

# Supplementary Information

## Two Structural Scenarios for Protein Stabilization by PEG

Shu-Han Chao,<sup>†</sup> Sam S. Matthews,<sup>§</sup> Ryan Paxman,<sup>§</sup> Aleksei Aksimentiev,<sup>†,^,\*</sup> Martin Gruebele,<sup>†,¶,^,\*</sup> and Joshua B. Price<sup>§,\*</sup>

<sup>†</sup>Department of Physics and Center for the Physics of Living Cells, University of Illinois, Urbana, Illinois 61801, United States

<sup>§</sup>Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT 84602, United States

<sup>¶</sup>Department of Chemistry, University of Illinois, Urbana, Illinois 61801, United States

<sup>^</sup>Center for Biophysics and Quantitative Biology, University of Illinois, Urbana, Illinois 61801, United States

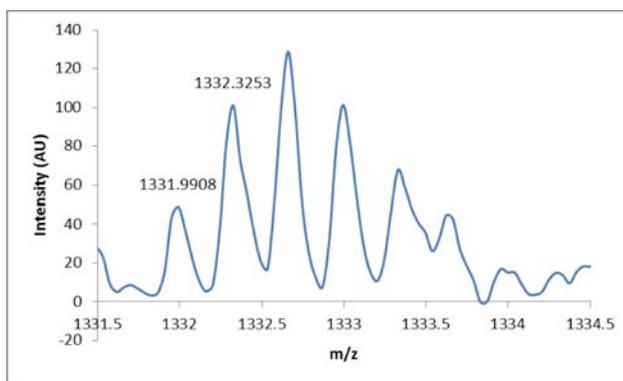
### 1. Protein Synthesis, Purification, and Characterization (detail)

PEGylated protein **S16/N19p/Y23** and non-PEGylated **S16/N19/Y23** were synthesized, purified, and characterized previously. PEGylated proteins **A16/N19p/Y23**, **S16/N19p/F23**, and **A16/N19p/F23**, and their non-PEGylated counterparts **A16/N19/Y23**, **S16/N19/F23**, and **A16/N19/F23** were synthesized as C-terminal acids, by microwave-assisted solid-phase peptide synthesis, using a standard Fmoc N $\alpha$  protection strategy as described previously.<sup>1-2</sup> Briefly, 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU, purchased from Advanced ChemTech) and N-hydroxybenzotriazole hydrate (HOBT, purchased from Advanced ChemTech) were used to activate Fmoc-protected amino acids in preparation for coupling reactions. Deprotection reactions were performed with 20% piperidine in N,N-dimethylformamide. Fmoc-Gly-loaded Novasyn WANG resin and all Fmoc-protected  $\alpha$ -amino acids with acid-labile side-chain protecting groups were purchased from EMD Biosciences, except for Fmoc-Asn(PEG4)-OH (*N*<sup>2</sup>-fluorenylmethoxycarbonyl-*N*<sup>4</sup>-[11-methoxy-3,6,9-trioxaundecyl]-L-asparagine), which was synthesized as described previously.<sup>1</sup> Proteins were globally deprotected and cleaved from the resin by stirring the resin for 4h in a solution of trifluoroacetic acid (2 mL) containing phenol (0.125 g), water (125  $\mu$ L), thioanisole (125  $\mu$ L), ethanedithiol 62.5  $\mu$ L) and triisopropylsilane (25  $\mu$ L). Proteins were precipitated from the concentrated TFA solution by addition of diethyl ether

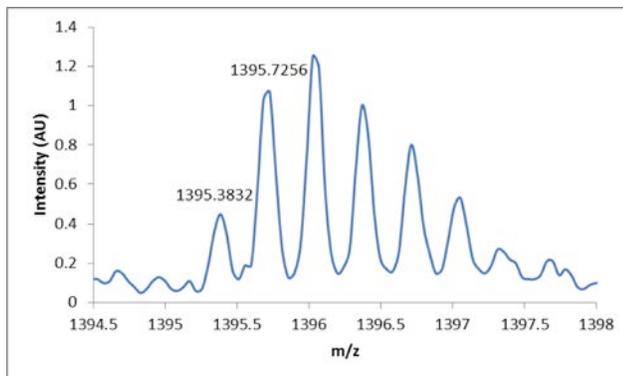
(~40 ml), dissolved in 1:1 H<sub>2</sub>O/MeCN, frozen, and lyophilized to remove volatile impurities. The resulting powder was stored at -20°C until purification.

Proteins were purified by preparative reverse-phase high performance liquid chromatography (HPLC) on a C18 column using a linear gradient of water in acetonitrile with 0.1% v/v TFA. HPLC fractions containing the desired protein product were pooled, frozen, and lyophilized. Proteins were identified by electrospray ionization time of flight mass spectrometry (ESI-TOF, Figures S1-S6), and purity was assessed by analytical HPLC (Figures S7-S12).

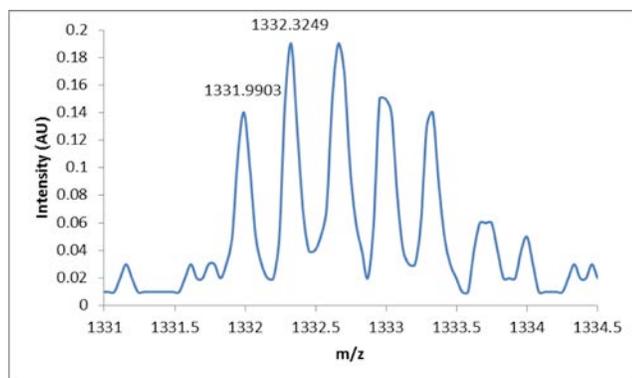
## 2. ESI-TOF Data



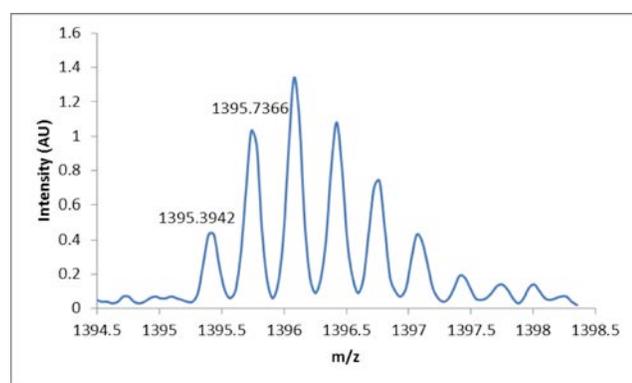
**Figure S1:** ESI-TOF spectrum for protein **A16/N19/Y23**. Expected  $[M+3H]^{3+}/3 = 1332.0071$  Da. Observed  $[M+3H]^{3+}/3 = 1331.9908$



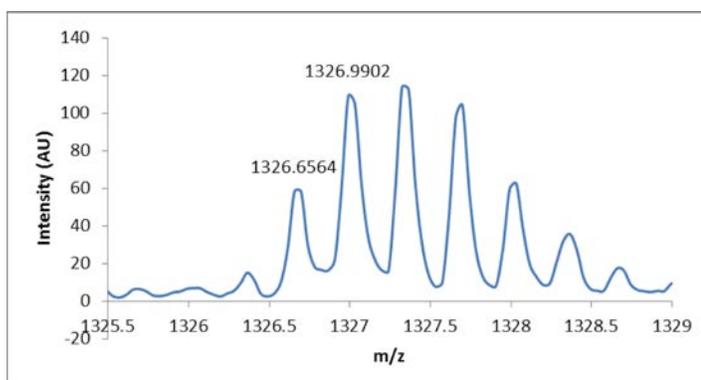
**Figure S2:** ESI-TOF spectrum for PEGylated protein **A16/N19p/Y23**. Expected  $[M+3H]^{3+}/3 = 1395.3806$  Da. Observed  $[M+3H]^{3+}/3 = 1395.3832$



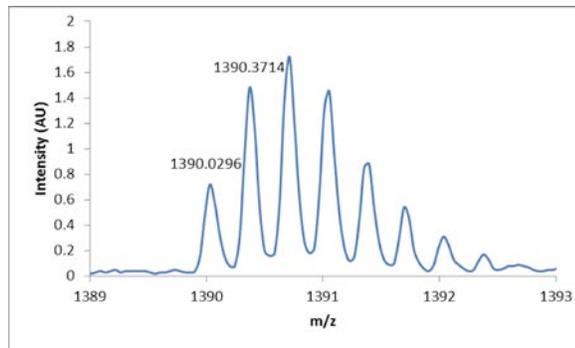
**Figure S3:** ESI-TOF spectrum for protein **S16/N19/F23**. Expected  $[M+3H]^{3+}/3 = 1332.0071$  Da. Observed  $[M+3H]^{3+}/3 = 1331.9903$



**Figure S4:** ESI-TOF spectrum for PEGylated protein **S16/N19p/F23**. Expected  $[M+3H]^{3+}/3 = 1395.3806$  Da. Observed  $[M+3H]^{3+}/3 = 1395.3942$

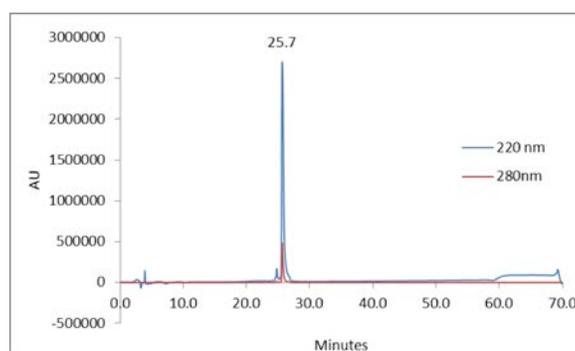


**Figure S5:** ESI-TOF spectrum for the Pin WW domain protein **A16/N19/F23**. Expected  $[M+3H]^{3+}/3 = 1326.6755$  Da. Observed  $[M+3H]^{3+}/3 = 1326.6564$

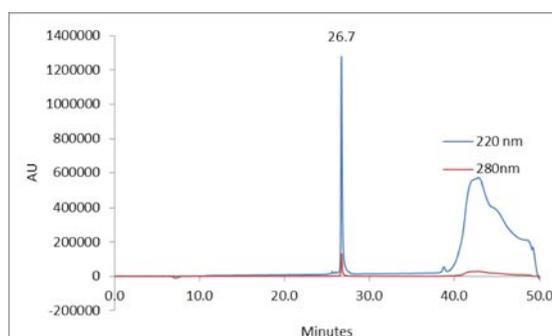


**Figure S6:** ESI-TOF spectrum for the Pin WW domain protein **A16/N19p/F23**. Expected  $[M+3H]^{3+}/3 = 1390.0490$  Da. Observed  $[M+3H]^{3+}/3 = 1390.0296$

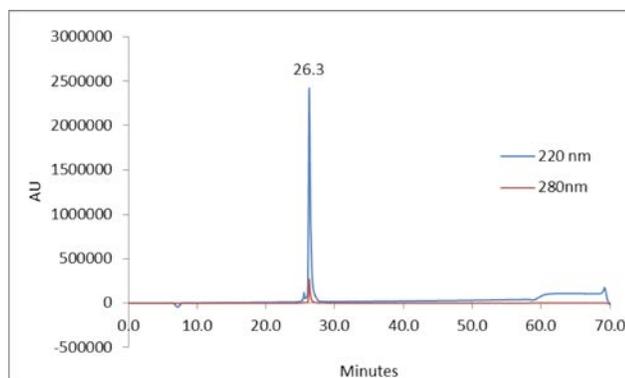
### 3. HPLC Data



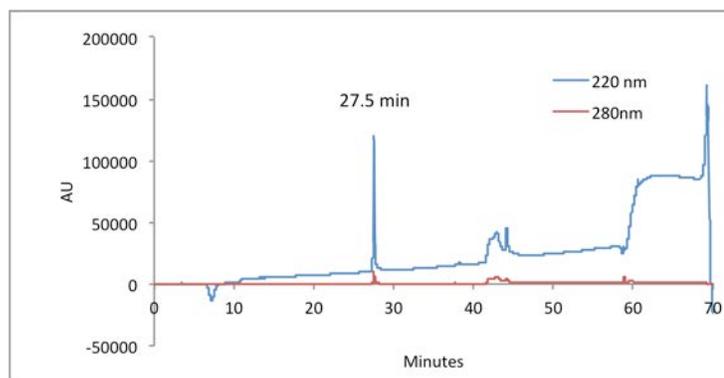
**Figure S7:** Analytical HPLC Data for Pin WW domain protein **A16/N19/Y23**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H<sub>2</sub>O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.



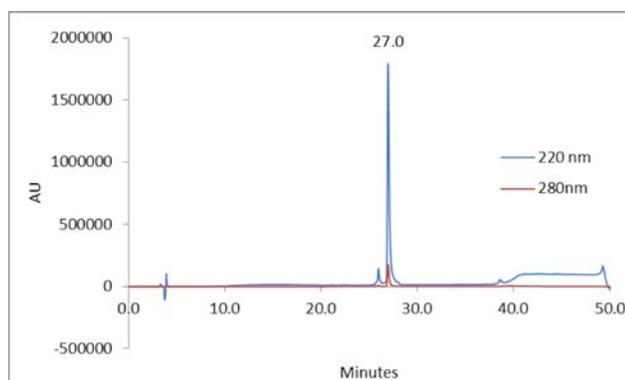
**Figure S8:** Analytical HPLC Data for Pin WW domain protein **A16/N19p/Y23**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H<sub>2</sub>O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B, flow rate 1 ml/min).



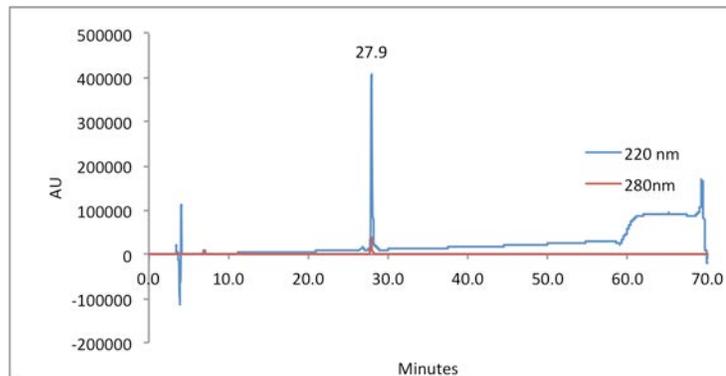
**Figure S9:** Analytical HPLC Data for Pin WW domain protein **S16/N19/F23**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H<sub>2</sub>O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.



**Figure S10:** Analytical HPLC Data for Pin WW domain protein **S16/N19p/F23**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H<sub>2</sub>O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.



**Figure S11:** Analytical HPLC Data for Pin WW domain protein **A16/N19/F23**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H<sub>2</sub>O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.



**Figure S12:** Analytical HPLC Data for Pin WW domain protein **A16/N19p/F23**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H<sub>2</sub>O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

#### 4. Variable Temperature Circular Dichroism Spectropolarimetry

Variable temperature CD data for **S16/N19/Y23** and **S16/N19p/Y23** are from previously performed experiments. Measurements were made with an Aviv 420 Circular Dichroism Spectropolarimeter, using quartz cuvettes with a path length of 0.1 cm. Variable temperature CD data were obtained at least in triplicate for 100  $\mu$ M solutions of **A16/N19/Y23**, **A16/N19p/Y23**, **S16/N19/F23**, **S16/N19p/F23**, **A16/N19/F23**, and **A16/N19p/F23** by monitoring molar ellipticity at 227 nm from 1 to 95°C at 2 °C intervals, with 120 s equilibration time between data points and 30 s averaging times. Data from the three (or more) replicate variable temperature CD experiments on each protein were fit globally to the following model for two-state thermally induced unfolding transitions:

$$[\theta] = \frac{(D_0 + D_1 \cdot T) + K_f(N_0 + N_1 \cdot T)}{1 + K_f}, \quad (S1)$$

where T is temperature in Kelvin, D<sub>0</sub> is the y-intercept and D<sub>1</sub> is the slope of the post-transition baseline; N<sub>0</sub> is the y-intercept and N<sub>1</sub> is the slope of the pre-transition baseline; and K<sub>f</sub> is the temperature-dependent folding equilibrium constant. K<sub>f</sub> is related to the temperature-dependent free energy of folding  $\Delta G_f(T)$  according to the following equation:

$$K_f = \exp\left[\frac{-\Delta G_f(T)}{RT}\right], \quad (S2)$$

where R is the universal gas constant (0.0019872 kcal/mol/K).  $\Delta G_f(T)$  was fit to the following equation:

$$\Delta G_f = \frac{\Delta H(T_m) \cdot (T_m - T)}{T_m} + \Delta C_p \cdot (T - T_m - T \cdot \ln\left[\frac{T}{T_m}\right]), \quad (S3)$$

where the fit parameters are  $T_m$  (the midpoint of the unfolding transition; the temperature at which  $\Delta G_f = 0$ ),  $\Delta H(T_m)$  (the change in enthalpy upon folding at  $T_m$ , and  $\Delta C_p$  (the change in heat capacity upon folding). The parameters for equations S1–S3 were used to calculate the values of the folding free energy  $\Delta G_f$  for proteins in Table \_ of the main text. Variable temperature CD data for proteins **A16/N19/Y23**, **A16/N19p/Y23**, **S16/N19/F23**, **S16/N19p/F23**, **A16/N19/F23**, and **A16/N19p/F23** appear in Figure S13–S15.

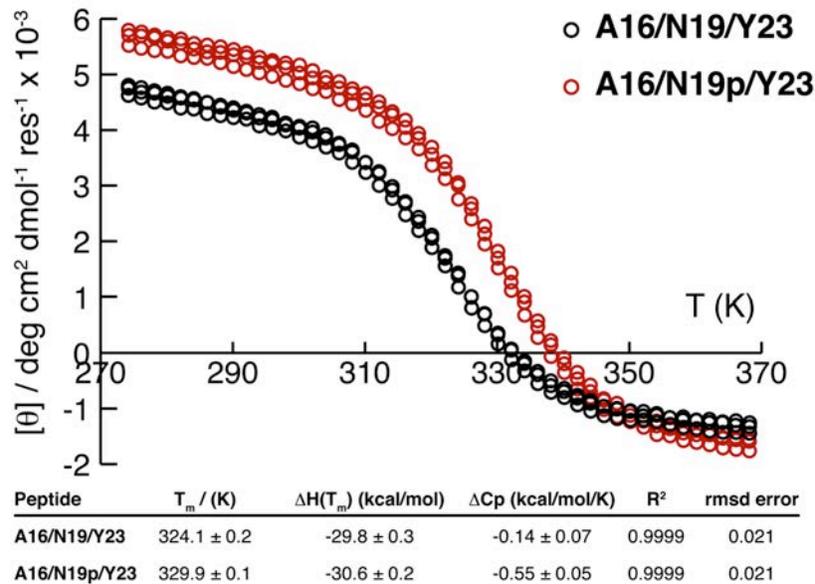
The enthalpy of folding  $\Delta H_f$  for a WW variant can be obtained at a given temperature  $T$  from the following equation, using the values of  $T_m$ ,  $\Delta H(T_m)$ , and  $\Delta C_p$  obtained from global fitting of variable temperature CD data:

$$\Delta H_f = \Delta H(T_m) + \Delta C_p(T - T_m) \quad (S4)$$

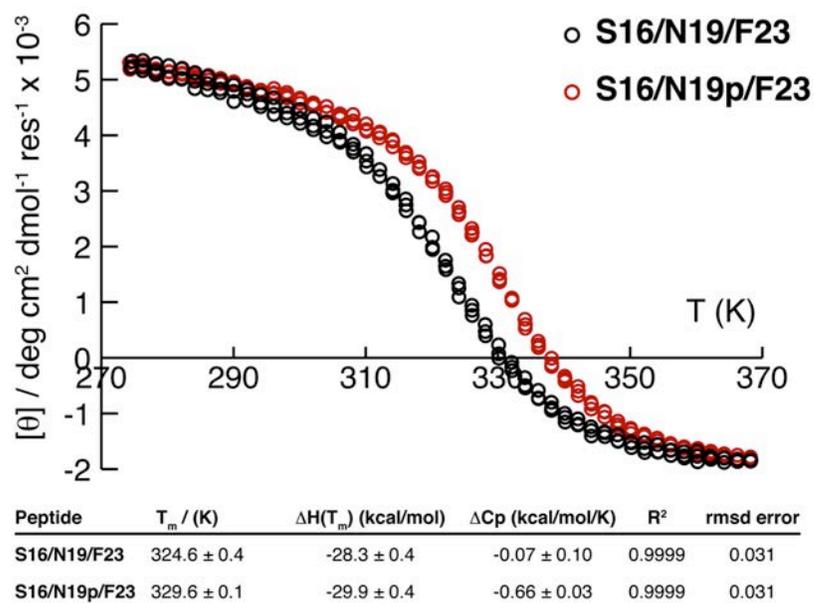
The entropy of folding can be also calculated from these parameters, using the following equation:

$$\Delta S = \frac{\Delta H(T_m)}{T_m} + \Delta C_p \ln \left[ \frac{T}{T_m} \right] \quad (S5)$$

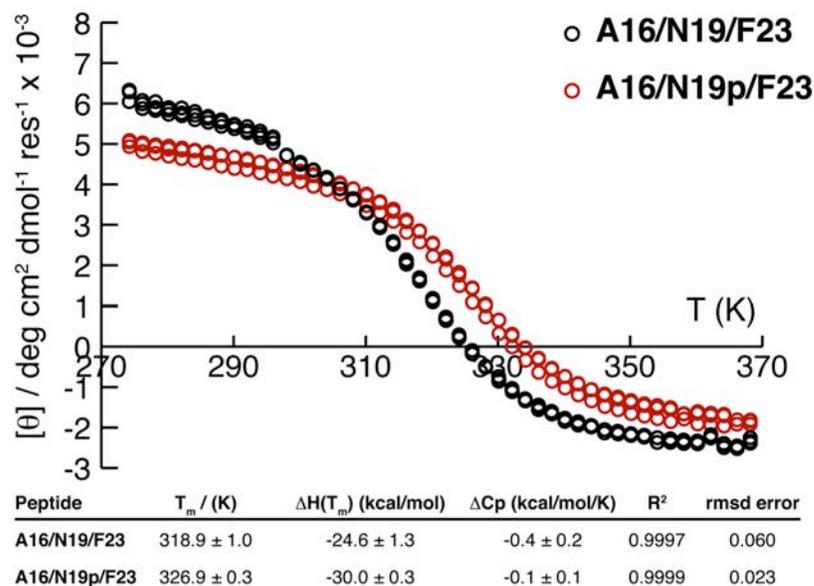
In this way, the  $\Delta G_f$  and  $\Delta \Delta G_f$  information from our variable temperature CD experiments can be parsed into entropic and enthalpic components.



**Figure S13:** Variable temperature CD Data for PEGylated protein **A16/N19p/Y23** and its non-PEGylated counterpart **A16/N19/Y23** at 100  $\mu$ M in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.

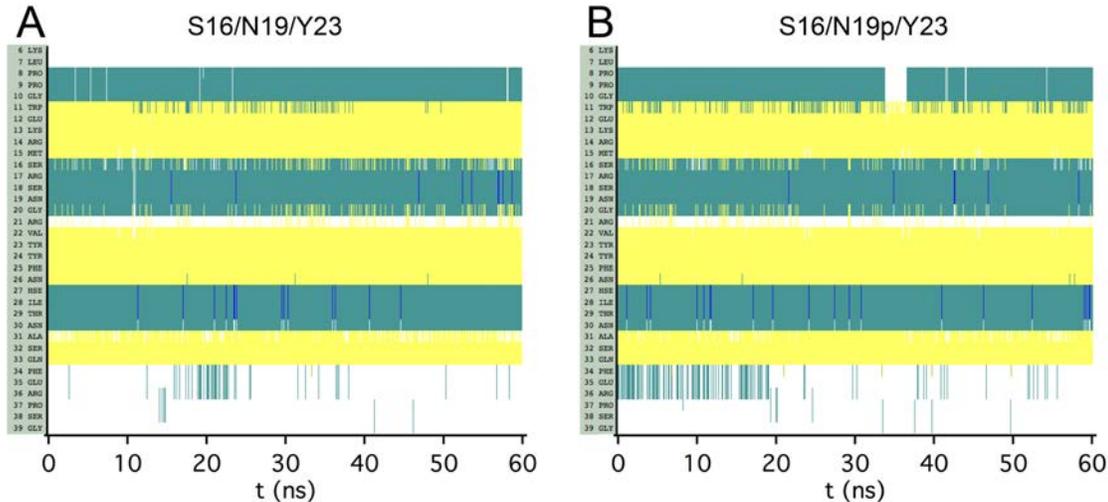


**Figure S14:** Variable temperature CD Data for PEGylated protein **S16/N19p/F23** and its non-PEGylated counterpart **S16/N19/F23** at 100  $\mu$ M in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.

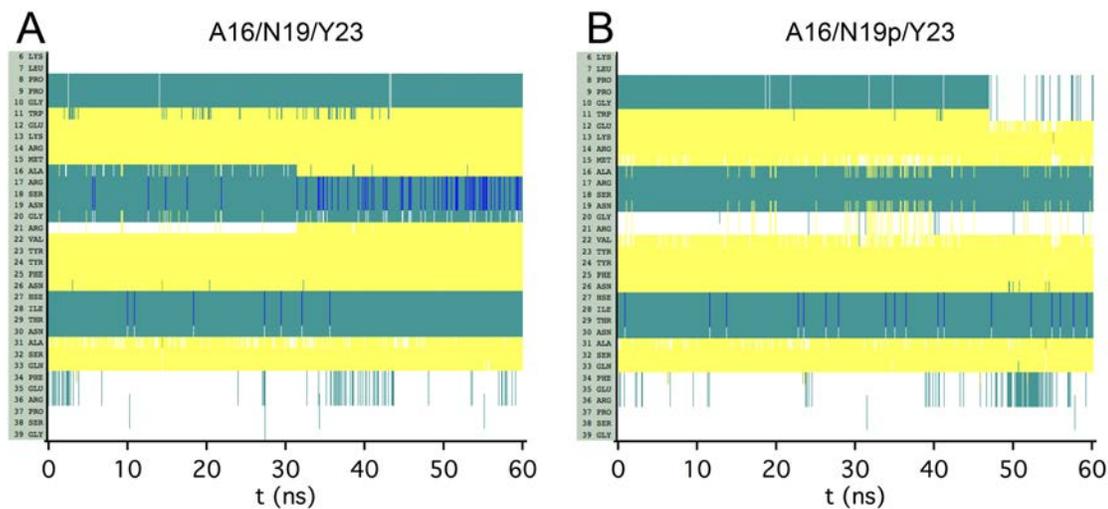


**Figure S15:** Variable temperature CD Data for PEGylated protein **A16/N19p/F23** and its non-PEGylated counterpart **A16/N19/F23** at 100  $\mu$ M in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.

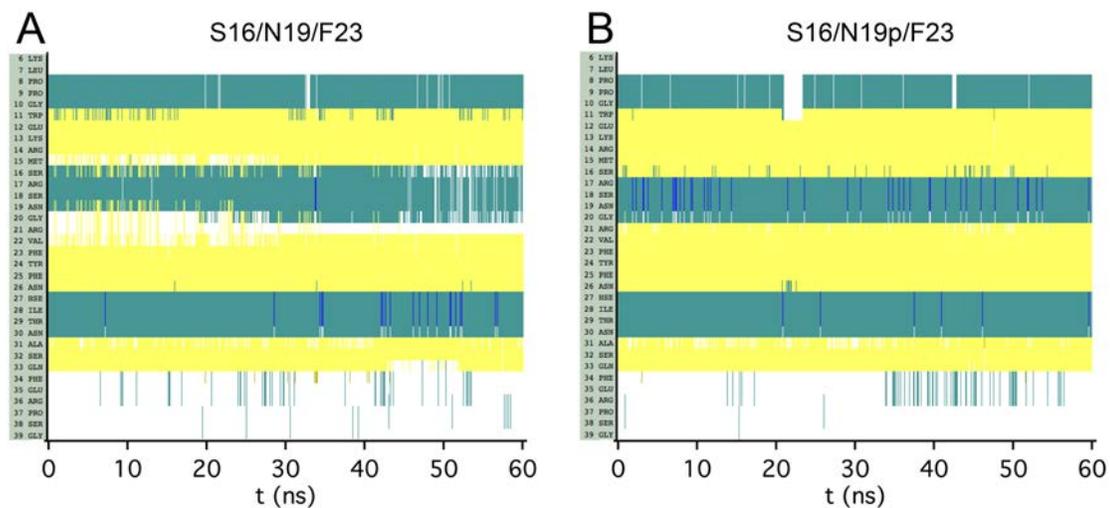
## 5. Secondary structure of all variants



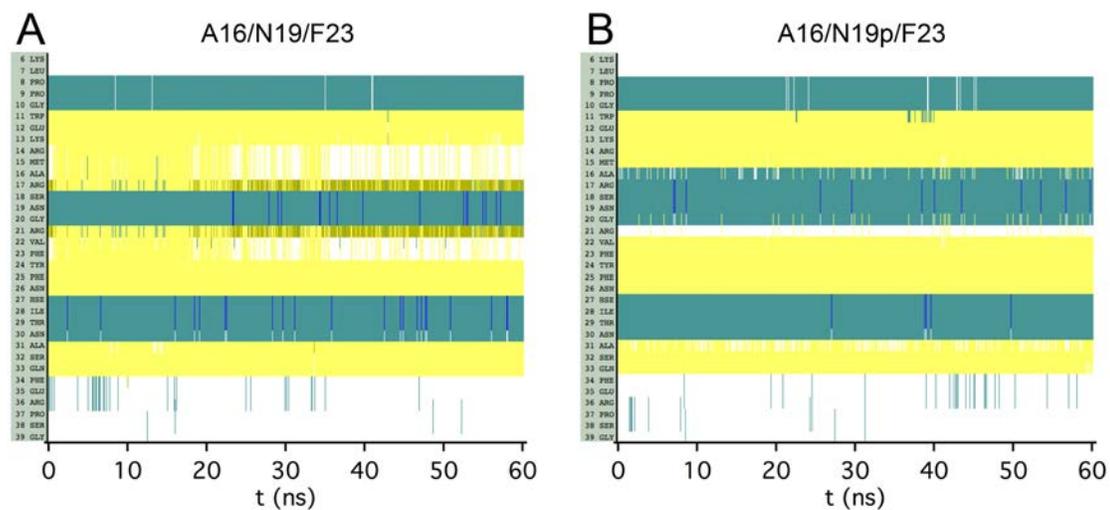
**Figure S16:** Secondary structure for PEGylated protein **S16/N19p/Y23** and its non-PEGylated counterpart **S16/N19/Y23** throughout the entire 60 ns trajectory. Color code: turn – cyan; coil – white;  $\beta$ -sheet – yellow; 3-10 helix – blue.



**Figure S17:** Secondary structure for PEGylated protein **A16/N19p/Y23** and its non-PEGylated counterpart **A16/N19/Y23** throughout the entire 60 ns trajectory.

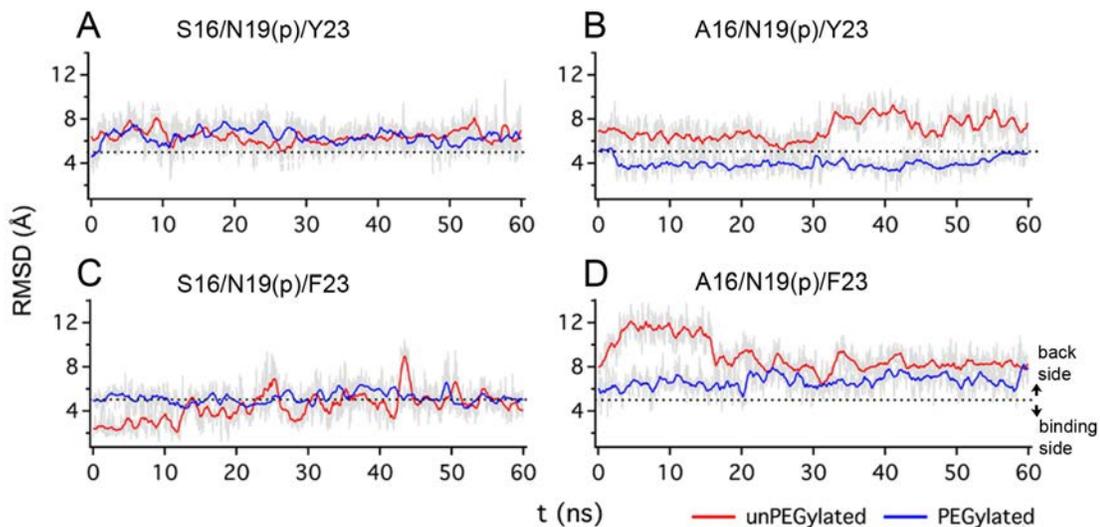


**Figure S18:** Secondary structure for PEGylated protein **S16/N19p/F23** and its non-PEGylated counterpart **S16/N19/F23** throughout the entire 60 ns trajectory.

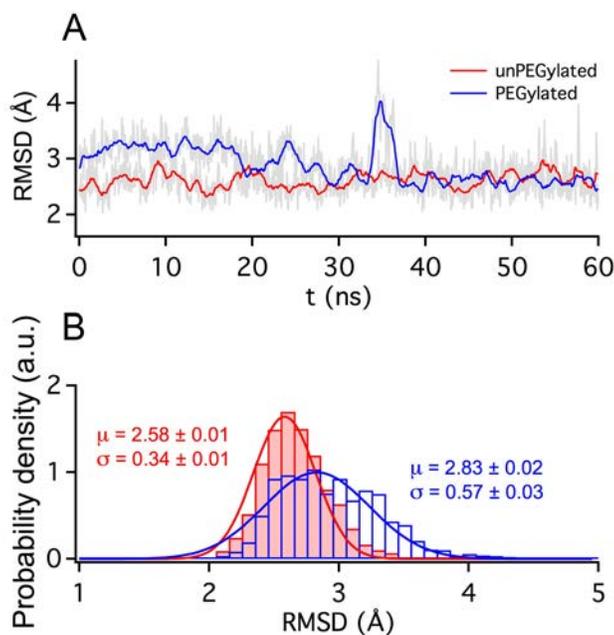


**Figure S19:** Secondary structure for PEGylated protein **A16/N19p/F23** and its non-PEGylated counterpart **A16/N19/F23** throughout the entire 60 ns trajectory.

## 6. RMSD

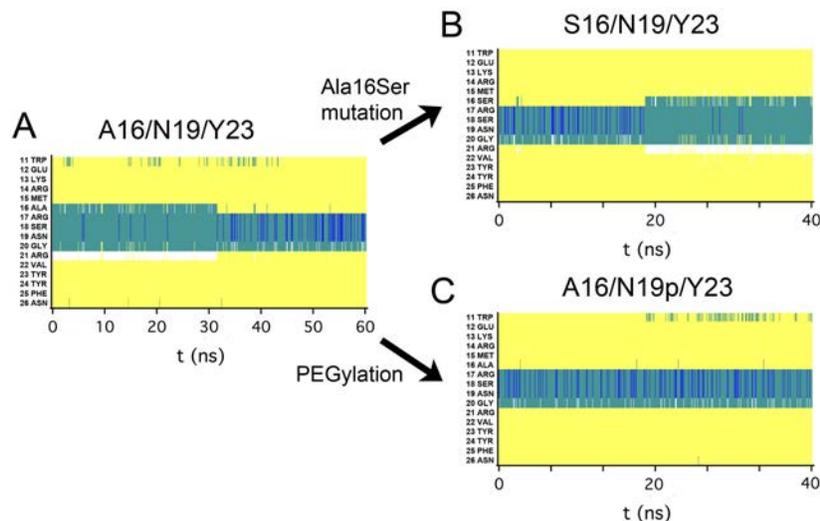


**Figure S20:** Asn19 RMSD of unPEGylated and PEGylated variants (A) **S16/N19/Y23**, (B) **A16/N19/Y23**, (C) **S16/N19/F23**, and (D) **A16/N19/F23F** are calculated every 50 ps (gray) and smoothed every 20 points for visualization (red and blue).

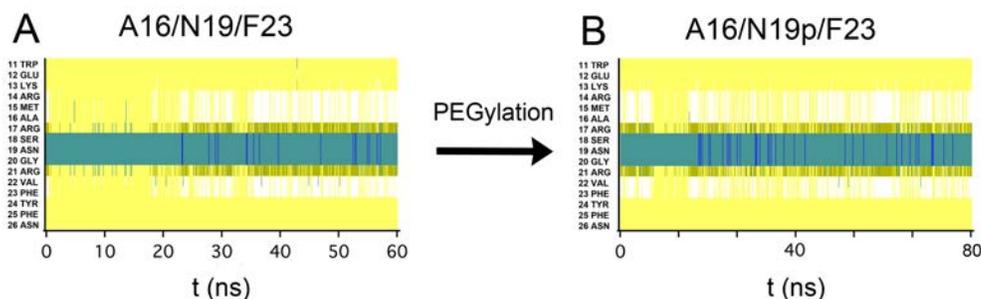


**Figure S21:** RMSD trajectories (A) and histograms (B) for entire pseudo wild type protein **S16/N19/Y23** before and after PEGylation. Data are calculated every 50 ps (gray) and smoothed every 20 points for visualization (red and blue). Histograms are fitted to Gaussian distribution with mean  $\mu$  and standard deviation  $\sigma$  shown on figure.

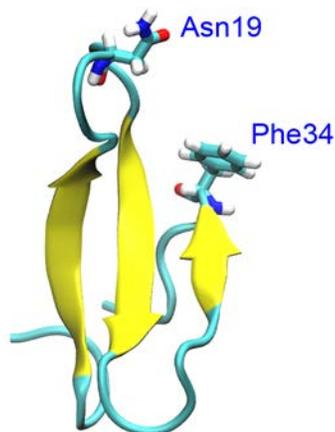
## 7. Secondary structure change after modification



**Figure S22:** Secondary structure change for **A16/N19/Y23**. (A) The occurrence of Asn19-Met15 interaction makes **A16/N19/Y23** has a structural transition at around 32 ns, which appears to be a more stable state for this mutant. In order to study the impact of mutation and PEGylation on the structure, we modified the variant to (B) **S16/N19/Y23**, (C) **A16/N19p/Y23** at the end of 60 ns simulation and resumed it. While Ala16Ser changes the structure back to its “native-like” state, the presence of PEG did not force the protein loop to switch back, but instead staying to what it was. This indicates although PEG can prevent structural transition due to Asn19-Met15 interaction (no transition in Figure S16B), the entropic force cannot win out the enthalpic force of Asn19-Met15 interaction once it happens. On the contrary, the mutation on position 16 has more impact to change it.



**Figure S23:** Secondary structure change for double mutant **A16/N19/F23**. We PEGylated the variant to (B) **A16/N19p/F23** at the end of 60 ns simulation and resumed it. The presence of PEG was unable to restore the secondary structure for entire 80 ns simulation.



**Figure S24:** Graphical representation of pseudo wild type hPin1 WW domain with two mutation sites highlighted.

## 8. Movies from MD simulations

**Movie jp502234s\_si\_002.avi:** 60 ns trajectory showing that PEG extends into the solvent. Water molecules and ions are hidden.

**Movie jp502234s\_si\_003.avi:** The interaction between Asn19 (red) and Met15 (orange) of A16/N19/Y23. Total trajectory length 60 ns. Water molecules and ions are hidden.

**Movie jp502234s\_si\_004.avi:** The interaction between PEG-4 (blue) and Arg21 (orange) of A16/N19p/Y23. Total trajectory length 60 ns. Water molecules and ions are hidden.

**Movie jp502234s\_si\_005.avi:** Repeated equilibration of PEG-45 attached at Asn19 between a solvated coil and a protein surface-coating state. Total trajectory length 50 ns. Water molecules and ions are hidden.

## References

- (1) Price, J. L.; Powers, E. T.; Kelly, J. W., N-PEGylation of a Reverse Turn is Stabilizing in Multiple Sequence Contexts unlike N-GlcNAcylation. *ACS Chem Biol* **2011**, *6*, 1188–1192.
- (2) Pandey, B. K.; Smith, M. S.; Torgerson, C.; Lawrence, P. B.; Matthews, S. S.; Watkins, E.; Groves, M. L.; Prigozhin, M. B.; Price, J. L., Impact of Site-Specific

PEGylation on the Conformational Stability and Folding Rate of the Pin WW Domain Depends Strongly on PEG Oligomer Length. *Bioconjugate Chem.* **2013**, *24*, 796-802.