2003,** *119* (6), 3559-3566.
47. Park, S.; Schulten, K., Calculating Potentials of Mean Force from Steered Molecular Dynamics Simulations. *The Journal of chemical physics* **2004,** *120* (13), 5946-5961.
48. Xiong, H.; Crespo, A.; Marti, M.; Estrin, D.; Roitberg, A. E., Free Energy Calculations with Non-Equilibrium Methods: Applications of the Jarzynski Relationship. *Theoretical Chemistry Accounts* **2006,** *116* (1-3), 338-346.

49. Chen, L. Y., Hybrid Steered Molecular Dynamics Approach to Computing Absolute Binding Free Energy of Ligand–Protein Complexes: A Brute Force Approach That Is Fast and Accurate. *Journal of chemical theory and computation* **2015,** *11* (4), 1928-1938.
50. Cuendet, M. A.; Michielin, O., Protein-Protein Interaction Investigated by Steered Molecular Dynamics: The Tcr-Pmhc Complex. *Biophysical journal* **2008,** *95* (8), 3575-3590.
51. Neumann, J.; Gottschalk, K.-E., The Effect of Different Force Applications on the Protein-Protein Complex Barnase-Barstar. *Biophysical journal* **2009,** *97* (6), 1687-1699.
52. Patel, J. S.; Berteotti, A.; Ronsisvalle, S.; Rocchia, W.; Cavalli, A., Steered Molecular Dynamics Simulations for Studying Protein–Ligand Interaction in Cyclin-Dependent Kinase 5. *Journal of chemical information and modeling* **2014,** *54* (2), 470-480.
53. Nicolini, P.; Frezzato, D.; Gellini, C.; Bizzarri, M.; Chelli, R., Toward Quantitative Estimates of Binding Affinities for Protein–Ligand Systems Involving Large Inhibitor Compounds: A Steered Molecular Dynamics Simulation Route. *Journal of computational chemistry* **2013,** *34* (18), 1561-1576.
54. Bell, G. I., Models for the Specific Adhesion of Cells to Cells. *Science* **1978,** *200* (4342), 618-627.
55. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K. P., String V10: Protein–Protein Interaction Networks, Integrated over the Tree of Life. *Nucleic acids research* **2014,** *43* (D1), D447-D452.

Figures

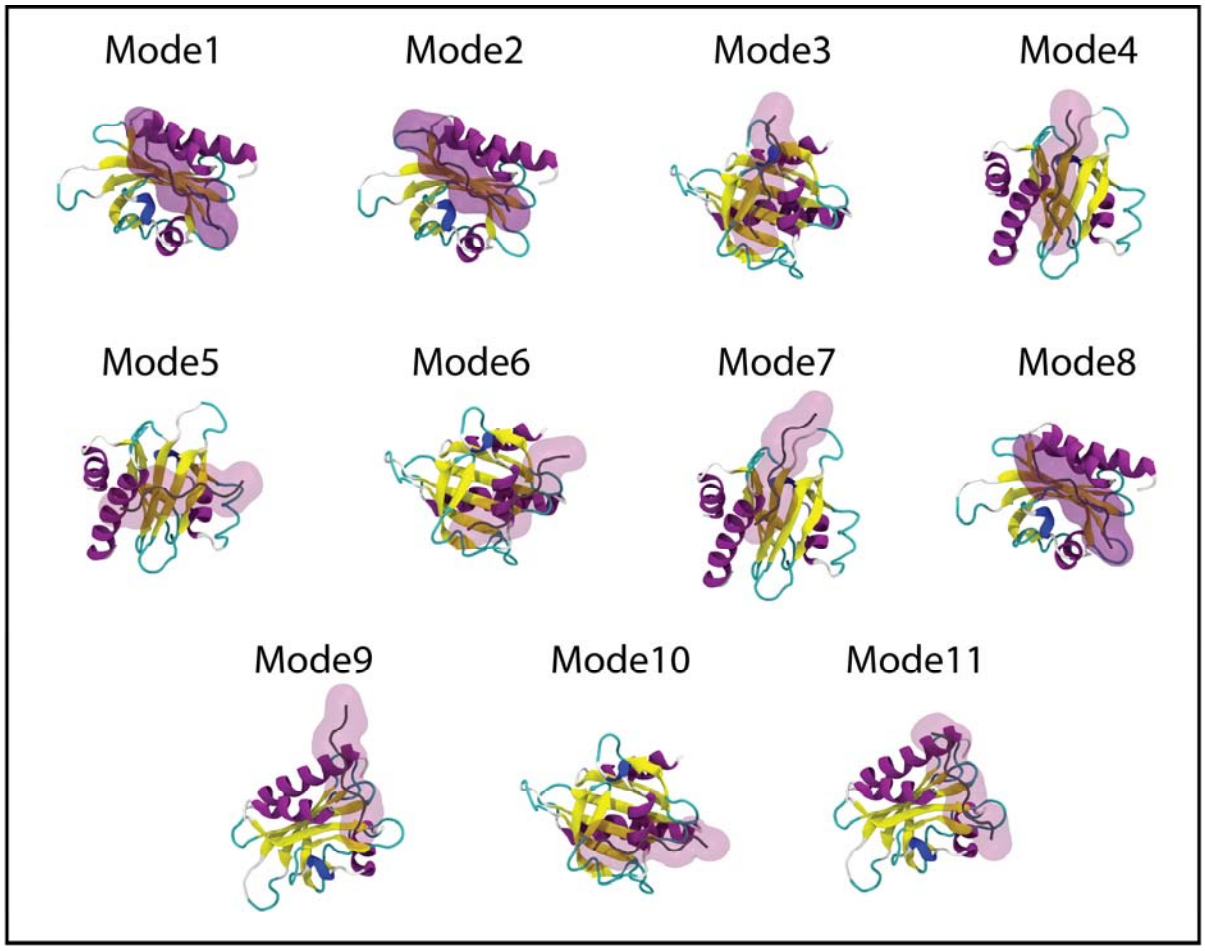

Figure 1. Top 10 scored human profilin (PFN1)-polyproline 10 (P10) docked structures from docking program GRAMM-X (Mode1 to Mode10). Mode11 is taken from the crystal structure (PDB ID: 1AWI).

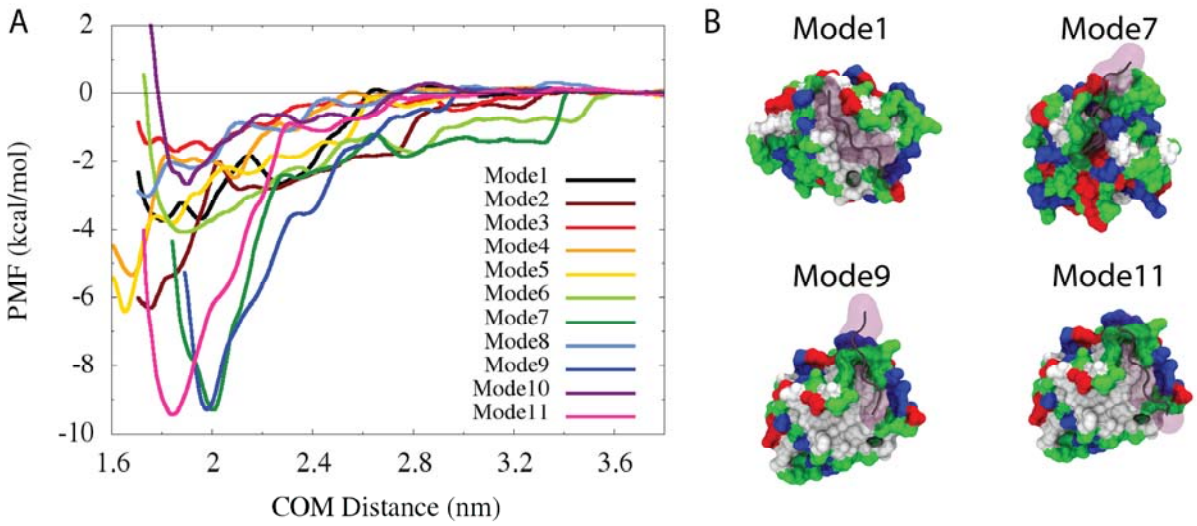

Figure 2. (A) PMF curves from umbrella sampling. (B) Representative binding modes where PFN1 is colored by molecular surface and residue types (hydrophobic: white; hydrophilic: green; positive: blue; and negative: red). P10 is highlighted with purple shade. The top scored mode1 mainly utilizes a hydrophobic interface. Meanwhile, the binding modes with the lowest binding free energies (mode7, mode9 and mode11) utilize a mixed hydrophobic/hydrophilic interface.

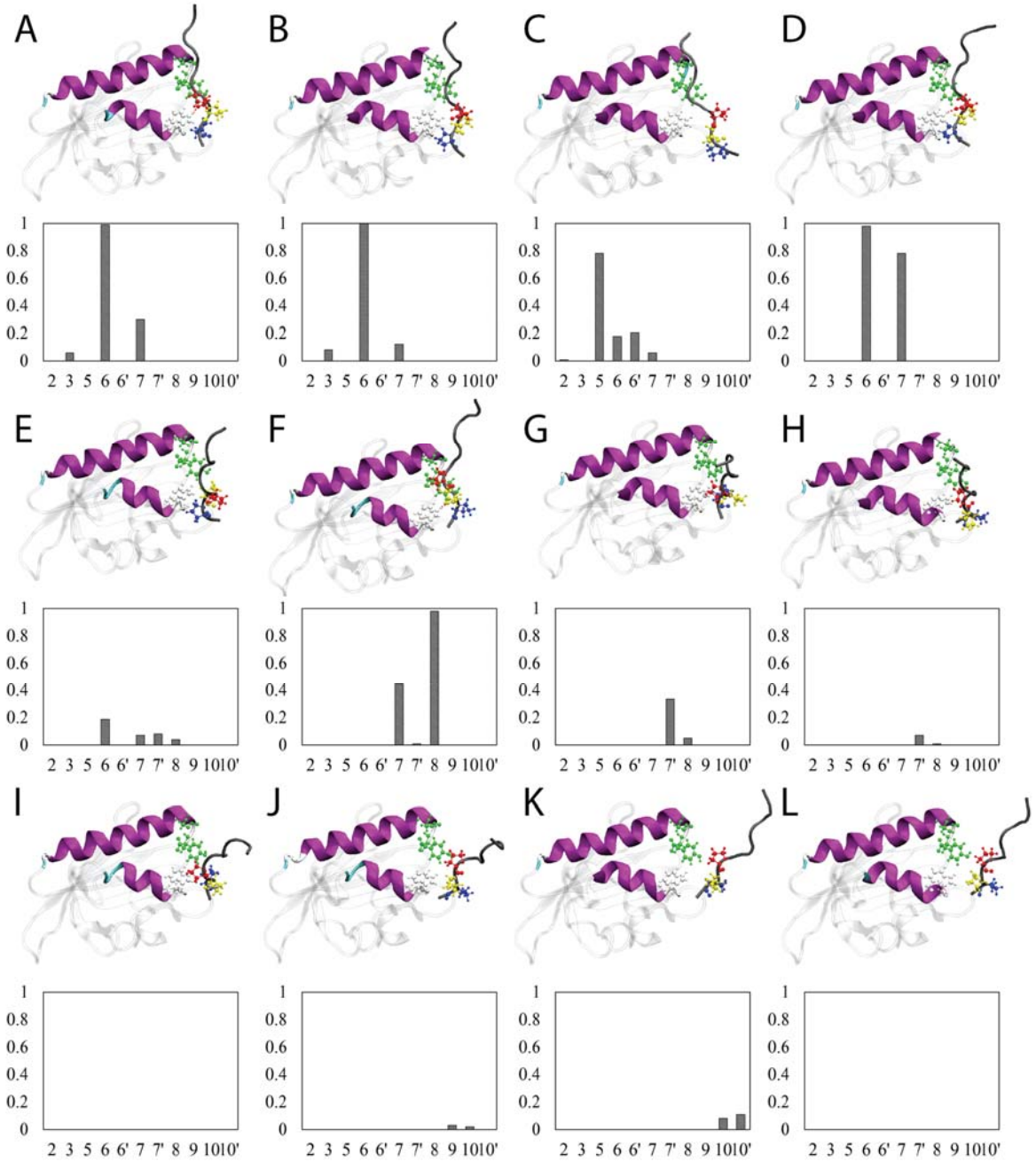

Figure 3. Illustrations of PFN1-P10 binding mechanism (mode9). Representative structures from window 1.6 nm to 2.7 nm are shown from (A) to (L). Only N-terminal α helix and C-terminal α helix of PFN1 are shown in visible secondary structure representations. P10 is shown as a black string representation. Typical residues involved in the interfacial hydrogen bonds are shown in bead and stick models (PFN1-W3 (white), PFN1-S137 (green), PFN1-Y139 (green), P10-P7 (red), P10-P8 (yellow) and P10-P9 (blue)). The average occurrence frequency of hydrogen bonds

are plotted under each of the structures. The x axis lists all dominant hydrogen bond pairs: 2 stands for S137-P2 (the order is PFN1-P10, omitted thereafter); 3 stands for S137-P3; 5 stands for Y139-P5; 6 stands for Y139-P6; 6' stands for W3-P6; 7 stands for W3-P7; 7' stands for Y139-P7; 8 stands for W3-P8; 9 stands for G2-P9; 10 stands for Y139-P10; 10' stands for S137-P10.

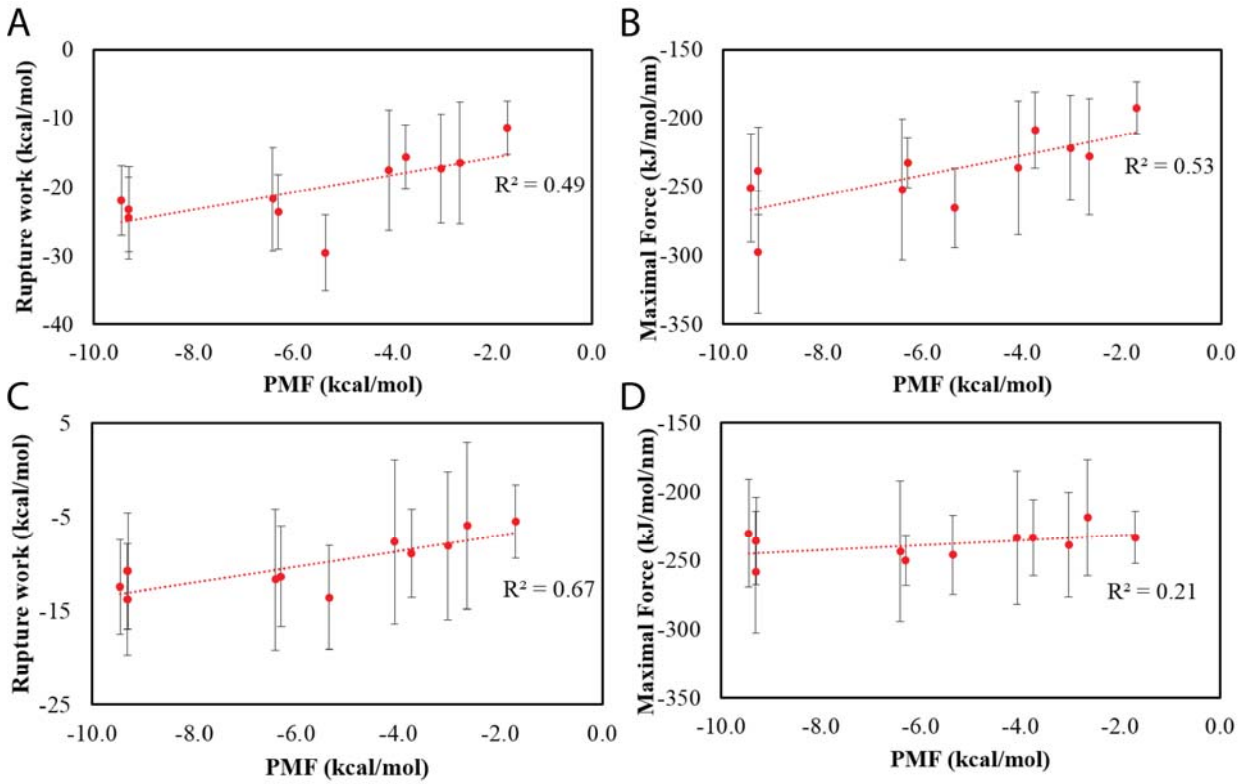

Figure 4. Correlation between SMD results (rupture work, maximal force) and PMF differences. (A) Rupture work calculated from 1 nm/ns SMD. (B) Maximal force recorded from 1 nm/ns SMD. (C) Rupture work calculated from 0.1 nm/ns SMD. (D) Maximal force recorded from 0.1 nm/ns SMD.