

Context-Dependent Effects of Asparagine Glycosylation on Pin WW Folding Kinetics and Thermodynamics

*Joshua L. Price,¹ Dalit Shental-Bechor,² Apratim Dhar,³ Maurice J. Turner,¹ Evan T. Powers,¹
Martin Gruebele,³ Yaakov Levy*,² and Jeffery W. Kelly*¹*

¹Departments of Chemistry and Molecular and Experimental Medicine, The Scripps Research
Institute, La Jolla, CA 92037; ²Department of Structural Biology, Weizmann Institute of Science,
Rehovot, Israel, 76100; ³Center for Biophysics and Computational Biology and Departments of
Chemistry and Physics, University of Illinois, Urbana, Illinois 61801

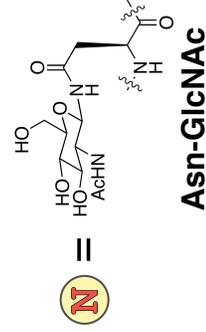
jkelly@scripps.edu; koby.levy@weizmann.ac.il

Supporting Information

Protein Synthesis	S3
General	S3
Purification and Characterization	S5
Removal of Acetate Protecting Groups on Asn-linked GlcNAc Residues in Glycosylated Pin WW Domain Proteins	S5
MALDI.....	S7
HPLC.....	S21
Circular Dichroism Spectroscopy	S34
Laser Temperature Jump Experiments	S41
Native Topology Model Calculations	S56
Potential of Mean Force.....	S58
Folding/Unfolding rate	S60
Ca Model	S60
References	S61

Table S1. Peptide Sequences and MALDI-TOF Data

Protein	Sequence	Expected [M+H] ⁺ (g/mol)	Observed [M+H] ⁺ (g/mol)
Pin WW	H ₂ N-KLPPGWEKRMSSGRVYFNFHI TNASQFERPSG-COOH	3983.0	3982.4
14	H ₂ N-KLPPGWEKRMSSGRVYFNFHI TNASQFERPSG-COOH	3940.9	3942
14g	H ₂ N-KLPPGWEK(N)MSRSSGRVYFNFHI TNASQFERPSG-COOH	4144.0	4146
17	H ₂ N-KLPPGWEKRMSSGRVYFNFHI TNASQFERPSG-COOH	3940.9	3939.1
17g	H ₂ N-KLPPGWEKRM(S)SSGRVYFNFHI TNASQFERPSG-COOH	4144.0	4145
18	H ₂ N-KLPPGWEKRMSSGRVYFNFHI TNASQFERPSG-COOH	4010.0	4011
18g	H ₂ N-KLPPGWEKRM(S)SSGRVYFNFHI TNASQFERPSG-COOH	4213.1	4213
19	H ₂ N-KLPPGWEKRMSSGRVYFNFHI TNASQFERPSG-COOH	4010.0	4010.2
19g	H ₂ N-KLPPGWEKRM(S)SSGRVYFNFHI TNASQFERPSG-COOH	4213.1	4213
20	H ₂ N-KLPPGWEKRMSSNRVYFNFHI TNASQFERPSG-COOH	4040.0	4039
20g	H ₂ N-KLPPGWEKRM(S)NRVYFNFHI TNASQFERPSG-COOH	4243.1	4243.9
23	H ₂ N-KLPPGWEKRMSSGRVYFNFHI TNASQFERPSG-COOH	3934.0	3932.4
23g	H ₂ N-KLPPGWEKRMSSGRVYFNFHI TNASQFERPSG-COOH	4137.0	4139.1
26g	H ₂ N-KLPPGWEKRMSSGRVYFNFHI TNASQFERPSG-COOH	4186.1	4188
30g	H ₂ N-KLPPGWEKRMSSGRVYFNFHI TNASQFERPSG-COOH	4186.1	4190
33	H ₂ N-KLPPGWEKRMSSGRVYFNFHI TNASQFERPSG-COOH	3969.0	3972
33g	H ₂ N-KLPPGWEKRMSSGRVYFNFHI TNASQFERPSG-COOH	4172.1	4172.1



Protein Synthesis

General

Proteins were synthesized as C-terminal acids, employing a solid phase peptide synthesis approach using a standard Fmoc N α protecting group strategy either manually (Proteins **14**, **14g**, **17**, **18**, **18g**, **19**, **20**, **23**, **23g**, **26g**, **30g**, **33**, and **33g**) or via a combination of manual and automated methods (Proteins **17g**, **19g**, and **20g** were synthesized on an Applied Biosystems 433A automated peptide synthesizer except for the manual coupling of Fmoc-Asn(Ac₃GlcNAc)-OH; see below). Amino acids were activated by 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). Fmoc-Gly-loaded Novasyn TGT resin and all Fmoc-protected α -amino acids with acid-labile side-chain protecting groups were purchased from EMD Biosciences, including the glycosylated amino acid Fmoc-Asn(Ac₃GlcNAc)-OH (N- α -Fmoc-N- β -[3,4,6-tri-O-acetyl-2-(acetylamino)-deoxy-2- β -glucopyranosyl]-L-asparagine).^{1,2} Piperidine and N,N-diisopropylethylamine (DIEA) were purchased from Aldrich, and N-methyl pyrrolidinone (NMP) was purchased from Applied Biosystems.

A general protocol for manual solid phase peptide synthesis follows: Fmoc-Gly-loaded NovaSyn TGT resin (217 mg, 50 μ mol at 0.23 mmol/g resin loading) was aliquotted into a fritted polypropylene syringe and allowed to swell in CH₂Cl₂ and dimethylformamide (DMF). Solvent was drained from the resin using a vacuum manifold. To remove the Fmoc protecting group on the resin-linked amino acid, 2.5 mL of 20% piperidine in DMF was added to the resin, and the resulting mixture was stirred at room temperature for 5 minutes. The deprotection solution was drained from the resin with a vacuum manifold. Then, an additional 2.5 mL of 20% piperidine in

DMF was added to the resin, and the resulting mixture was stirred at room temperature for 15 minutes. The deprotection solution was drained from the resin using a vacuum manifold, and the resin was rinsed five times with DMF.

For coupling of an activated amino acid to a newly deprotected amine on resin, the desired Fmoc-protected amino acid (250 μmol , 5 eq) and HBTU (250 μmol , 5 eq) were dissolved by vortexing in 2.5 mL 0.1 M HOBt (250 μmol , 5 eq) in NMP. To the dissolved amino acid solution was added 87.1 μmol DIEA (500 μmol , 10 eq). [Only 1.5 eq of amino acid were used during the coupling of the expensive Fmoc-Asn(Ac3GlcNAc)-OH monomer, and the required amounts of HBTU, HOBt, and DIEA were adjusted accordingly.] The resulting mixture was vortexed briefly and allowed to react for at least 1 min. The activated amino acid solution was then added to the resin, and the resulting mixture was stirred at room temperature for at least 1 h. Selected amino acids were double coupled as needed to allow the coupling reaction to proceed to completion. Following the coupling reaction, the activated amino acid solution was drained from the resin with a vacuum manifold, and the resin was subsequently rinsed five times with DMF. The cycles of deprotection and coupling were alternately repeated to give the desired full-length protein.

Acid-labile side-chain protecting groups were globally removed and proteins were cleaved from the resin by stirring the resin for ~ 4 h in a solution of phenol (0.5 g), water (500 μL), thioanisole (500 μL), ethanedithiol (250 μL), and triisopropylsilane (100 μL) in trifluoroacetic acid (TFA, 8 mL). Following the cleavage reaction, the TFA solution was drained from the resin, the resin was rinsed with additional TFA, and the resulting solution was concentrated under Ar. Proteins were precipitated from the concentrated TFA solution by addition of diethyl

ether (~45 mL). Following centrifugation, the ether was decanted, and the pellet was stored at -20 °C until purification.

Purification and Characterization

Immediately prior to purification, the crude protein was dissolved in either 1:1 water:acetonitrile, DMSO, or 8 M GdnHCl. Proteins were purified by preparative reverse-phase HPLC on a C18 column using a linear gradient of water in acetonitrile with 0.2% v/v TFA. HPLC fractions containing the desired protein product were pooled, frozen, and lyophilized. Proteins were identified by matrix-assisted laser desorption/ionization time-of-flight spectrometry (MALDI-TOF, Table S1; MALDI spectra appear below in Figure S1–S17), and purity was analyzed by analytical HPLC (Figures S17–S34).

Removal of Acetate Protecting Groups on Asn-linked GlcNAc Residues in Glycosylated Pin WW Domain Proteins

Acetate protecting groups were removed from the 3-, 4-, and 6-hydroxyl groups on the Asn-linked GlcNAc residues in proteins **14g**, **17g**, **18g**, **19g**, **20g**, **23g**, **26g**, **30g**, and **33g** by hydrazinolysis. Originally, this hydrazinolysis reaction was performed prior to cleavage from the resin by stirring the resin in 3 mL of a 1:6 v/v solution of hydrazine in methanol as described previously.³ We found that this method frequently resulted in the presence of an undesired side-product that closely co-eluted with the desired product during HPLC purification. MALDI-TOF data showed a difference of 14 mass units between peaks corresponding to the desired and undesired products. The undesired side-reaction of a single protected glutamate side-chain with hydrazine to form the corresponding hydrazide would be consistent with this mass difference. HPLC peaks corresponding to the putative hydrazide impurity could be removed by HPLC to

afford proteins **14g**, **18g**, **23g**, **26g**, **30g**, and **33g** in sufficient purity for biophysical characterization. However, proteins **17g**, **19g**, and **20g** were still contaminated with the putative hydrazide impurity even after multiple rounds of purification.

Consequently, **17g**, **19g**, and **20g** were re-synthesized, cleaved from resin, and purified without performing the hydrazinolysis reaction. Then, another previously described method was used to deprotect the Asn-linked GlcNAc hydroxyl groups in proteins **17g**, **19g**, and **20g**.⁴ Briefly, the purified lyophilized protein was dissolved in a solution of 5% hydrazine solution in 60 mM aqueous dithiothreitol and allowed to stand at room temperature for ~1 h with intermittent agitation. The deprotection reaction was quenched by the addition of a solution of tris(2-carboxyethyl)phosphine hydrochloride (10 mg/mL) and TFA (0.2% v/v) in water. The deprotected protein was then purified by HPLC and characterized by MALDI as described above.

After the glycosylated proteins described in this manuscript had been prepared, we found that HPLC purification prior to the hydrazinolysis reaction is unnecessary: the reaction appears to proceed to completion efficiently even for relatively insoluble crude proteins. After the reaction is quenched, solvent can be removed by lyophilization, and purification of the crude deprotected protein proceeds smoothly. We found that lyophilized crude deprotected glycoproteins are often relatively insoluble in water/acetonitrile solutions before HPLC purification. Such crude glycoproteins dissolve readily in aqueous 8 M guanidine hydrochloride, and can be injected onto preparative HPLC columns without difficulty. Interestingly, the resulting purified deprotected glycoproteins are generally quite soluble in water.

MALDI

MALDI spectra for proteins **PinWW**, **14–33g** are shown in Figures S1–S8.

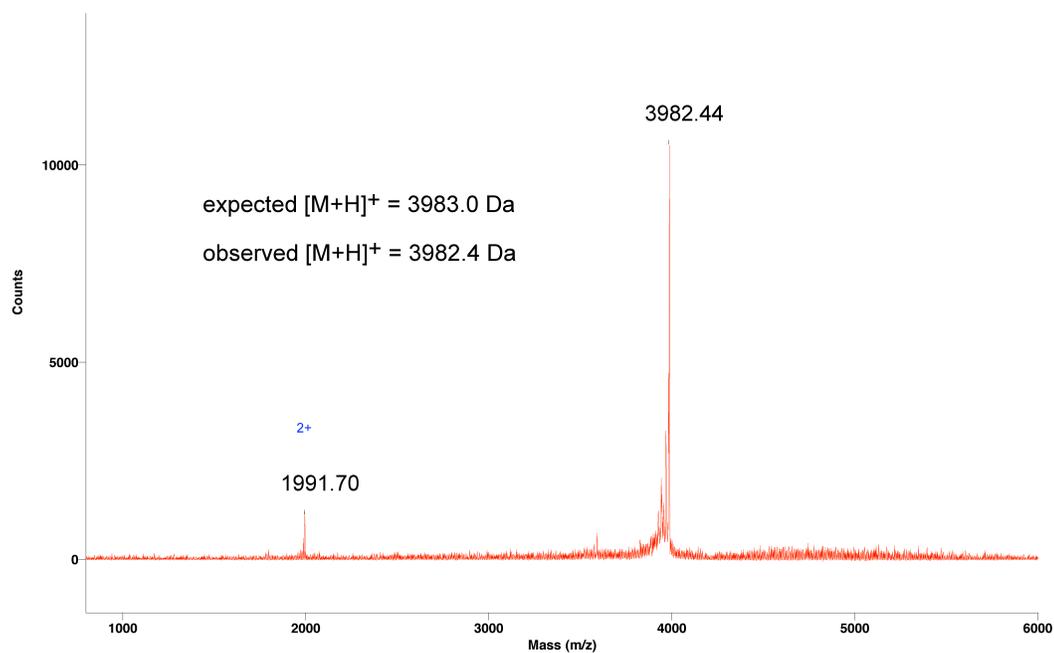


Figure S1. MALDI-TOF spectrum for the **Pin WW** domain protein. Expected [M+H]⁺ = 3983.0 Da. Observed [M+H]⁺ = 3982.4 Da.

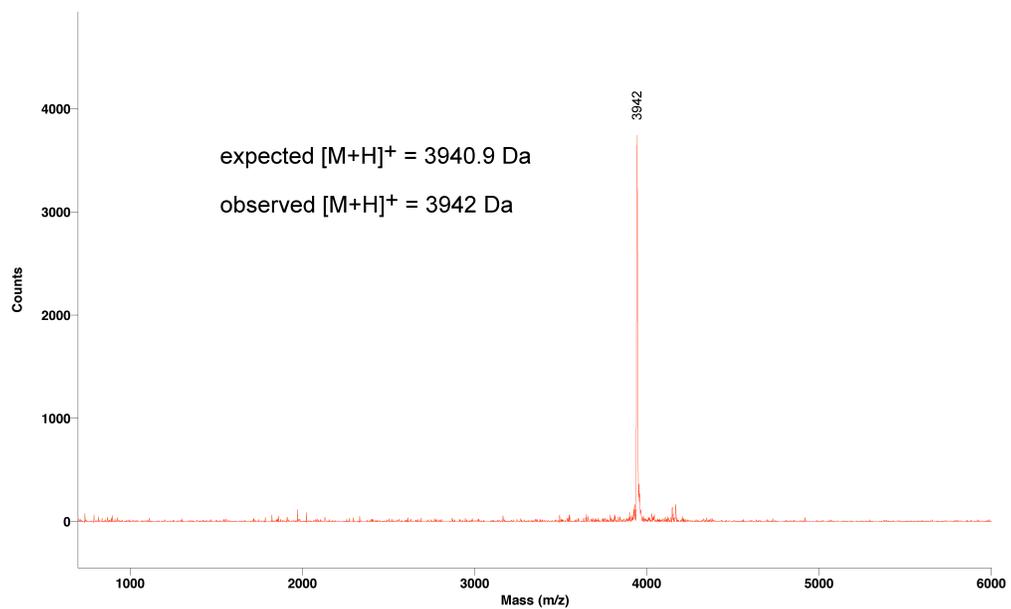


Figure S2. MALDI-TOF spectrum for Pin WW domain protein **14**. Expected $[M+H]^+ = 3940.9$ Da. Observed $[M+H]^+ = 3942$ Da.

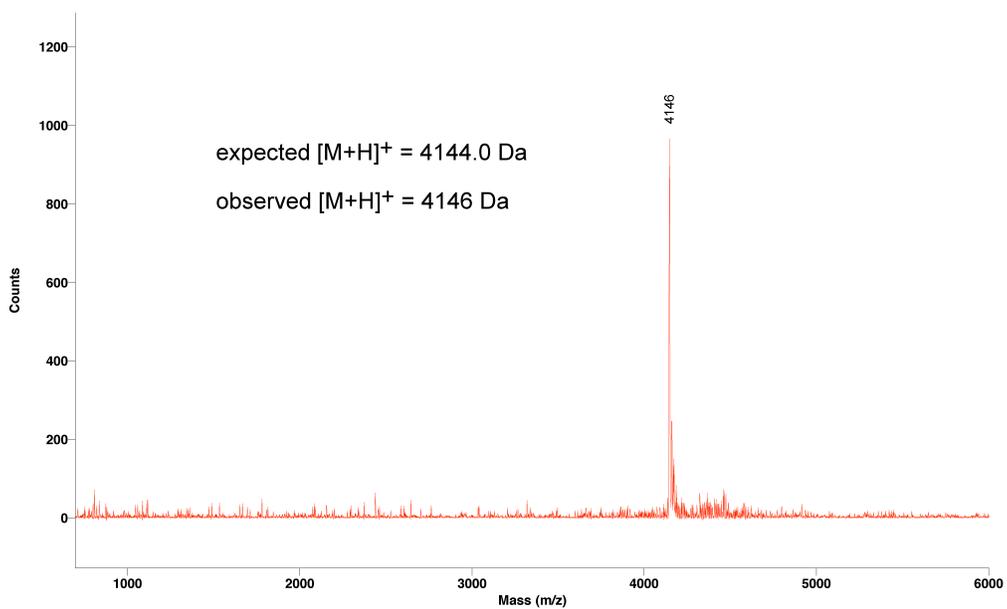


Figure S3. MALDI-TOF spectrum for Pin WW domain protein **14g**. Expected $[M+H]^+ = 4144.0$ Da. Observed $[M+H]^+ = 4146$ Da.

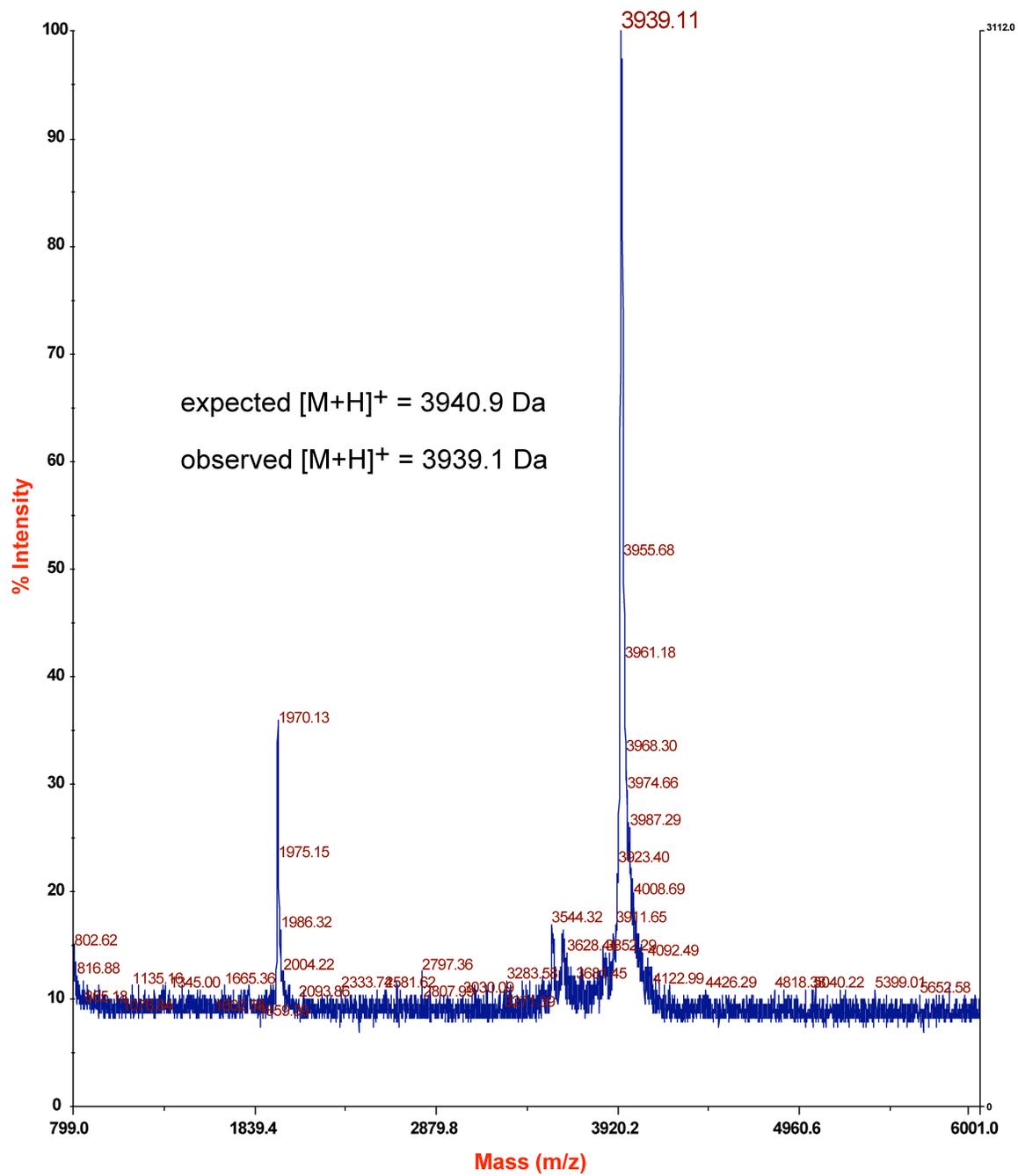


Figure S4. MALDI-TOF spectrum for Pin WW domain protein 17. Expected $[M+H]^+ = 3940.9$ Da. Observed $[M+H]^+ = 3939.1$ Da.

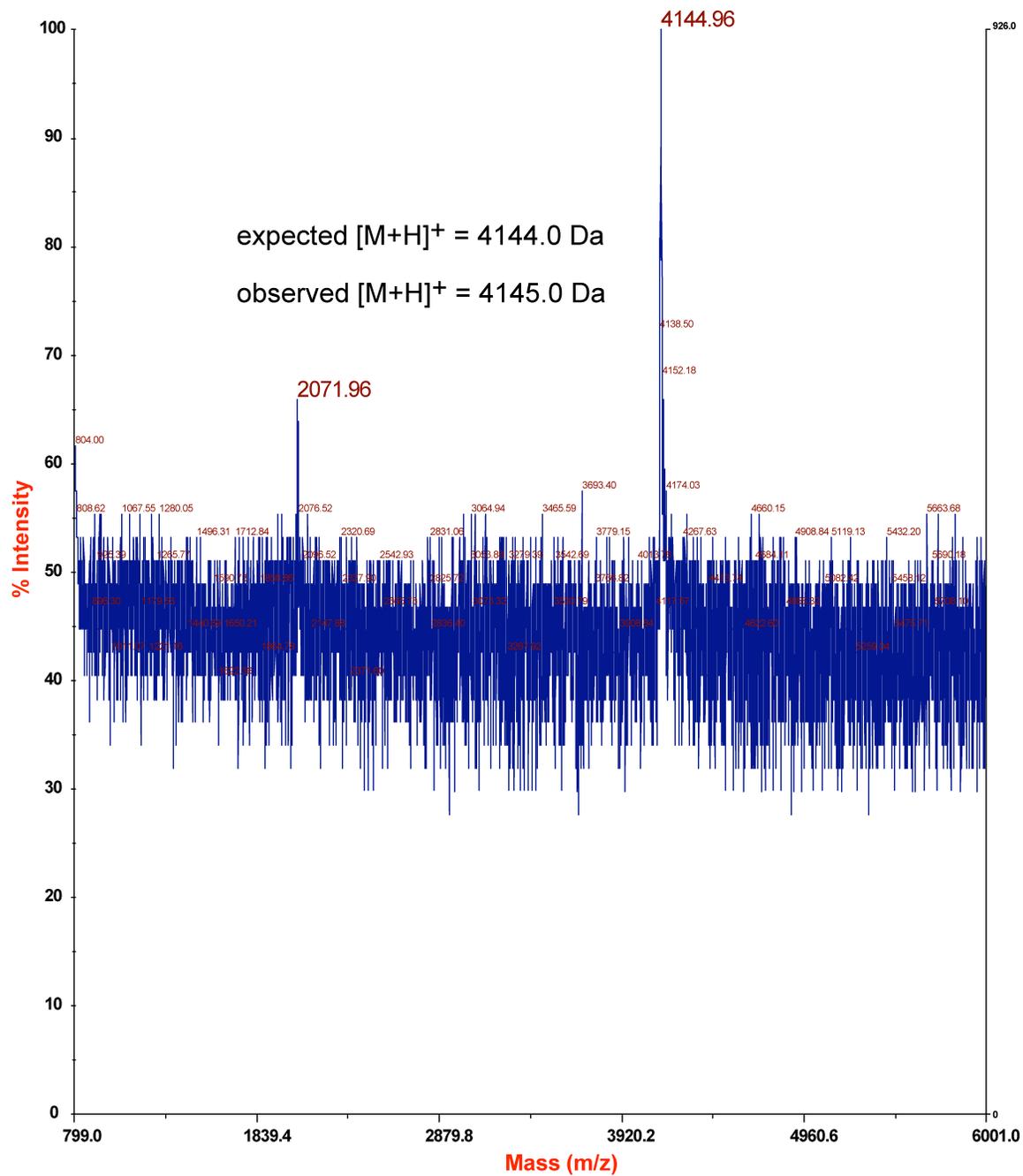


Figure S5. MALDI-TOF spectrum for Pin WW domain protein **17g**. Expected $[M+H]^+ = 4144.0$ Da. Observed $[M+H]^+ = 4145.0$ Da.

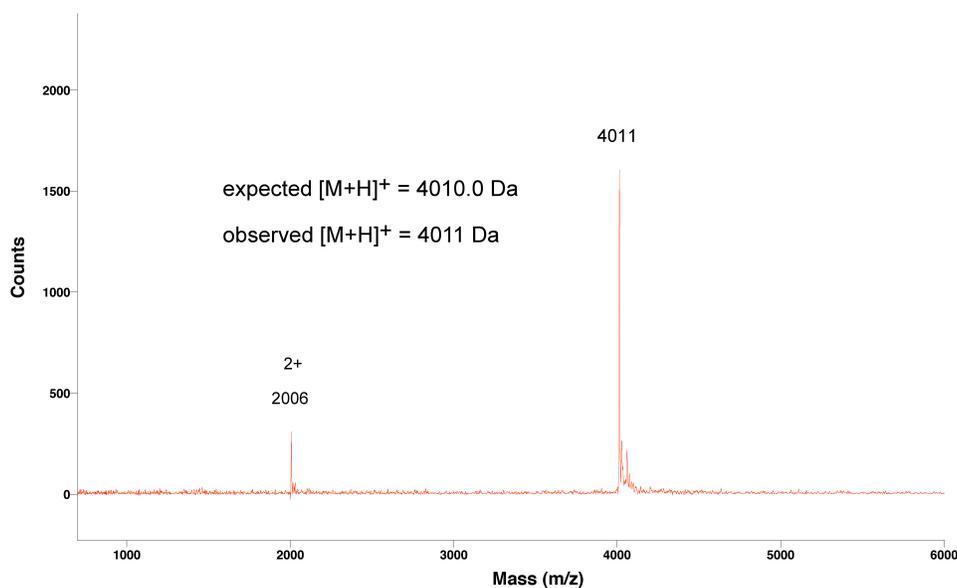


Figure S6. MALDI-TOF spectrum for Pin WW domain protein **18**. Expected $[M+H]^+ = 4010.0$ Da. Observed $[M+H]^+ = 4011$ Da.

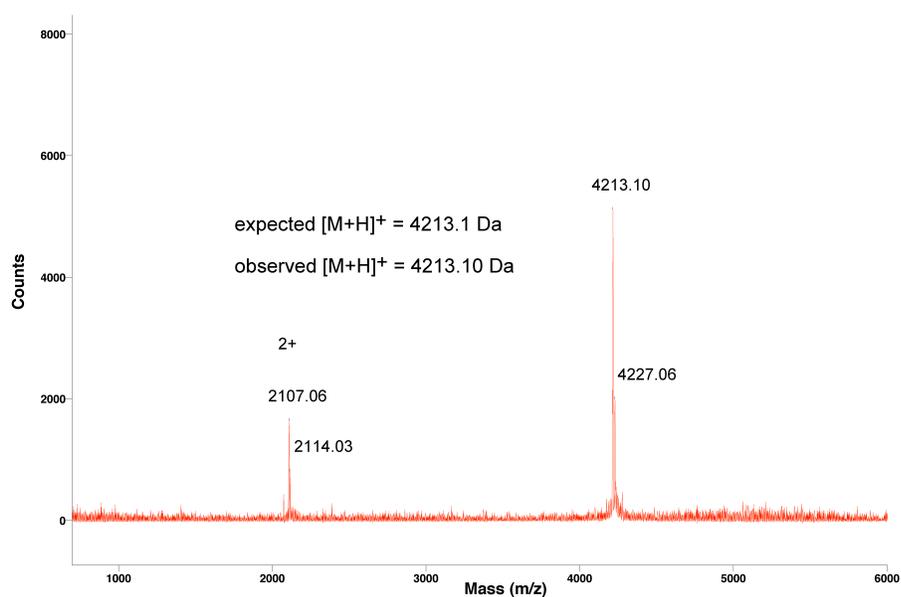


Figure S7. MALDI-TOF spectrum for Pin WW domain protein **18g**. Expected $[M+H]^+ = 4213.1$ Da. Observed $[M+H]^+ = 4213.1$ Da. The presence of a peak at 4227.06 Da represents a difference of 14 mass units from the $[M+H]^+$ peak, and most likely corresponds to the product of an undesired reaction between hydrazine and a protected glutamate side-chain (during the hydrazinolysis of the acetate protecting groups on the sugar) to form a hydrazide (the difference in mass between a hydrazide and a carboxylic acid is 14 mass units). The analytical HPLC trace (see Figure S24) for protein **18g** suggests that the amount of this putative undesired hydrazide is small. In any case, the mono-exponential nature of the temperature jump kinetic data for **18g** (see Figure S47) suggests that the desired product and the putative undesired hydrazide have indistinguishable folding properties.

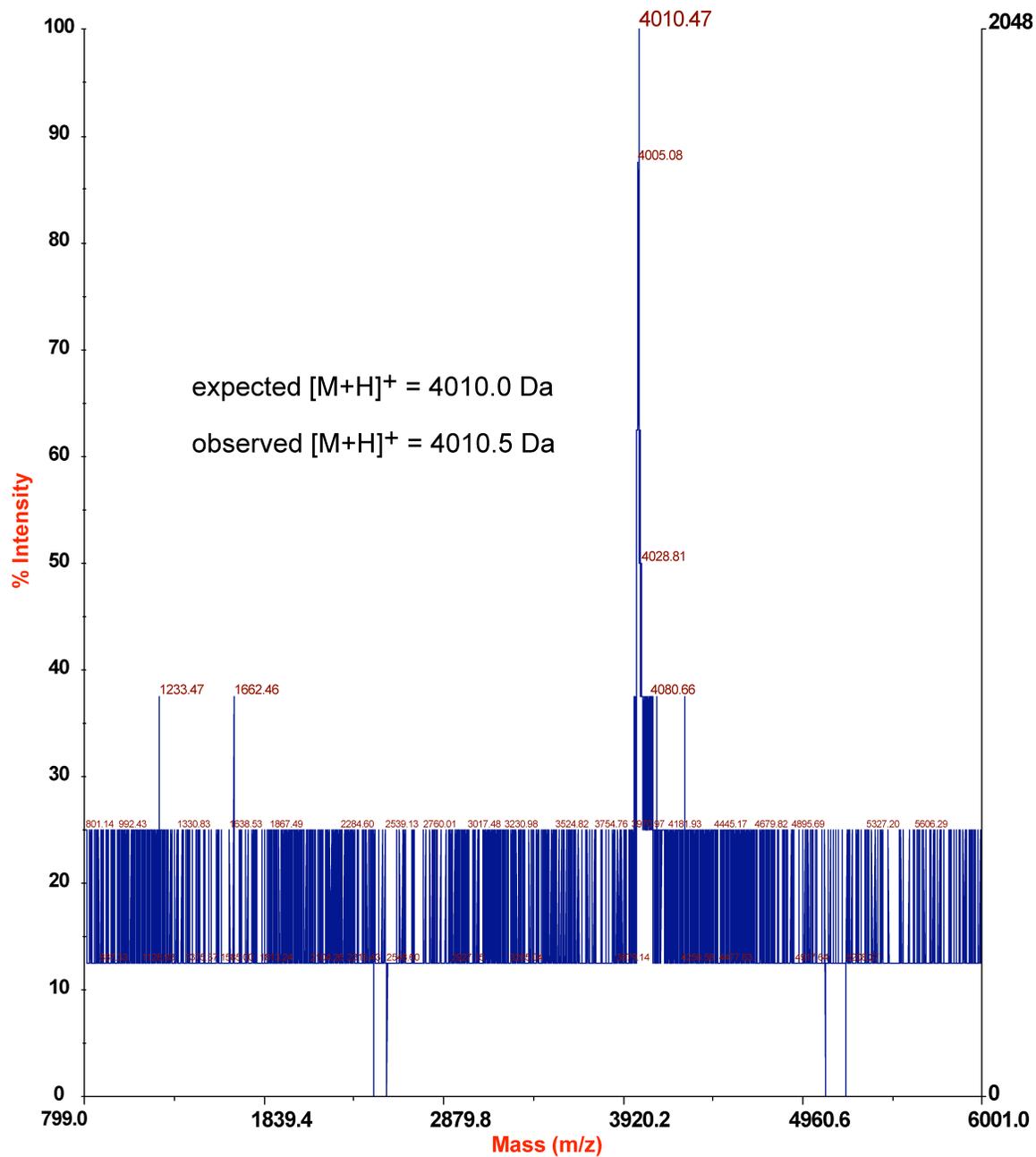
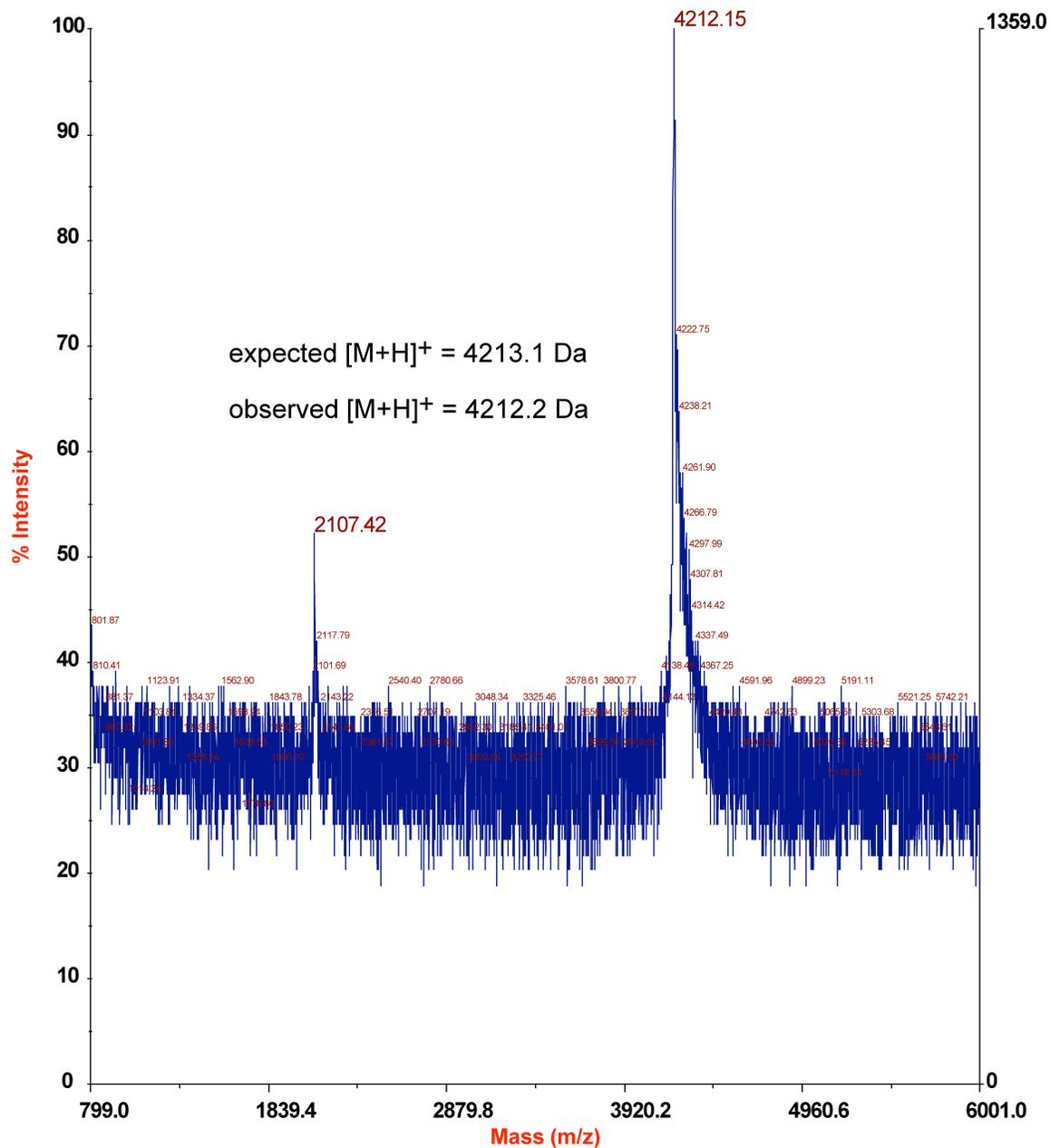


Figure S8. MALDI-TOF spectrum for Pin WW domain protein **19**. Expected $[M+H]^+ = 4010.0$ Da. Observed $[M+H]^+ = 4010.5$ Da.



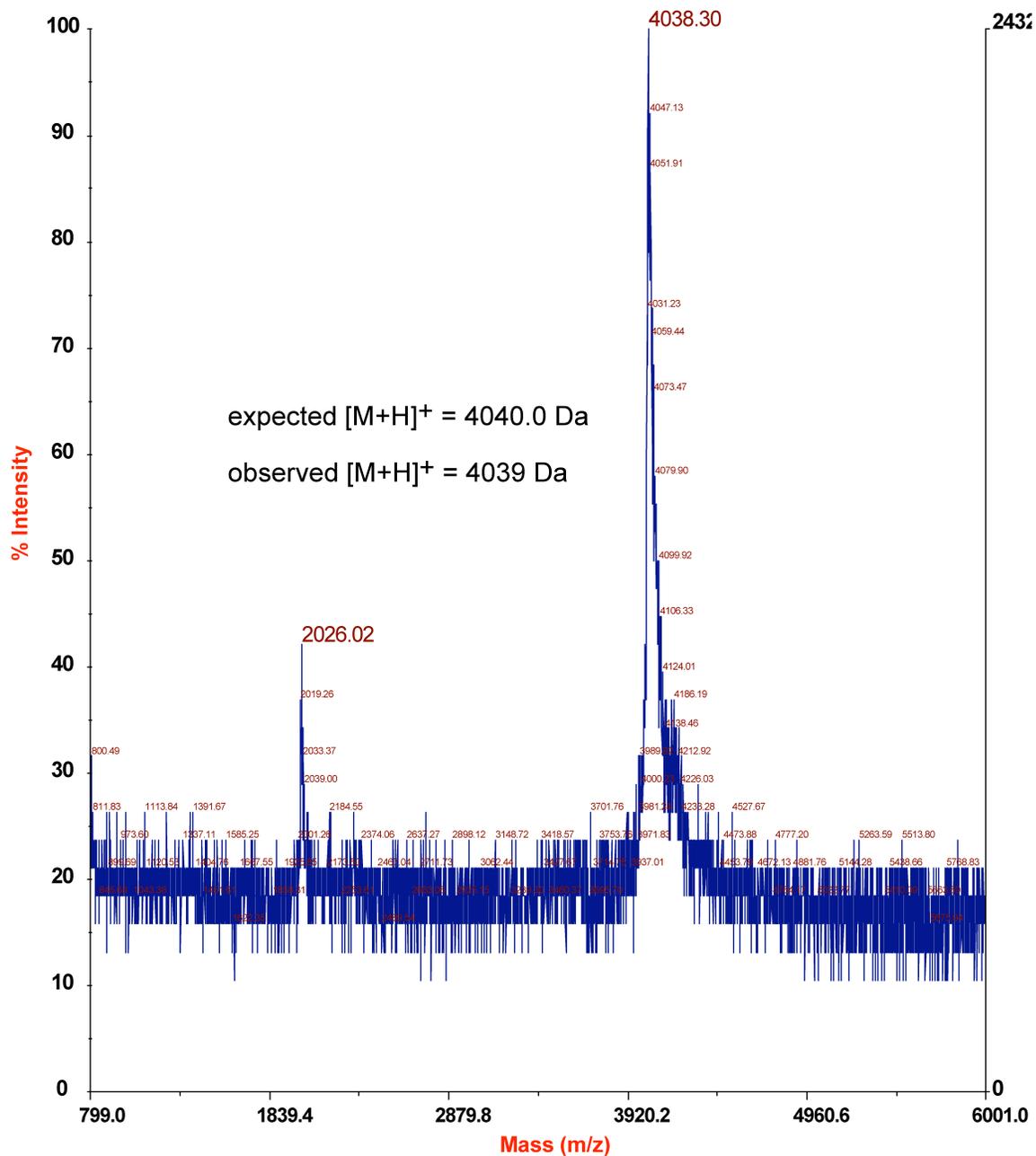


Figure S10. MALDI-TOF spectrum for Pin WW domain protein **20**. Expected $[M+H]^+ = 4040.0$ Da. Observed $[M+H]^+ = 4039$ Da.

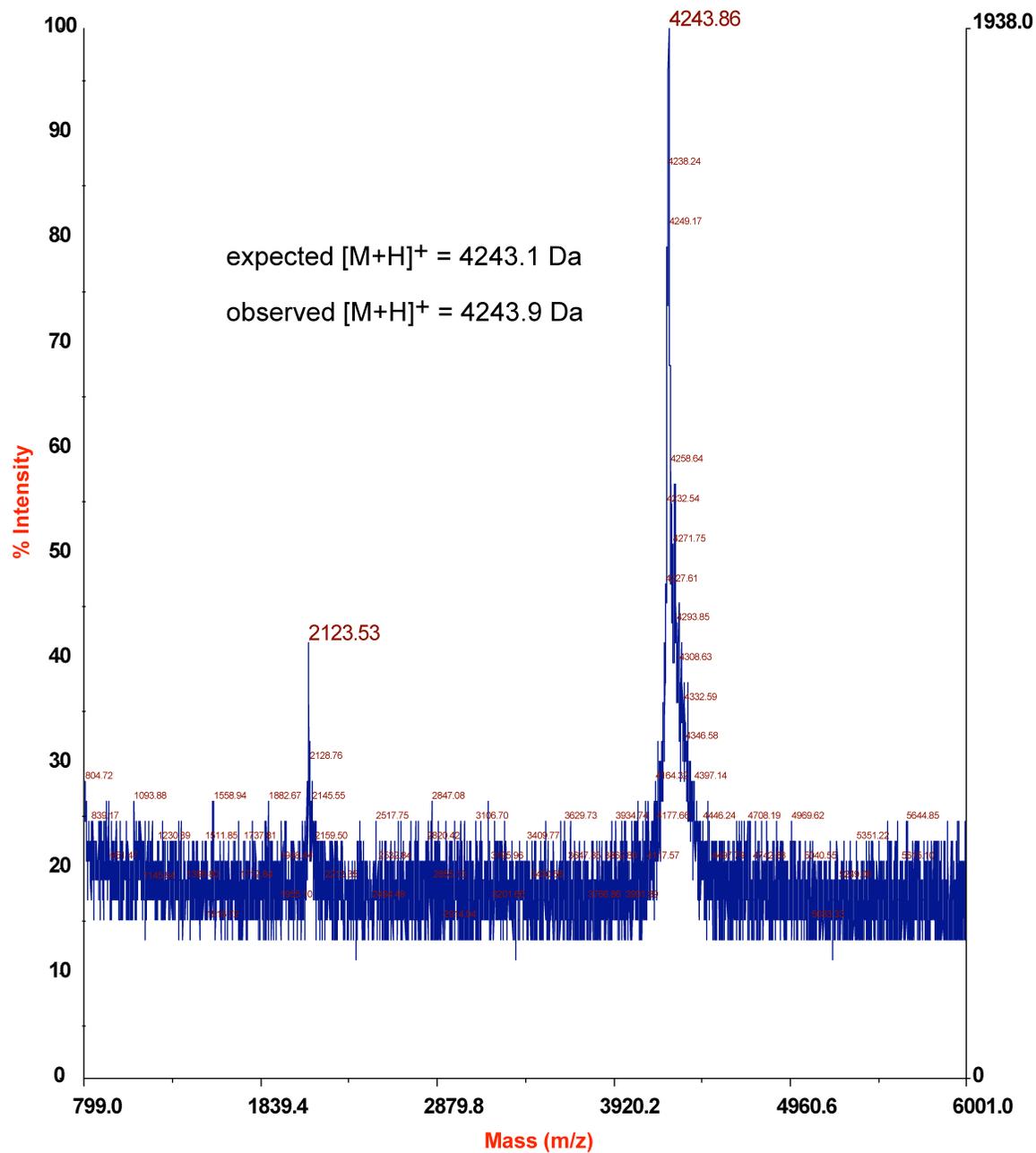


Figure S11. MALDI-TOF spectrum for Pin WW domain protein **20g**. Expected $[M+H]^+ = 4243.1$ Da. Observed $[M+H]^+ = 4243.9$ Da.

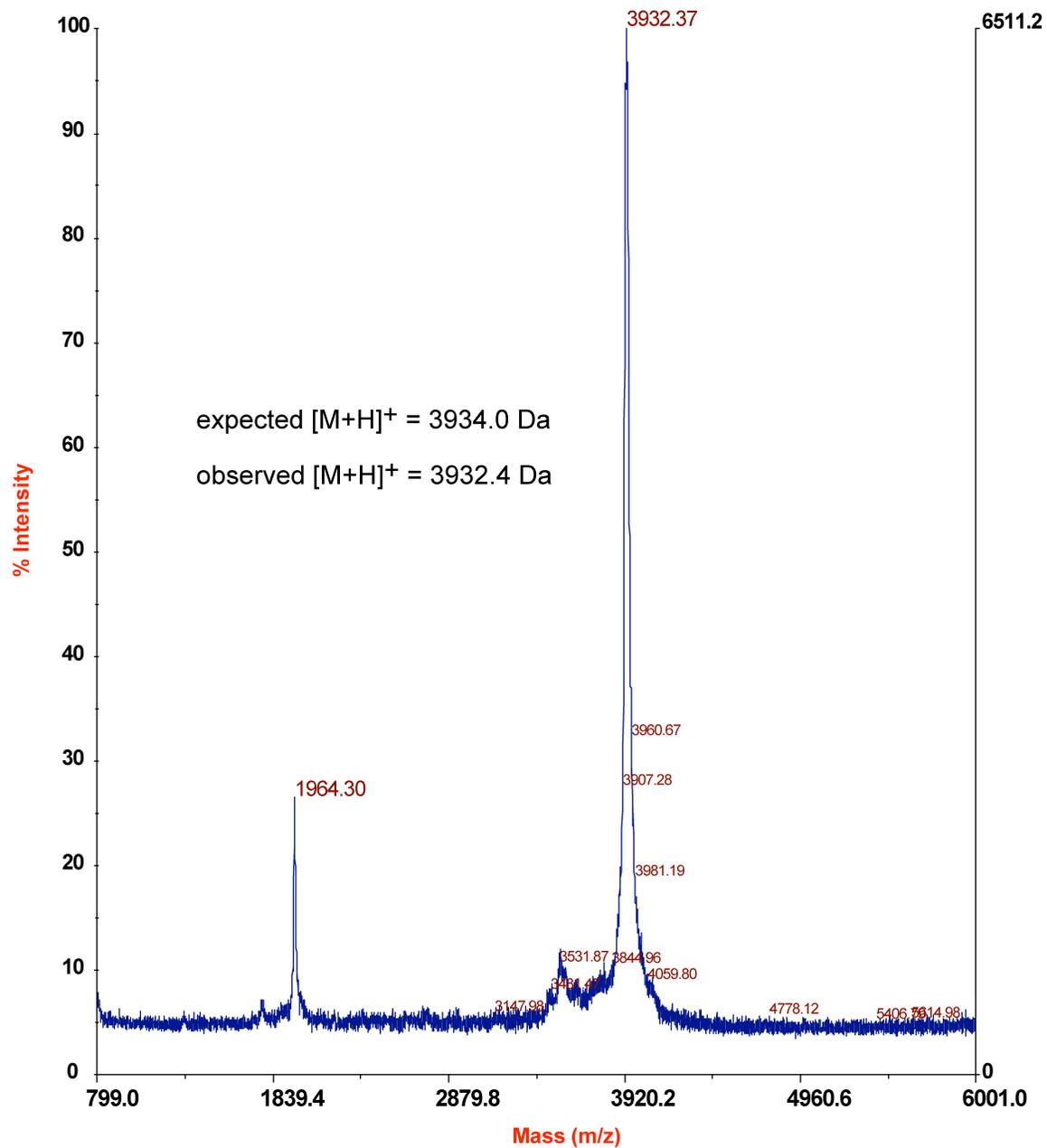


Figure S12. MALDI-TOF spectrum for Pin WW domain protein **23**. Expected $[M+H]^+ = 3934.0$ Da. Observed $[M+H]^+ = 3932.4$ Da.

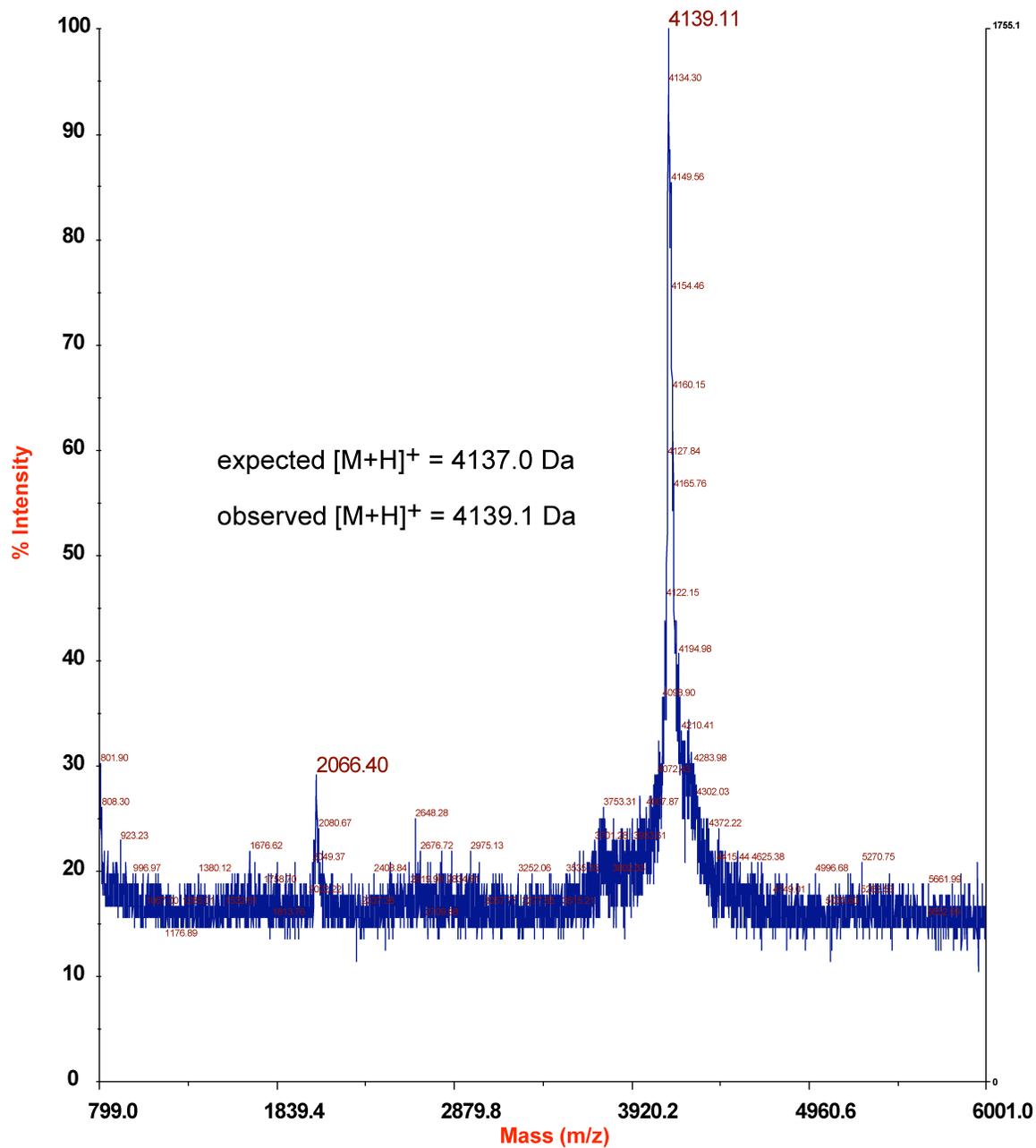


Figure S13. MALDI-TOF spectrum for Pin WW domain protein **23g**. Expected $[M+H]^+ = 4137.0$ Da. Observed $[M+H]^+ = 4139.1$ Da.

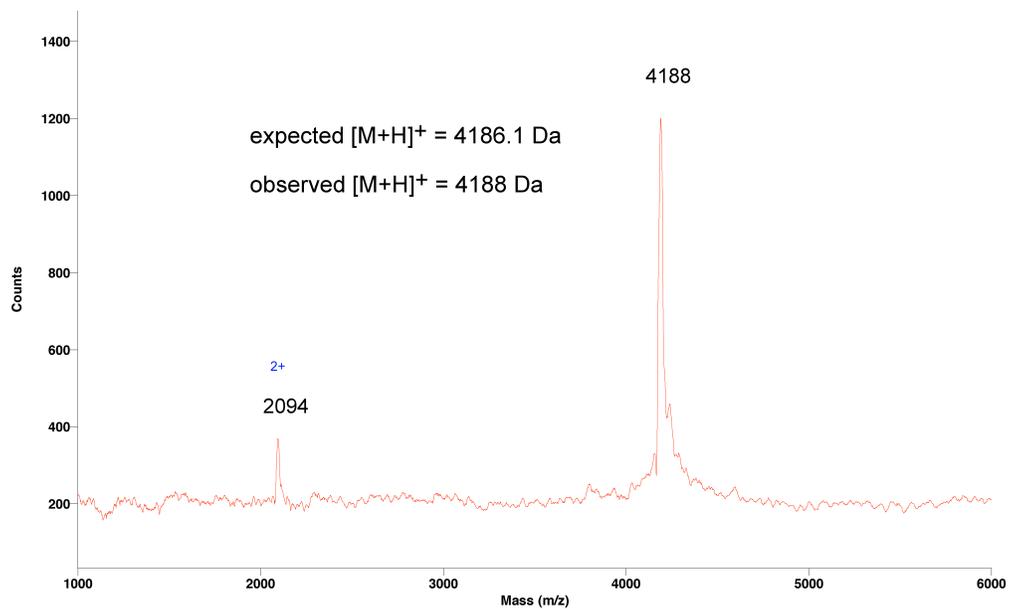


Figure S14. MALDI-TOF spectrum for Pin WW domain protein **26g**. Expected $[M+H]^+ = 4186.1$ Da. Observed $[M+H]^+ = 4188$ Da.

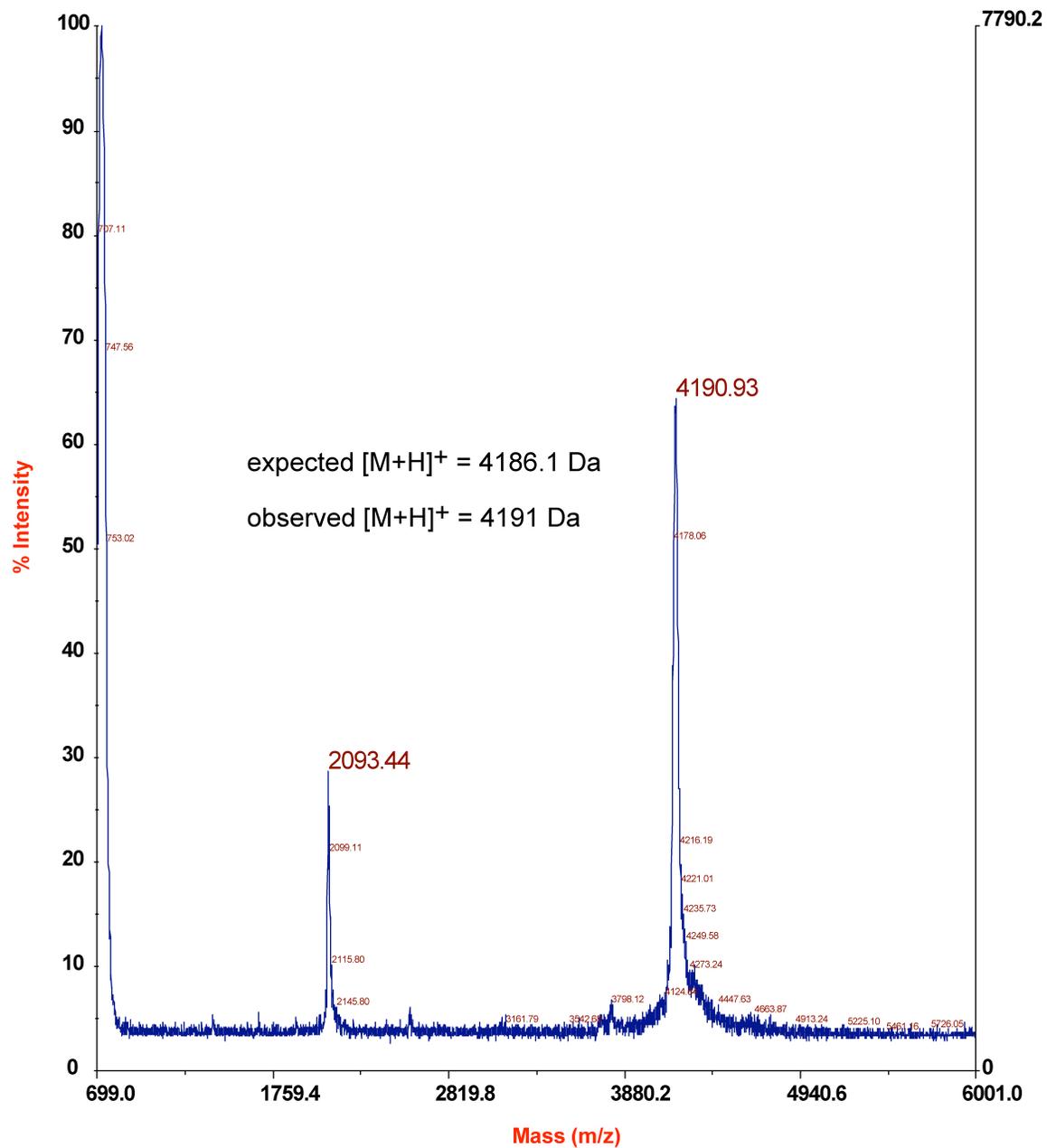


Figure S15. MALDI-TOF spectrum for Pin WW domain protein **30g**. Expected $[M+H]^+ = 4186.1$ Da. Observed $[M+H]^+ = 4191$ Da.

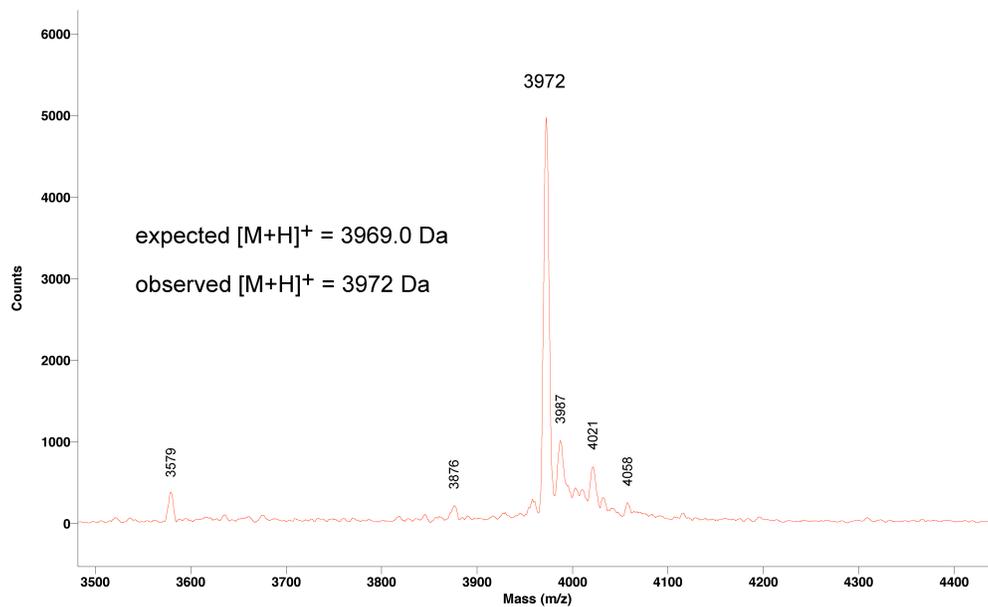


Figure S16. MALDI-TOF spectrum for Pin WW domain protein **33**. Expected $[M+H]^+ = 3969.0$ Da. Observed $[M+H]^+ = 3972$ Da.

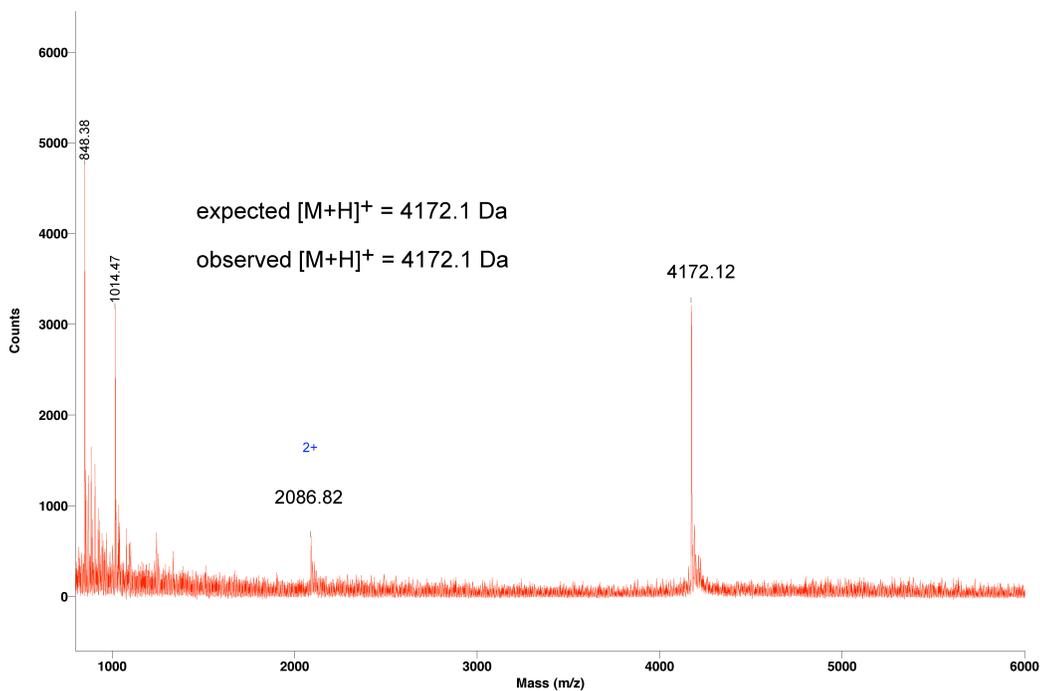


Figure S17. MALDI-TOF spectrum for Pin WW domain protein **33g**. Expected $[M+H]^+ = 4172.1$ Da. Observed $[M+H]^+ = 4172.1$ Da.

HPLC

HPLC traces for proteins **PinWW**, **14–33g** are shown in Figures S18–S34.

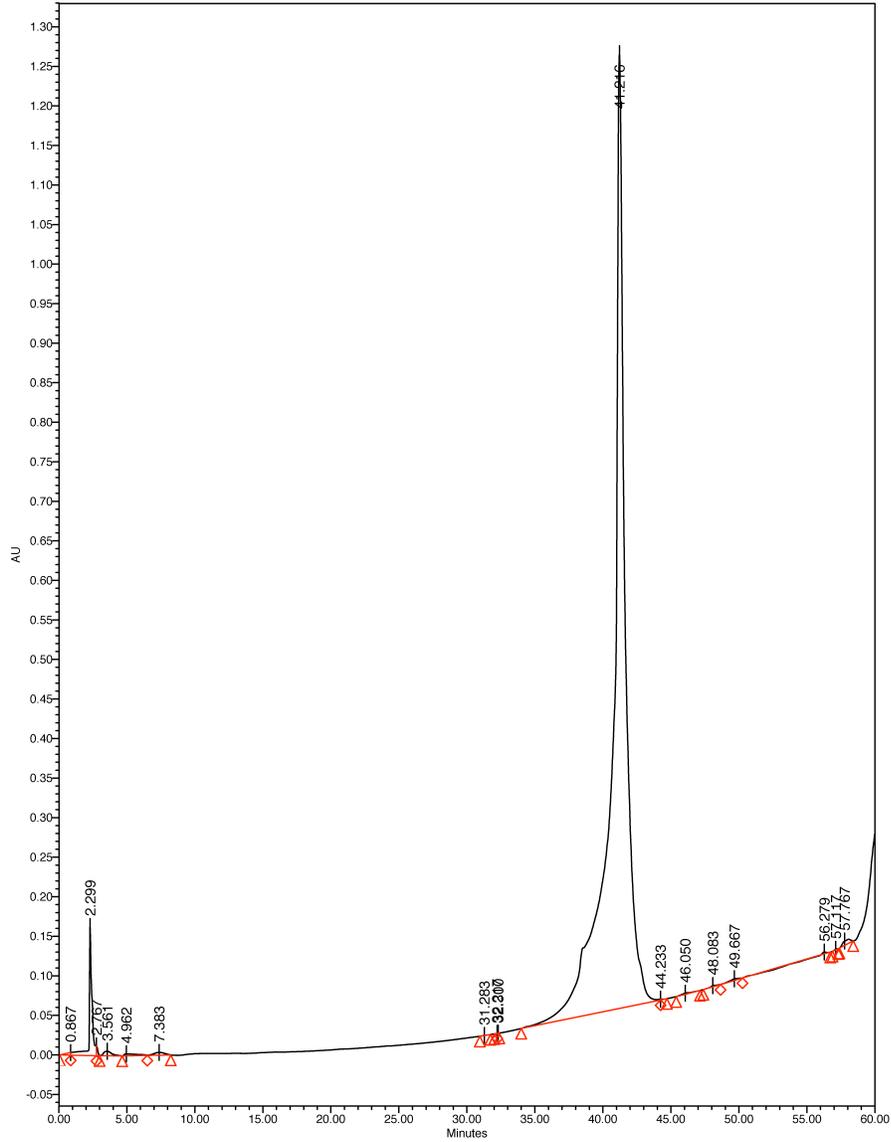


Figure S18. Analytical HPLC data for **Pin WW**. Protein solution was injected onto a C4 analytical column eluted using a linear gradient of 15–65% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 1 ml/min.

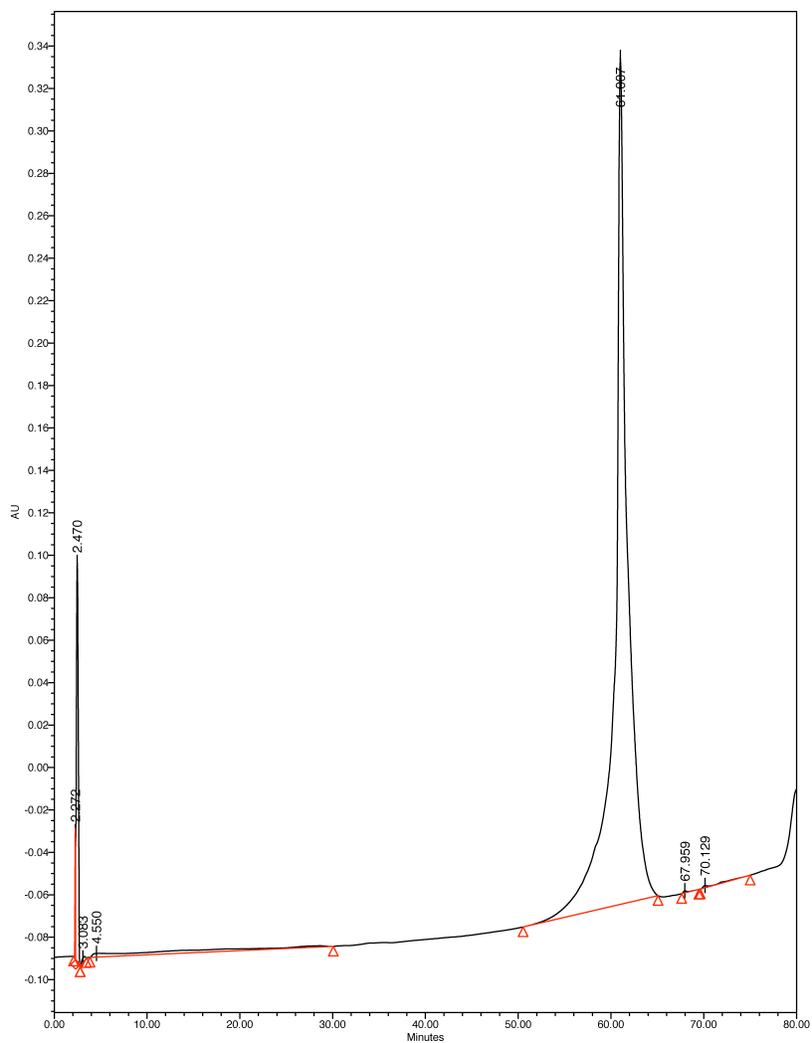


Figure S19. Analytical HPLC data for Pin WW domain protein **14**. Protein solution was injected onto a C4 analytical column eluted using a linear gradient of 0–65% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 65 minutes, with a flow rate of 1 mL/min.

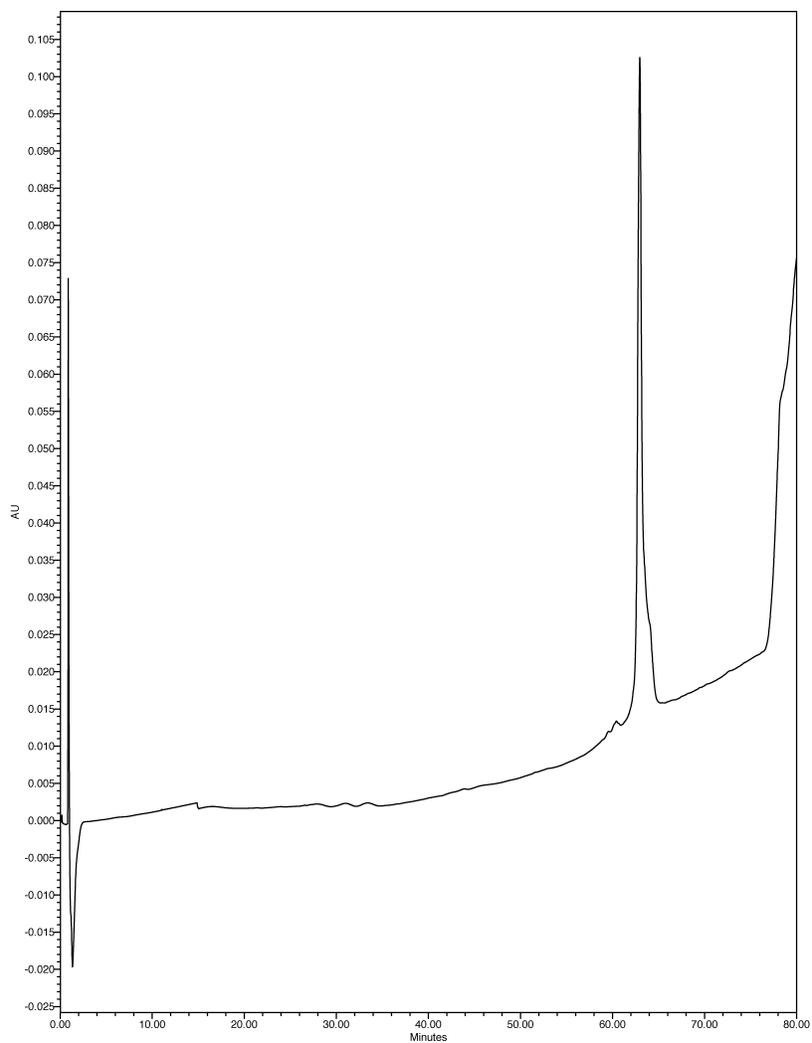


Figure S20. Analytical HPLC data for Pin WW domain protein **14g**. Protein solution was injected onto a C18 analytical column eluted using a linear gradient of 0–65% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 65 minutes, with a flow rate of 1 mL/min.

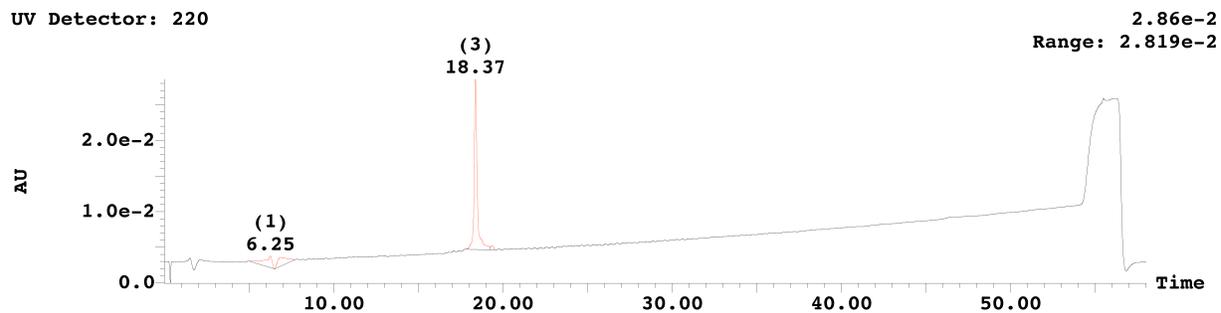


Figure S21. Analytical HPLC data for Pin WW domain protein **17**. Protein solution was injected onto a 4.6 mm x 50 mm C18 analytical column eluted using a linear gradient of 10–60% B (A = H₂O, 0.2% TFA; B = 95% acetonitrile, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 3 ml/min.

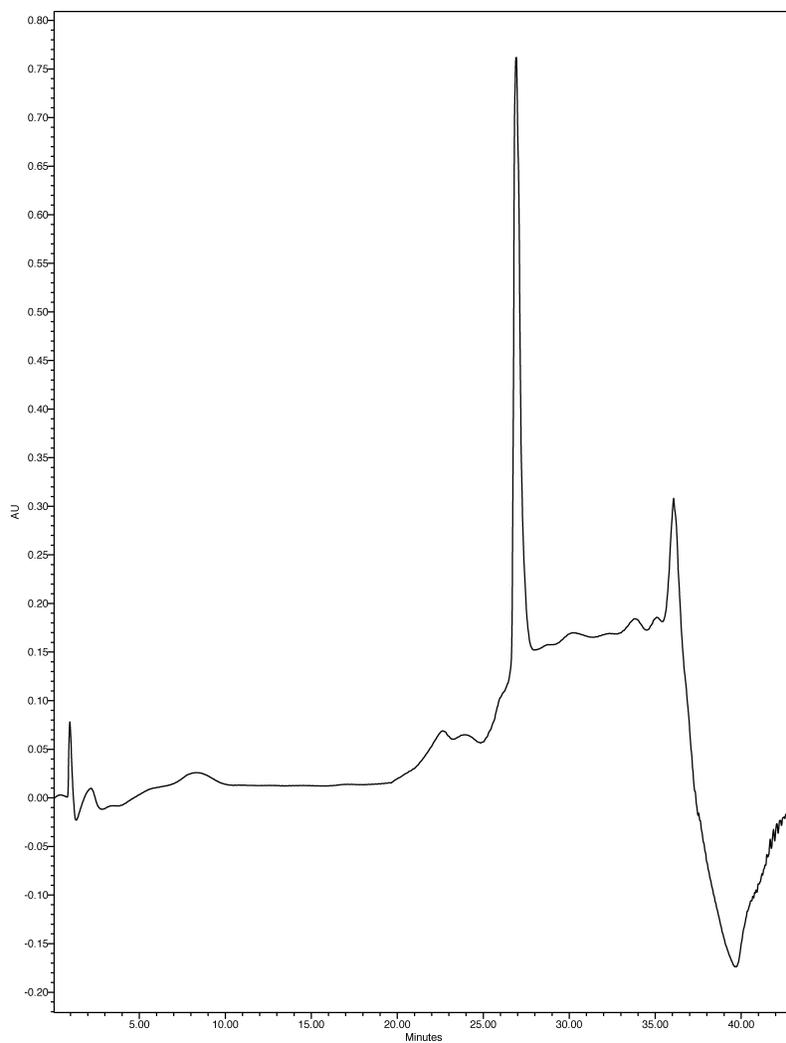


Figure S22. Analytical HPLC data for Pin WW domain protein **17g**. Protein solution was injected onto a C18 analytical column eluted using a linear gradient of 50–60% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 20 minutes, with a flow rate of 1 ml/min.

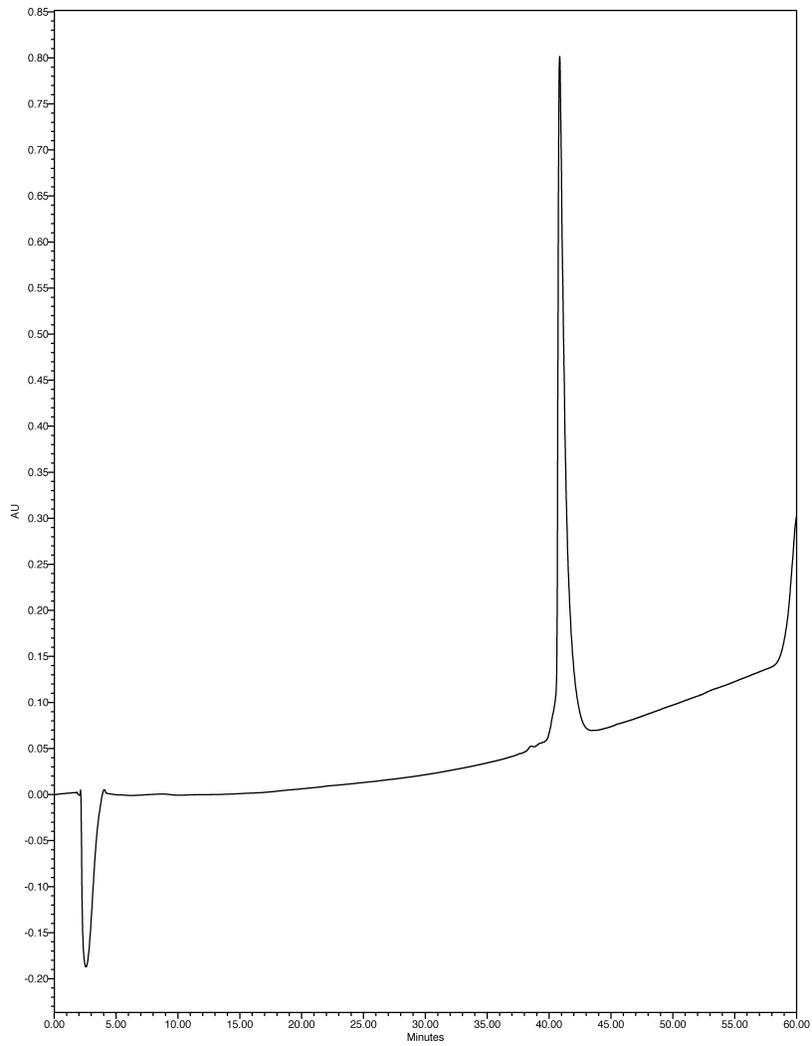


Figure S23. Analytical HPLC data for Pin WW domain protein **18**. Protein solution was injected onto a C4 analytical column eluted using a linear gradient of 15–65% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 1 ml/min.

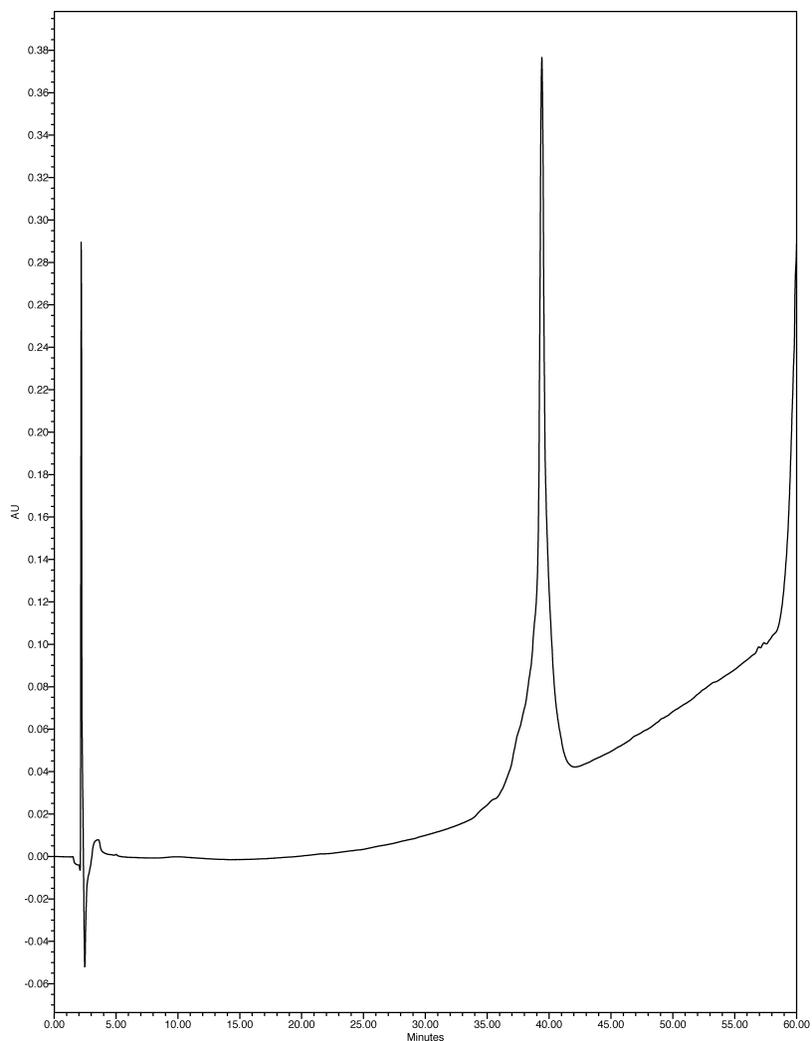


Figure S24. Analytical HPLC data for Pin WW domain protein **18g**. Protein solution was injected onto a C4 analytical column eluted using a linear gradient of 15–65% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 1 ml/min.

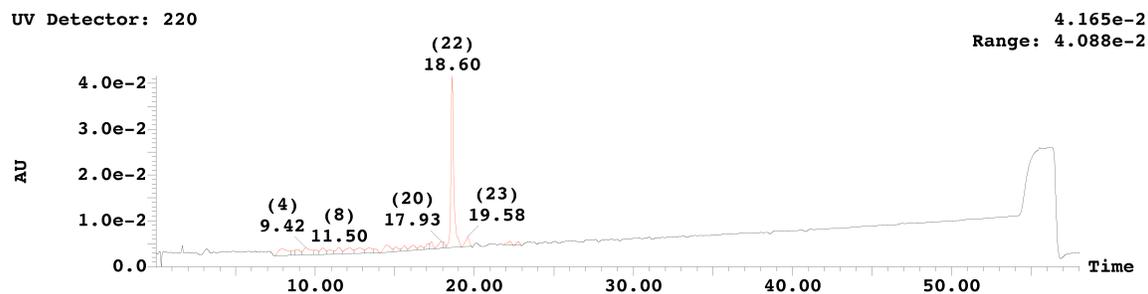


Figure S25. Analytical HPLC data for Pin WW domain protein **19**. Protein solution was injected onto a 4.6 mm x 50 mm C18 analytical column eluted using a linear gradient of 10–60% B (A = H₂O, 0.2% TFA; B = 95% acetonitrile, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 3 ml/min.

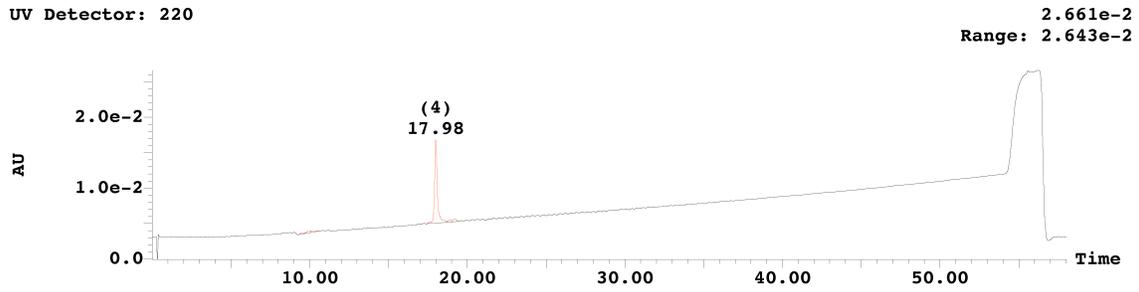


Figure S26. Analytical HPLC data for Pin WW domain protein **19g**. Protein solution was injected onto a 4.6 mm x 50 mm C18 analytical column eluted using a linear gradient of 10–60% B (A = H₂O, 0.2% TFA; B = 95% acetonitrile, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 3 ml/min.

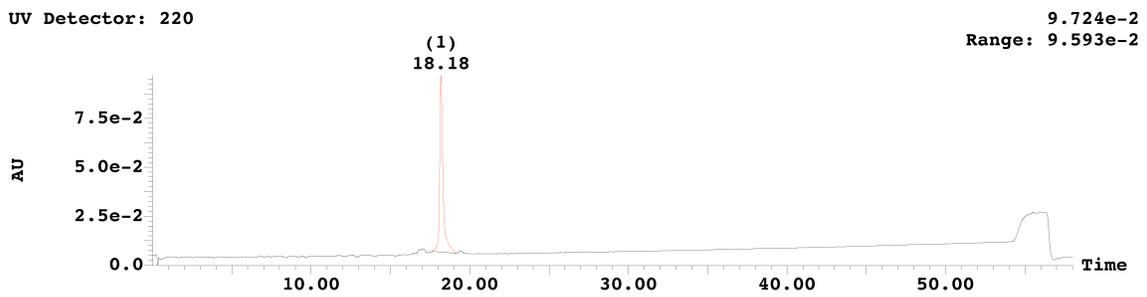


Figure S27. Analytical HPLC data for Pin WW domain protein **20**. Protein solution was injected onto a 4.6 mm x 50 mm C18 analytical column eluted using a linear gradient of 10–60% B (A = H₂O, 0.2% TFA; B = 95% acetonitrile, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 3 ml/min.

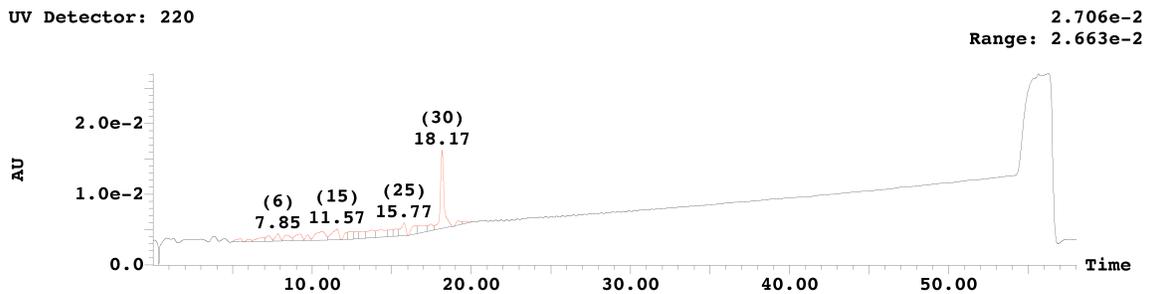


Figure S28. Analytical HPLC data for Pin WW domain protein **20g**. Protein solution was injected onto a 4.6 mm x 50 mm C18 analytical column eluted using a linear gradient of 10–60% B (A = H₂O, 0.2% TFA; B = 95% acetonitrile, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 3 ml/min.

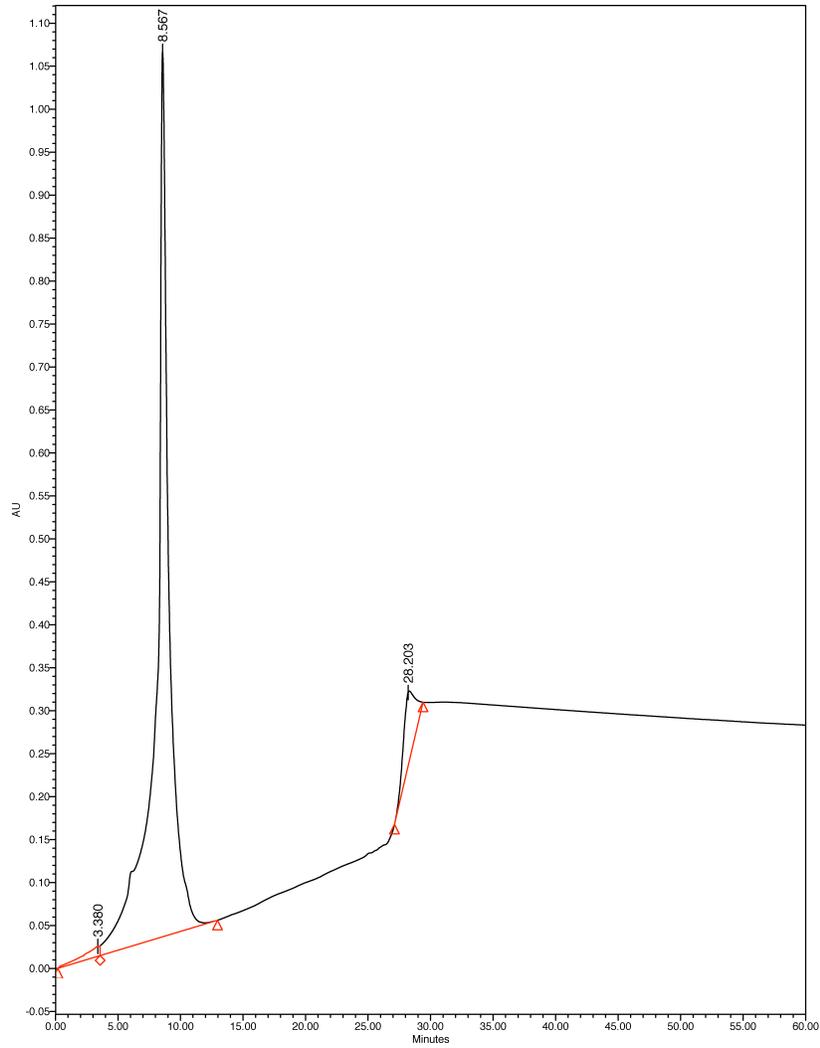


Figure S29. Analytical HPLC data for Pin WW domain protein **23**. Protein solution was injected onto a C4 analytical column eluted using a linear gradient of 15–65% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 1 ml/min. This trace only shows the last few minutes of that gradient (because of a mistake in the data collection software).

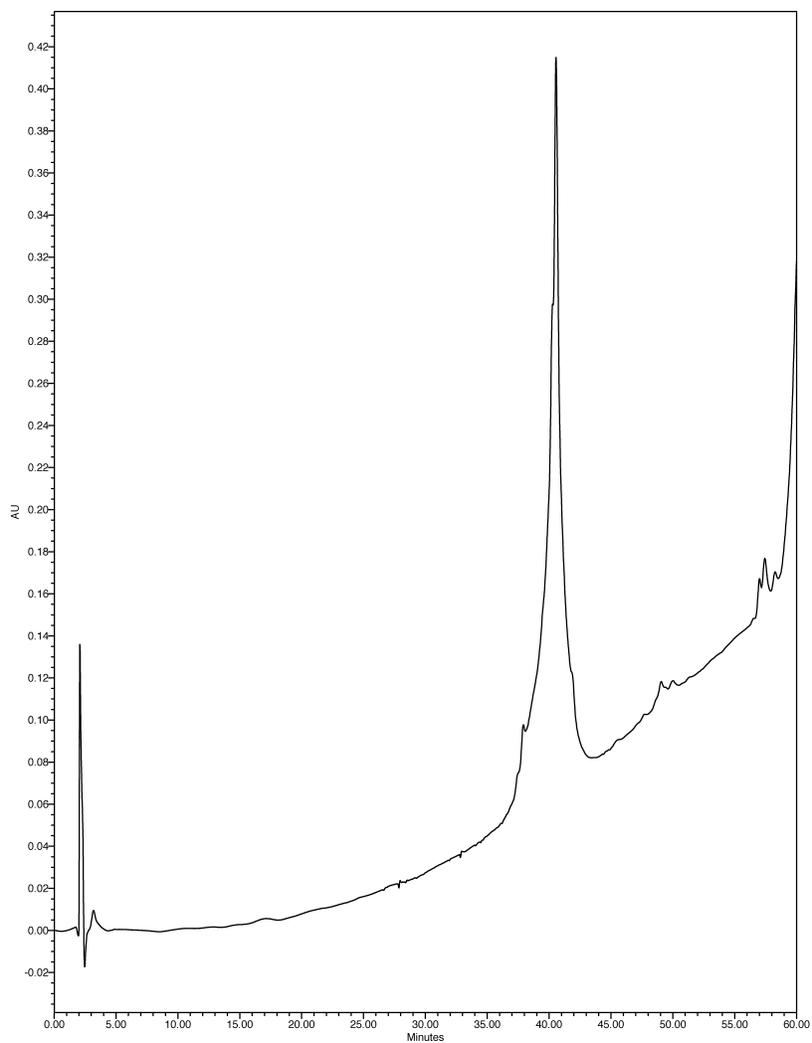


Figure S30. Analytical HPLC data for Pin WW domain protein **23g**. Protein solution was injected onto a C4 analytical column eluted using a linear gradient of 15–65% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 1 ml/min.

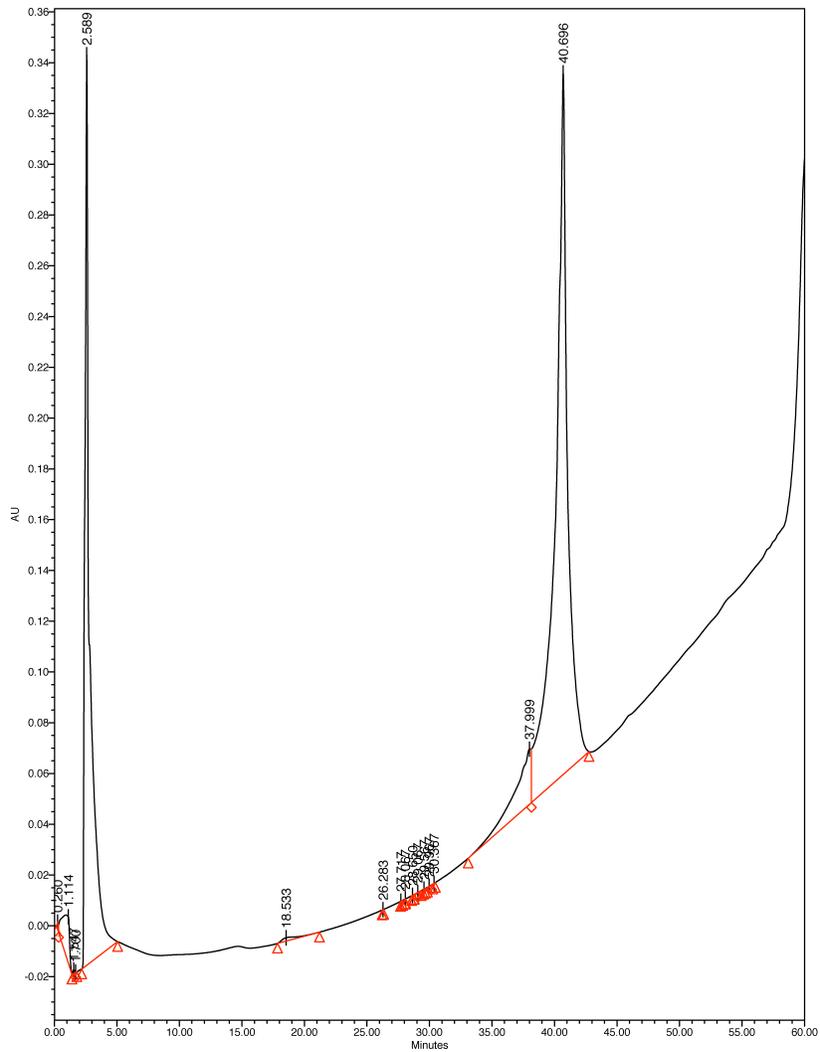


Figure S31. Analytical HPLC data for Pin WW domain protein **26g**. Peptide solution was injected onto a C4 analytical column eluted using a linear gradient of 15–65% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 1 ml/min.

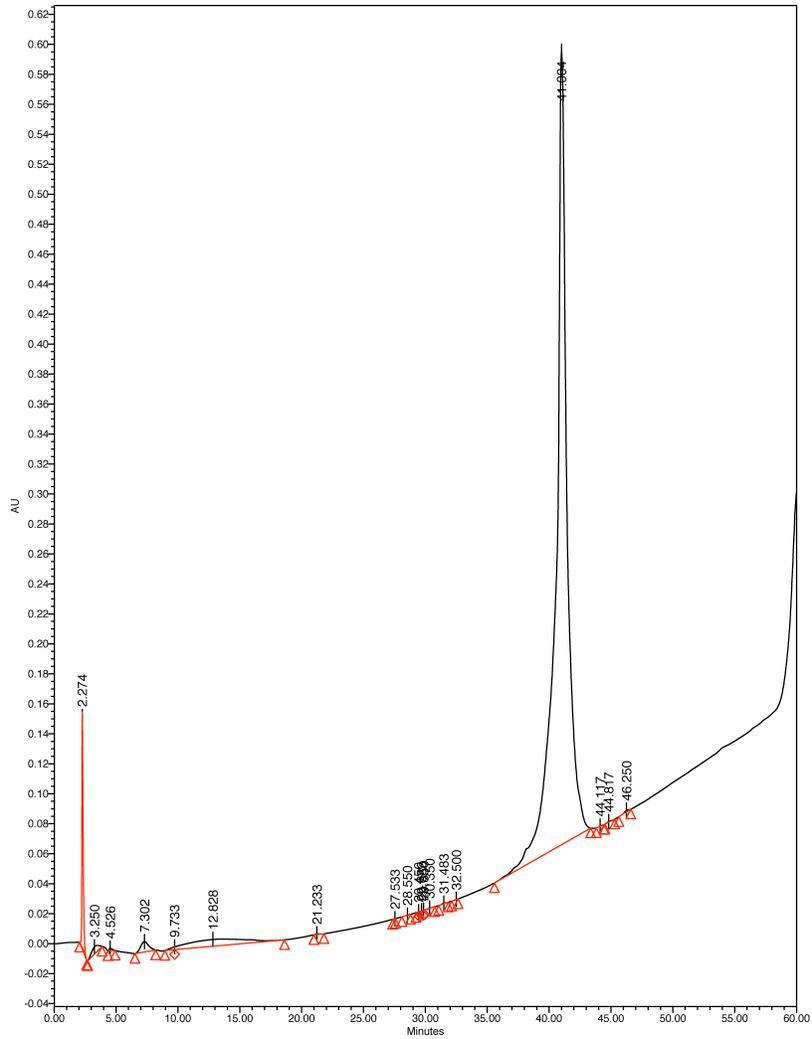


Figure S32. Analytical HPLC data for Pin WW domain protein **30g**. Peptide solution was injected onto a C4 analytical column eluted using a linear gradient of 15–65% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 1 ml/min.

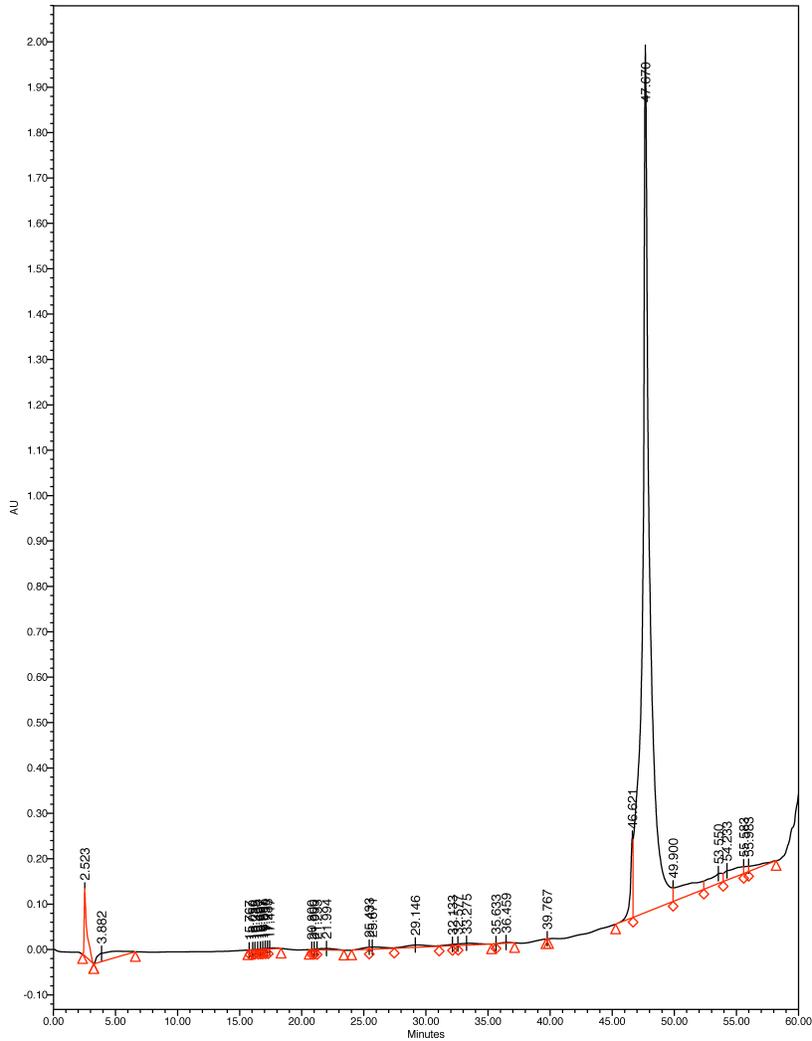


Figure S33. Analytical HPLC data for Pin WW domain protein **33**. Peptide solution was injected onto a C4 analytical column eluted using a linear gradient of 15–65% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 50 minutes, with a flow rate of 1 ml/min.

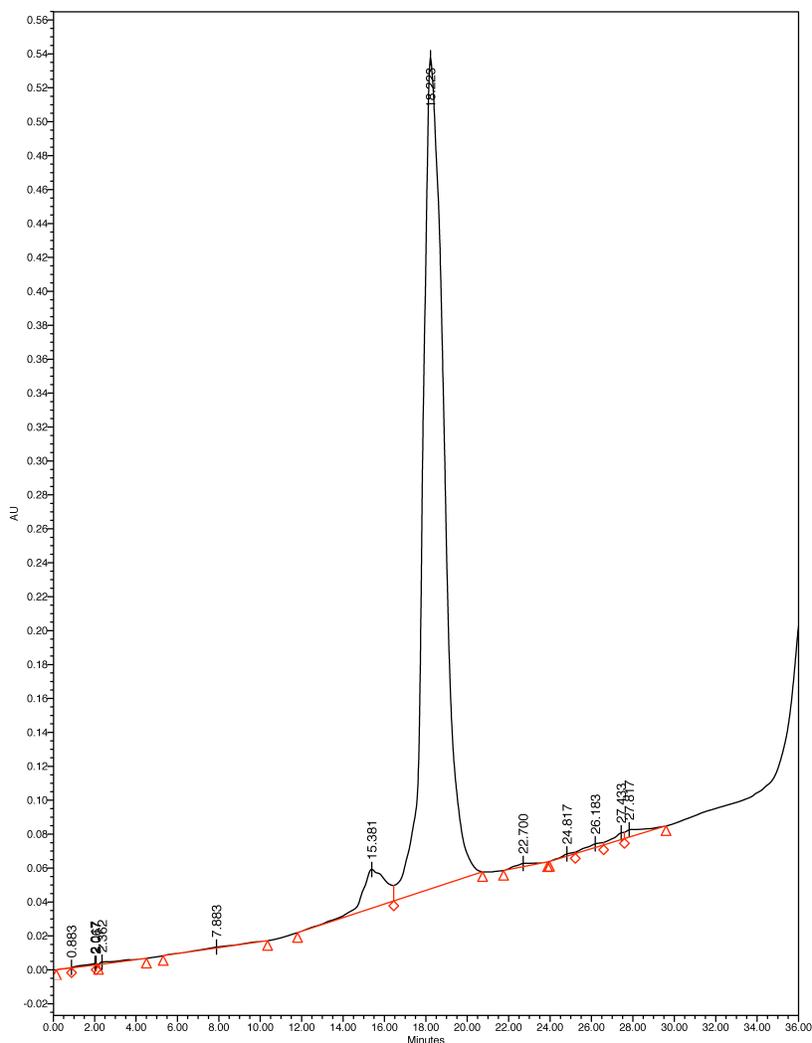


Figure S34. Analytical HPLC data for Pin WW domain protein **33g**. Peptide solution was injected onto a C4 analytical column eluted using a linear gradient of 39–65% B (A = H₂O, 0.2% TFA; B = 95% MeOH, 5% H₂O, 0.2 % TFA) over 26 minutes, with a flow rate of 1 ml/min.

Circular Dichroism Spectroscopy

Measurements were made with an Aviv 62A DS Circular Dichroism Spectrometer, using quartz cuvettes with a 0.1 cm path length. Protein solutions were prepared in 10 mM sodium phosphate buffer, pH 7, and protein concentrations were determined spectroscopically based on tyrosine and tryptophan absorbance at 280 nm in 6 M guanidine hydrochloride + 20 mM sodium phosphate ($\epsilon_{\text{Trp}} = 5690 \text{ M}^{-1}\text{cm}^{-1}$, $\epsilon_{\text{Tyr}} = 1280 \text{ M}^{-1}\text{cm}^{-1}$).⁵ CD spectra were obtained by monitoring

molar ellipticity from 340 to 200 nm, with 5 second averaging times. Variable temperature CD data were obtained by monitoring molar ellipticity at 227 nm from 0.2 to 98.2 °C at 2 °C intervals, with 90 s equilibration time between data points and 30 s averaging times.

Variable temperature CD data were fit to the following model for two-state thermally induced unfolding transitions (in cases where clear pre- and post-transition baselines were visible):

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

where T is temperature in Kelvin, D_0 is the y-intercept and D_1 is the slope of the post-transition baseline; N_0 is the y-intercept and N_1 is the slope of the pre-transition baseline; and K_f is the temperature-dependent folding equilibrium constant. K_f is related to the temperature-dependent free energy of folding $\Delta G_f(T)$ according to the following equation:

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

where R is the universal gas constant (0.0019872 kcal/mol/K). The midpoint of the thermal unfolding transition (or melting temperature T_m) was calculated by fitting $\Delta G_f(T)$ to either of two equations. The first equation is derived from the van't Hoff relationship:

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

where $\Delta H(T_m)$ is the enthalpy of folding at the melting temperature and ΔC_p is the heat capacity of folding ($\Delta H(T_m)$, ΔC_p , and T_m are parameters of the fit). The second equation represents $\Delta G_f(T)$ as a Taylor series expansion about the melting temperature:

$$\Delta G_f(T) = \Delta G_0 + \Delta G_1(T - T_m) + \Delta G_2(T - T_m)^2, \quad (\text{S4})$$

in which ΔG_0 , ΔG_1 , and ΔG_2 are parameters of the fit and T_m a constant obtained from the van't Hoff fit. The T_m values displayed in the main text for each protein were obtained by averaging

the T_m values from each of three or more replicate variable temperature CD experiments on the same protein.

CD spectra and variable-temperature CD data for proteins **Pin WW**, **14-33g** appear in Figures S35-S43, along with parameters from equations S1–S4 that were used to fit the variable temperature CD data. The standard error for each fitted parameter is also shown. These standard parameter errors were used to estimate the uncertainty in the average melting temperatures shown in Table 1 of the main text, along with the uncertainty in the folding and unfolding rates show in Table 2 of the main text by propagation of error.

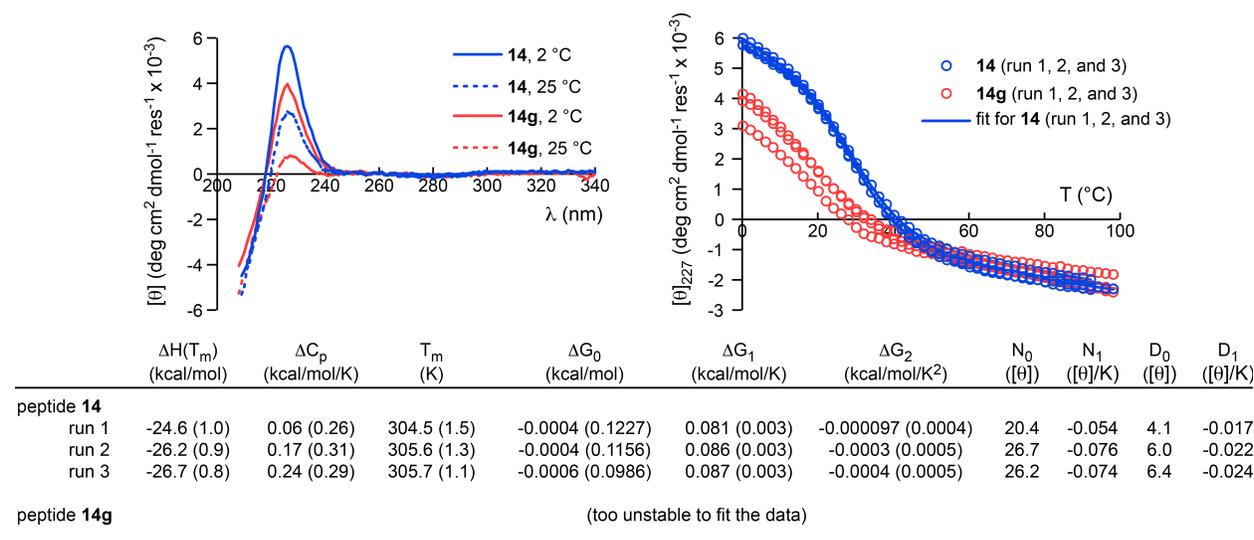


Figure S35. CD spectra and variable temperature CD data for Pin WW domain protein **14** (which has Asn at position 14) and Pin WW domain glycoprotein **14g** (which has an Asn-linked GlcNAc residue at position 14) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S1-S4) appear in the table, along with parameter standard errors in parentheses.

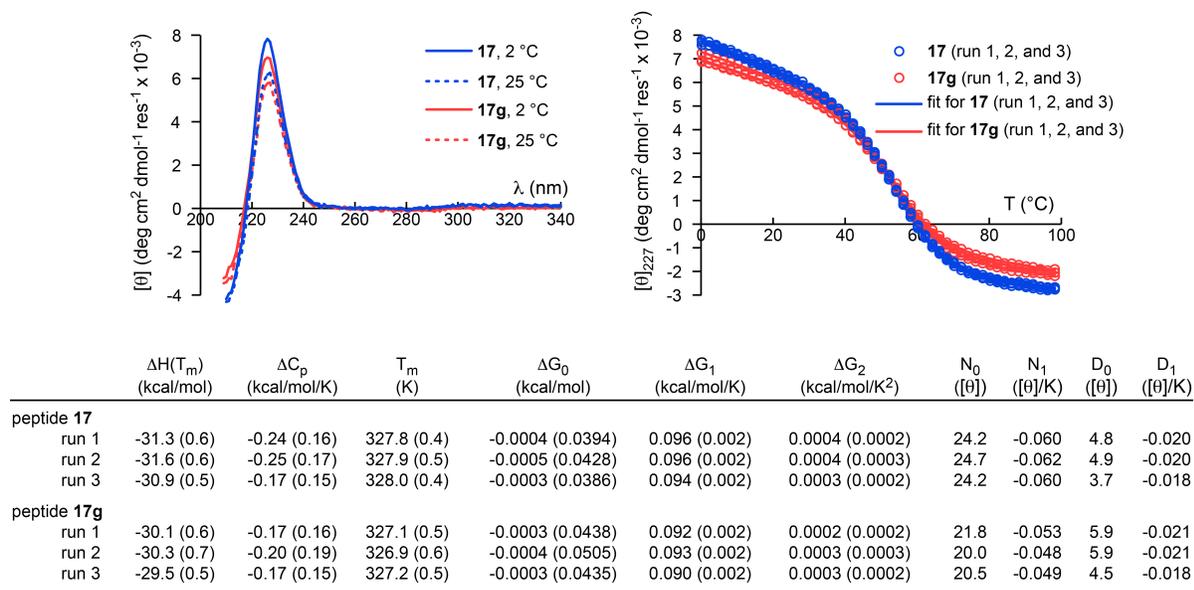


Figure S36. CD spectra and variable temperature CD data for Pin WW domain protein **17** (which has Asn at position 17) and Pin WW domain glycoprotein **17g** (which has an Asn-linked GlcNAc residue at position 17) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S1-S4) appear in the table, along with parameter standard errors in parentheses.

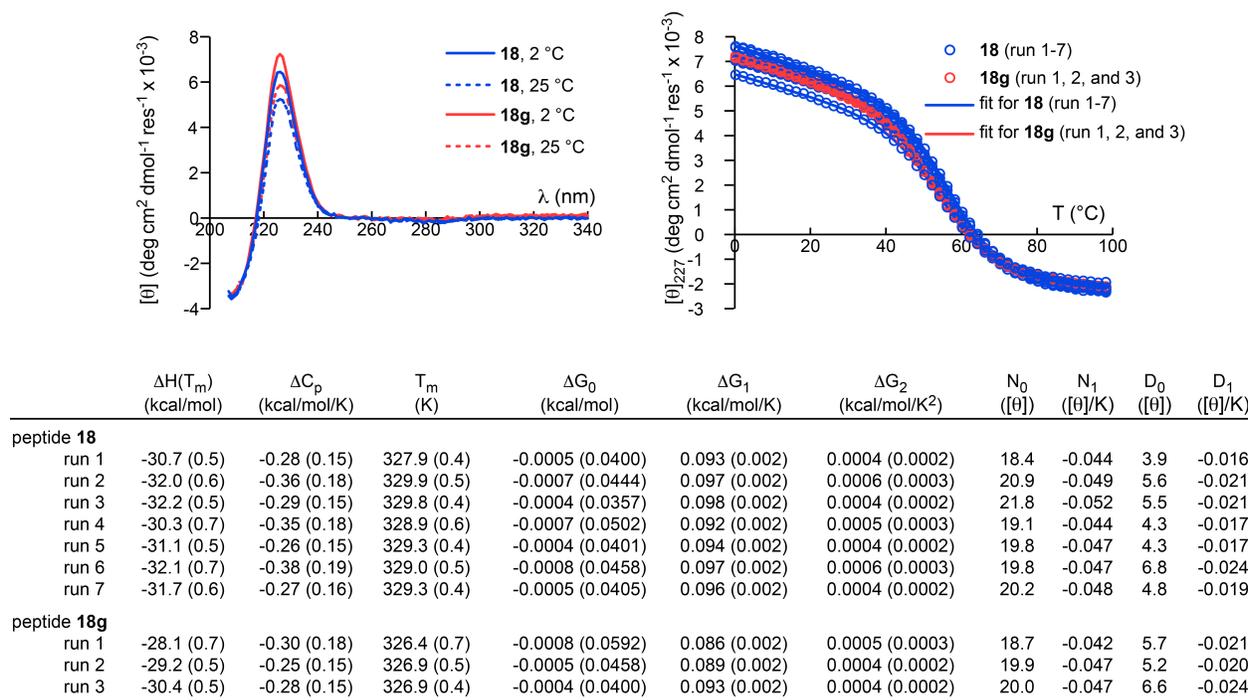
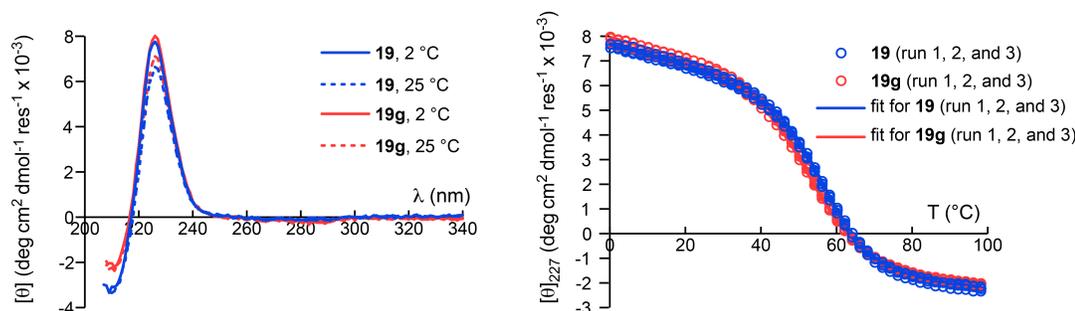
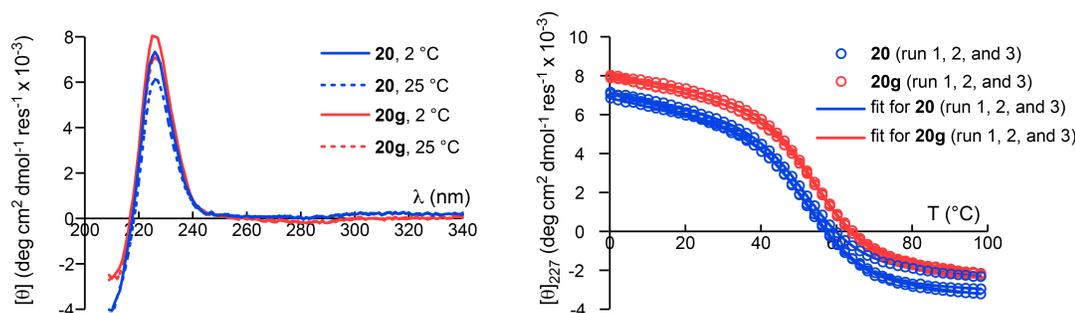


Figure S37. CD spectra and variable temperature CD data for Pin WW domain protein **18** (which has Asn at position 18) and Pin WW domain glycoprotein **18g** (which has an Asn-linked GlcNAc residue at position 18) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S1-S4) appear in the table, along with parameter standard errors in parentheses.



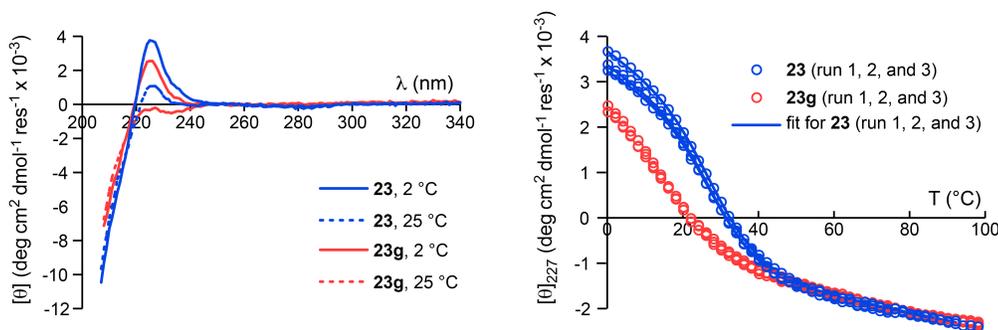
	$\Delta H(T_m)$ (kcal/mol)	ΔC_p (kcal/mol/K)	T_m (K)	ΔG_0 (kcal/mol)	ΔG_1 (kcal/mol/K)	ΔG_2 (kcal/mol/K ²)	N_0 ([θ])	N_1 ([θ]/K)	D_0 ([θ])	D_1 ([θ]/K)
peptide 19										
run 1	-33.1 (0.6)	-0.37 (0.18)	329.3 (0.4)	-0.0006 (0.0404)	0.101 (0.002)	0.0006 (0.0003)	18.7	-0.040	5.6	-0.021
run 2	-33.1 (0.5)	-0.38 (0.15)	329.5 (0.4)	-0.0007 (0.0343)	0.100 (0.002)	0.0006 (0.0002)	18.8	-0.041	5.4	-0.021
run 3	-32.9 (0.6)	-0.41 (0.17)	329.5 (0.4)	-0.0008 (0.0391)	0.100 (0.002)	0.0006 (0.0002)	18.8	-0.040	5.2	-0.021
peptide 19g										
run 1	-30.3 (0.6)	-0.14 (0.15)	326.7 (0.4)	-0.0003 (0.0406)	0.093 (0.002)	0.0002 (0.0002)	19.3	-0.041	4.6	-0.017
run 2	-30.3 (0.5)	-0.10 (0.12)	326.9 (0.4)	-0.0002 (0.0340)	0.093 (0.002)	0.0002 (0.0002)	19.9	-0.043	4.4	-0.017
run 3	-30.8 (0.7)	-0.25 (0.19)	326.1 (0.5)	-0.0005 (0.0475)	0.094 (0.002)	0.0004 (0.0003)	19.0	-0.041	6.6	-0.023

Figure S38. CD spectra and variable temperature CD data for Pin WW domain protein **19** (which has Asn at position 19) and Pin WW domain glycoprotein **19g** (which has an Asn-linked GlcNAc residue at position 19) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S1-S4) appear in the table, along with parameter standard errors in parentheses.



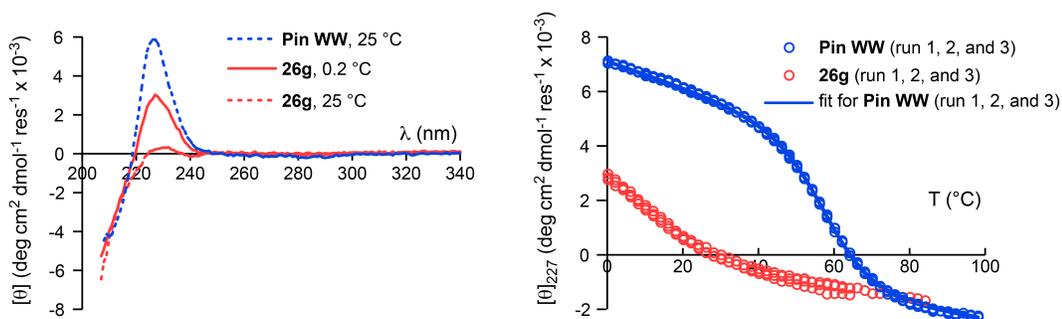
	$\Delta H(T_m)$ (kcal/mol)	ΔC_p (kcal/mol/K)	T_m (K)	ΔG_0 (kcal/mol)	ΔG_1 (kcal/mol/K)	ΔG_2 (kcal/mol/K ²)	N_0 ([θ])	N_1 ([θ]/K)	D_0 ([θ])	D_1 ([θ]/K)
peptide 20										
run 1	-30.5 (0.6)	-0.18 (0.17)	326.3 (0.5)	-0.0004 (0.0453)	0.093 (0.002)	0.0003 (0.0003)	20.3	-0.049	3.7	-0.019
run 2	-30.0 (0.4)	-0.25 (0.12)	325.4 (0.4)	-0.0004 (0.0325)	0.092 (0.001)	0.0004 (0.0002)	18.1	-0.040	4.6	-0.019
run 3	-30.0 (0.4)	-0.25 (0.12)	325.4 (0.4)	-0.0004 (0.0325)	0.092 (0.001)	0.0004 (0.0002)	18.1	-0.040	4.6	-0.019
peptide 20g										
run 1	-31.9 (0.7)	-0.36 (0.18)	327.0 (0.5)	-0.0006 (0.0431)	0.097 (0.002)	0.0006 (0.0003)	17.8	-0.036	8.7	-0.029
run 2	-30.8 (0.5)	-0.11 (0.12)	328.7 (0.4)	-0.0002 (0.0337)	0.094 (0.002)	0.0002 (0.0002)	19.3	-0.042	3.4	-0.015
run 3	-30.7 (0.4)	-0.20 (0.12)	328.1 (0.4)	-0.0003 (0.0334)	0.093 (0.002)	0.0003 (0.0002)	18.9	-0.040	4.7	-0.018

Figure S39. CD spectra and variable temperature CD data for Pin WW domain protein **20** (which has Asn at position 20) and Pin WW domain glycoprotein **20g** (which has an Asn-linked GlcNAc residue at position 20) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S1-S4) appear in the table, along with parameter standard errors in parentheses.



	$\Delta H(T_m)$ (kcal/mol)	ΔC_p (kcal/mol/K)	T_m (K)	ΔG_0 (kcal/mol)	ΔG_1 (kcal/mol/K)	ΔG_2 (kcal/mol/K ²)	N_0 ([θ])	N_1 ([θ]/K)	D_0 ([θ])	D_1 ([θ]/K)
peptide 23										
run 1	-26.7 (0.6)	0.22 (0.24)	303.3 (1.0)	-0.0004 (0.0870)	0.088 (0.002)	-0.0004 (0.0004)	19.7	-0.059	4.1	-0.018
run 2	-27.8 (1.5)	0.22 (0.55)	303.2 (1.9)	-0.0005 (0.1738)	0.092 (0.005)	-0.0004 (0.0009)	19.2	-0.058	4.6	-0.019
run 3	-27.2 (1.5)	0.01 (0.38)	302.3 (1.7)	-0.0003 (0.1461)	0.090 (0.005)	-0.000016 (0.00062)	16.7	-0.049	4.1	-0.018
peptide 23g	(too unstable to fit the data)									

Figure S40. CD spectra and variable temperature CD data for Pin WW domain protein **23** (which has Asn at position 23) and Pin WW domain glycoprotein **23g** (which has an Asn-linked GlcNAc residue at position 23) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S1-S4) appear in the table, along with parameter standard errors in parentheses.



	$\Delta H(T_m)$ (kcal/mol)	ΔC_p (kcal/mol/K)	T_m (K)	ΔG_0 (kcal/mol)	ΔG_1 (kcal/mol/K)	ΔG_2 (kcal/mol/K ²)	N_0 ([θ])	N_1 ([θ]/K)	D_0 ([θ])	D_1 ([θ]/K)
peptide Pin WW										
run 1	-33.5 (0.8)	-0.46 (0.23)	330.7 (0.5)	-0.0010 (0.0512)	0.101 (0.002)	0.0007 (0.0003)	20.1	-0.048	8.1	-0.028
run 2	-33.4 (0.8)	-0.42 (0.23)	330.7 (0.5)	-0.0009 (0.0534)	0.101 (0.002)	0.0006 (0.0003)	20.5	-0.049	7.4	-0.026
run 3	-32.4 (0.9)	-0.45 (0.23)	330.5 (0.6)	-0.0011 (0.0577)	0.098 (0.002)	0.0007 (0.0003)	20.1	-0.048	7.8	-0.027
peptide 26g	(too unstable to fit the data)									

Figure S41. CD spectra and variable temperature CD data for **Pin WW** (which has Asn at position 26) and **Pin WW** domain glycoprotein **26g** (which has an Asn-linked GlcNAc residue at position 26) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S1-S4) appear in the table, along with parameter standard errors in parentheses.

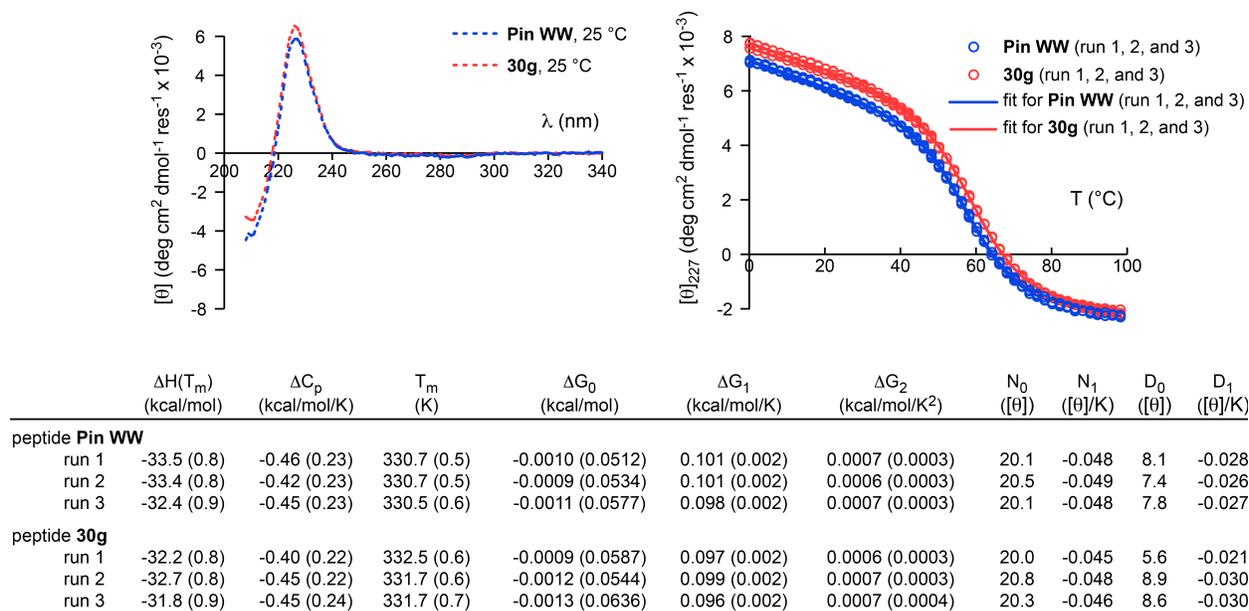


Figure S42. CD spectra and variable temperature CD data for **Pin WW** (which has Asn at position 30) and **Pin WW** domain glycoprotein **30g** (which has an Asn-linked GlcNAc residue at position 30) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S1-S4) appear in the table, along with parameter standard errors in parentheses.

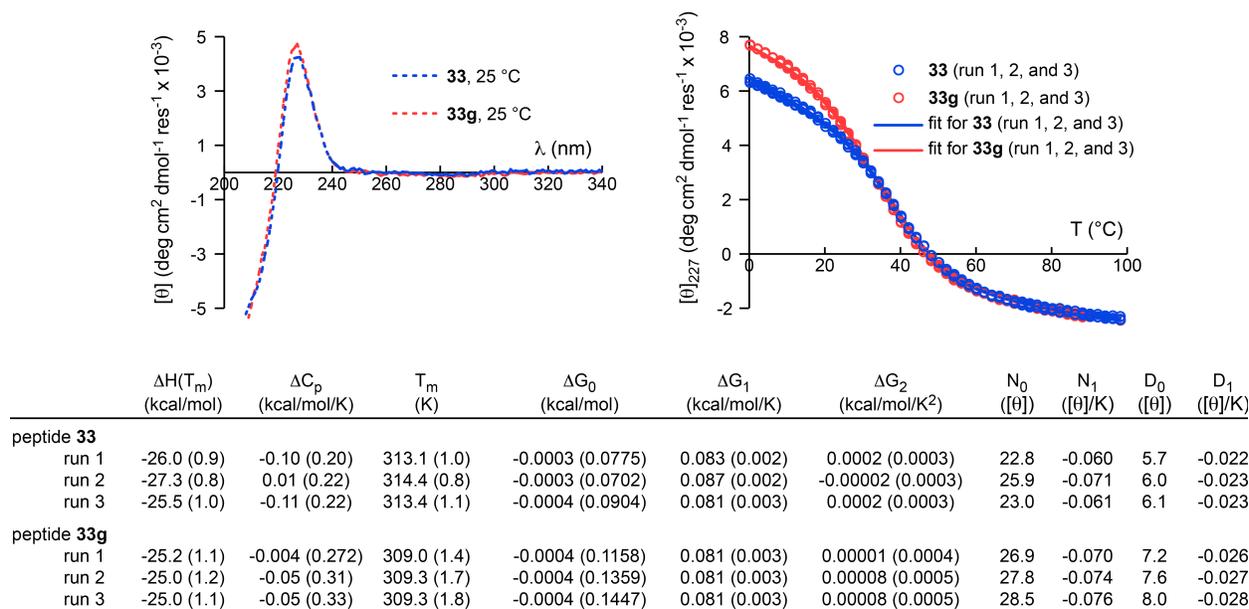


Figure S43. CD spectra and variable temperature CD data for **Pin WW** domain protein **33** (which has Asn at position 33) and **Pin WW** domain glycoprotein **33g** (which has an Asn-linked GlcNAc residue at position 33) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S1-S4) appear in the table, along with parameter standard errors in parentheses.

Laser Temperature Jump Experiments

Relaxation times following a rapid laser-induced temperature jump of ~ 12 °C were measured by monitoring Trp fluorescence of 100 μM solution of Pin WW domain proteins **17**, **17g**, **18**, **18g**, **19**, **19g**, **20**, **20g**, **Pin WW**, **30g**, **33**, and **33g** in 20 mM sodium phosphate (pH 7) using a nanosecond laser temperature jump apparatus, as described previously⁶⁻⁹ to monitor the fluorescence decay of a Trp residue in each protein after a laser-induced temperature jump (see Figures S44-S55) at each of several temperatures

The relaxation traces shown in Figures S44-S55 represent the average of at least 10 individual temperature-jump experiments, and were obtained by fitting the shape f of each fluorescence decay at time t to a linear combination of the fluorescence decay shapes before f_1 and after f_2 the temperature jump:

$$f(t) = a_1(t) \cdot f_1 + a_2(t) \cdot f_2, \quad (\text{S5})$$

where $a_1(t)$ and $a_2(t)$ are the coefficients of the linear combination describing the relative contributions of f_1 and f_2 to the shape of the fluorescence decay at time t .⁹ Then, the relaxation of the protein to equilibrium at the new temperature following the laser-induced temperature jump can be represented as $\chi_1(t)$:

$$\chi_1(t) = \frac{a_1(t)}{a_1(t) + a_2(t)}, \quad (\text{S6})$$

plotted as a function of time for each protein at several temperatures in Figures S44-S55.^{7,8}

The relaxation traces at each temperature were then fit to the following equation:

$$\chi(t) = C_1 \cdot \exp\left[\frac{-(t - x_0)}{\tau}\right] + C_2, \quad (\text{S7})$$

where C_1 and C_2 are constants describing the amplitude of the fluorescence decay, x_0 is a constant that adjusts the measured time to zero after the instantaneous temperature jump, and τ is

the relaxation time, which is the inverse of the observed rate constant k_{obs} ($k_{\text{obs}} = 1/\tau$). Using the temperature-dependent equilibrium constant K_f for each protein (from the variable temperature CD experiments; see equations S2-S4), folding k_f and unfolding k_u rate constants can be extracted from k_{obs} according to the following equations:

$$k_{\text{obs}} = k_f + k_u \quad (\text{S8})$$

$$K_f = \frac{k_f}{k_u} \quad (\text{S9})$$

$$k_f = k_{\text{obs}} \cdot \left[1 - \frac{1}{K_f + 1} \right] \quad (\text{S10})$$

The folding rates for each protein can then be fit as a function of temperature to the following Kramers model¹⁰⁻¹² equation:

$$k_f(T) = \nu(59 \text{ }^\circ\text{C}) \cdot \frac{\eta(59 \text{ }^\circ\text{C})}{\eta(T)} \exp \left[- \frac{\Delta G_0^\ddagger + \Delta G_1^\ddagger \cdot (T - T_m) + \Delta G_2^\ddagger \cdot (T - T_m)^2}{RT} \right], \quad (\text{S11})$$

in which the temperature-dependent free energy of activation ΔG_f^\ddagger is represented as a second order Taylor series expansion about the melting temperature T_m , and ΔG_0^\ddagger , ΔG_1^\ddagger , and ΔG_2^\ddagger are parameters of the fit (these parameters are given for each protein in Figures S44-S55). The pre-exponential term in equation S11 represents the viscosity-corrected frequency ν of the characteristic diffusional folding motion at the barrier^{13,14} (at 59 °C, $\nu = 5 \times 10^5 \text{ s}^{-1}$).¹⁵ $\eta(59 \text{ }^\circ\text{C})$ is the solvent viscosity at 59 °C and $\eta(T)$ is the solvent viscosity at temperature T, both calculated with equation S12:

$$\eta(T) = A \cdot 10^{\frac{B}{T-C}}, \quad (\text{S12})$$

where $A = 2.41 \times 10^5 \text{ Pa}\cdot\text{s}$, $B = 247.8 \text{ K}$, and $C = 140 \text{ K}$.¹⁶

The parameters for equations S4 and S11 were used to calculate the folding and unfolding rates for proteins **17**, **17g**, **18**, **18g**, **19**, **19g**, **20**, **20g**, **Pin WW**, **30g**, **33**, and **33g** shown in Table 2 of the main text, along with the folding and unfolding rate ratios shown in Table 3 of the main text. The low thermal stability of **14**, **14g**, **23**, **23g**, and **26g** precluded kinetic measurements on these proteins.

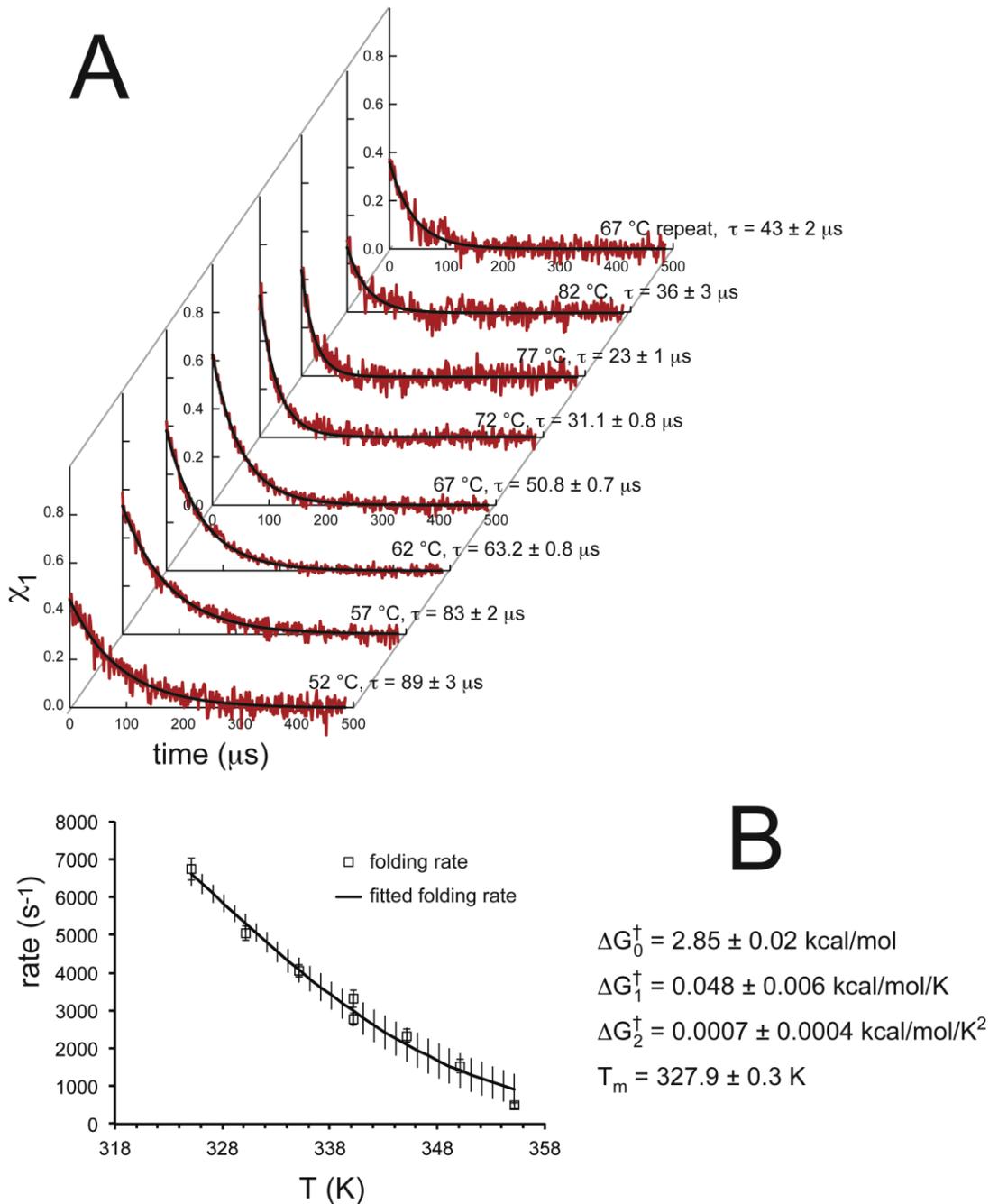


Figure S44. (A) Kinetic relaxation data for Pin WW domain protein **17** (which has Asn at position 17) at several temperatures following temperature jumps of $\sim 12^\circ\text{C}$. Red lines show the change in Trp fluorescence for **17** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW domain protein **17** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

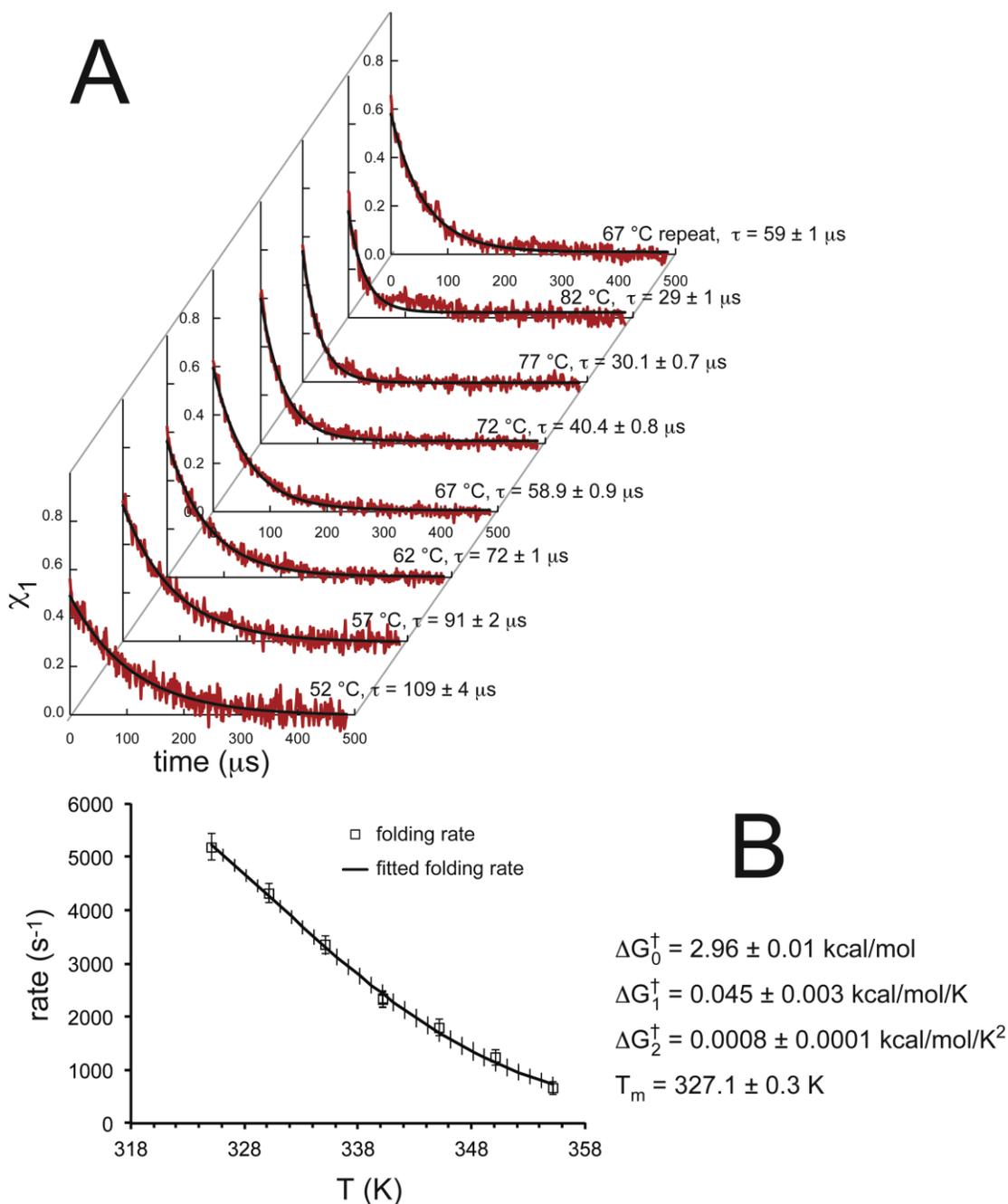


Figure S45. (A) Kinetic relaxation data for Pin WW domain glycoprotein **17g** (which has Asn-linked GlcNAc at position 17) at several temperatures following temperature jumps of ~ 12 °C. Red lines show the change in Trp fluorescence for **17g** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW domain glycoprotein **17g** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

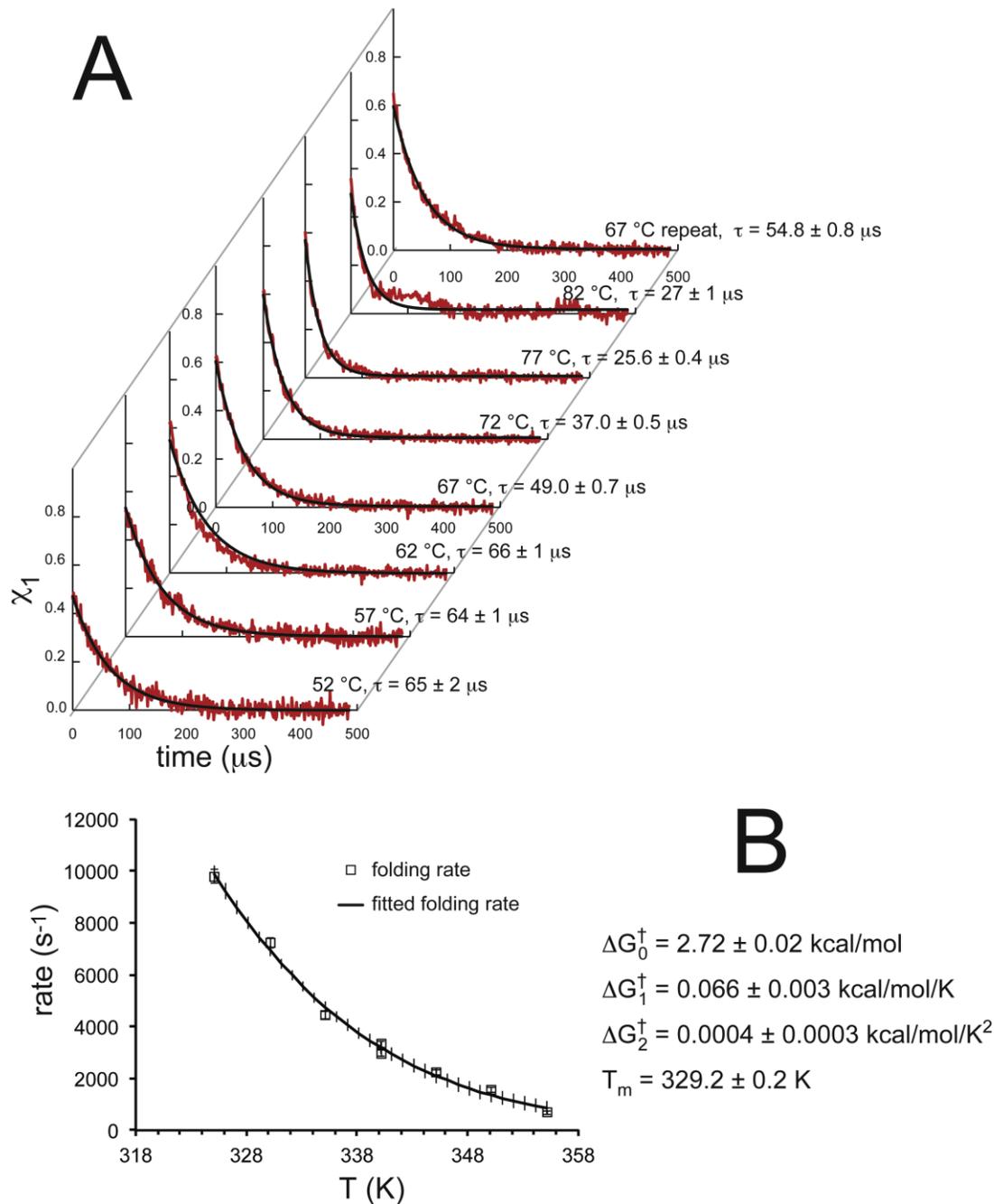


Figure S46. (A) Kinetic relaxation data for Pin WW domain protein **18** (which has Asn at position 18) at several temperatures following temperature jumps of $\sim 12^\circ\text{C}$. Red lines show the change in Trp fluorescence for **18** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW domain protein **18** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

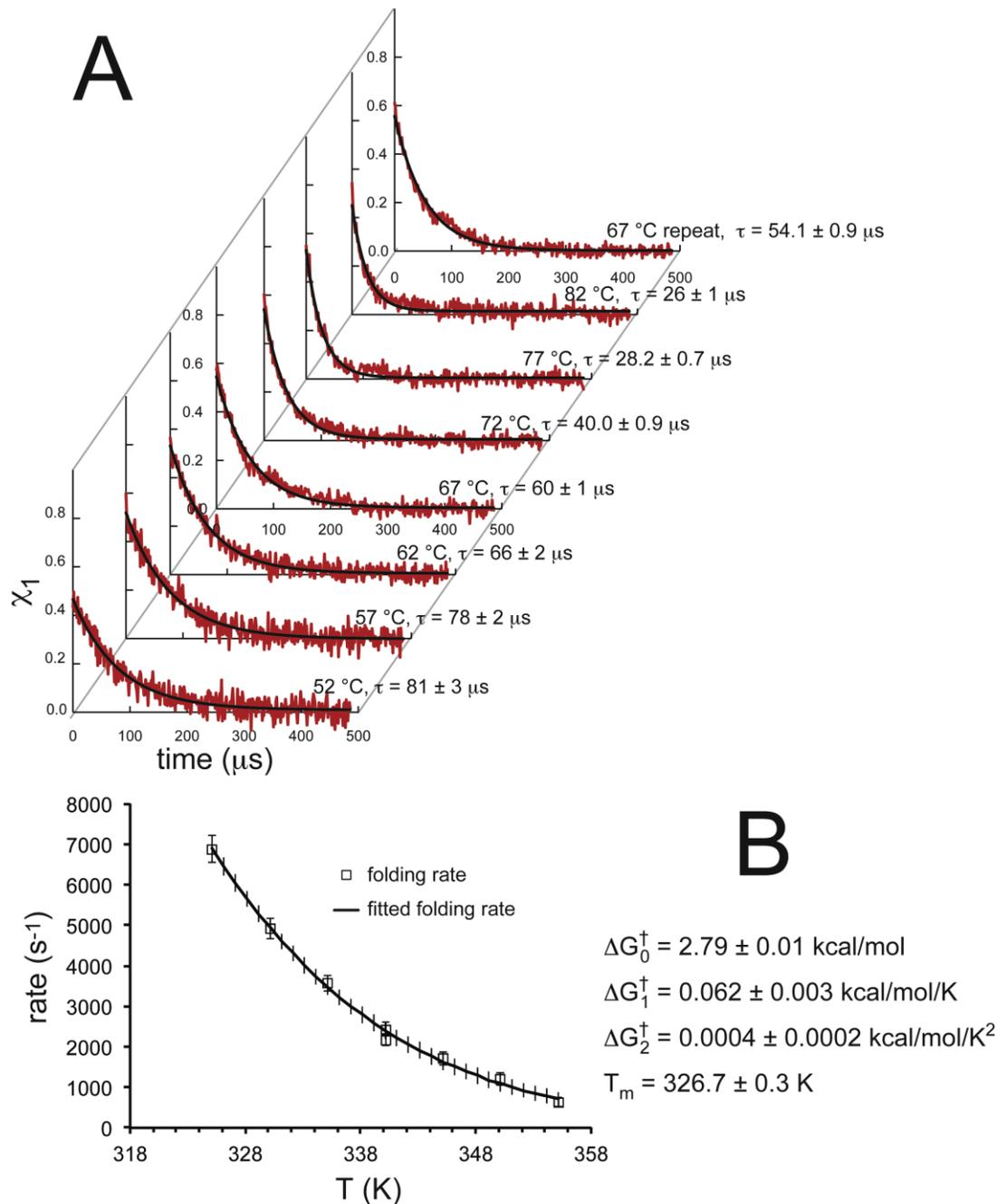


Figure S47. (A) Kinetic relaxation data for Pin WW domain glycoprotein **18g** (which has Asn-linked GlcNAc at position 18) several temperatures following temperature jumps of $\sim 12^\circ\text{C}$. Red lines show the change in Trp fluorescence for **18g** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW domain glycoprotein **18g** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

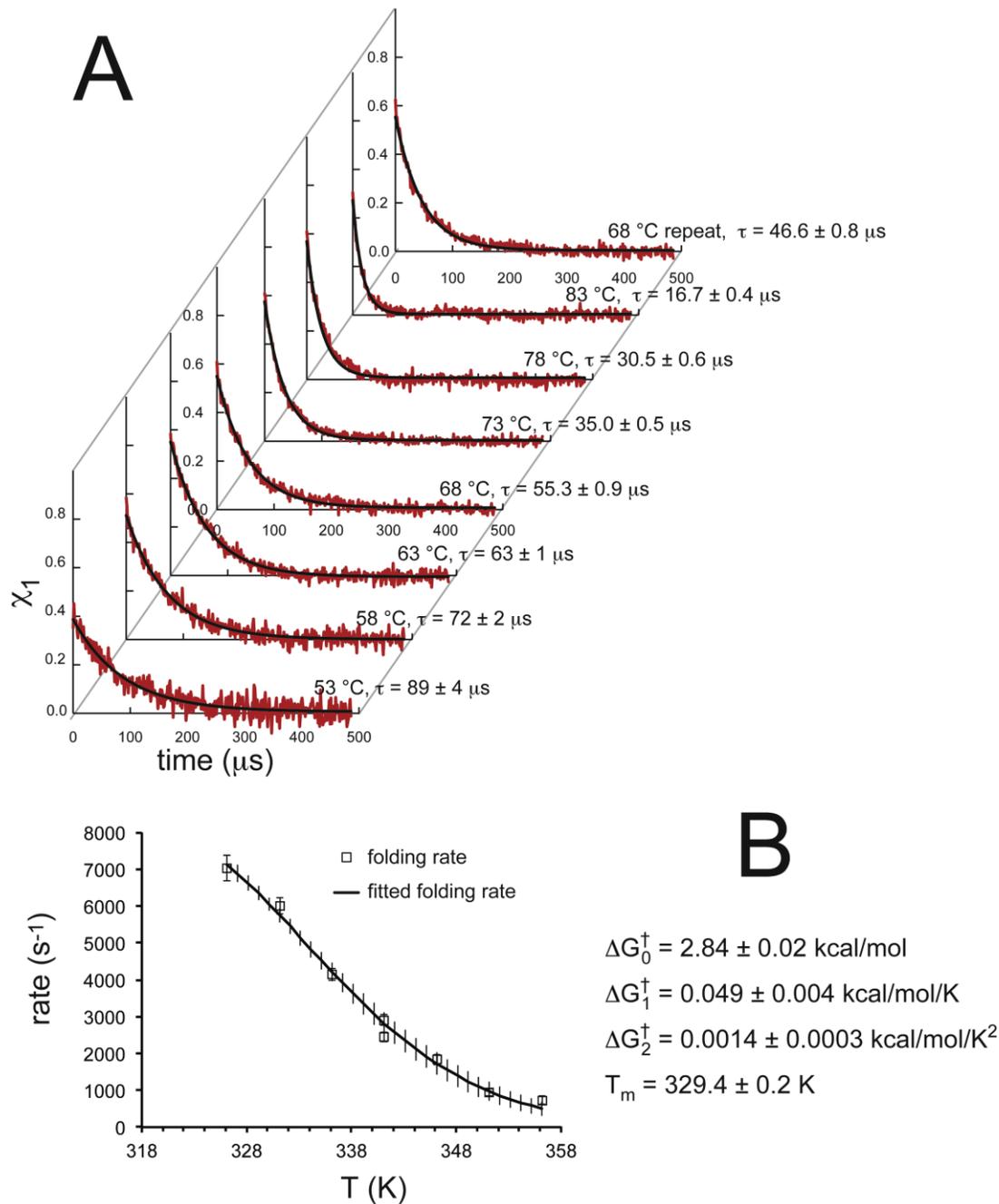


Figure S48. (A) Kinetic relaxation data for Pin WW domain protein **19** (which has Asn at position 19) at several temperatures following temperature jumps of $\sim 12^\circ\text{C}$. Red lines show the change in Trp fluorescence for **19** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW domain protein **19** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

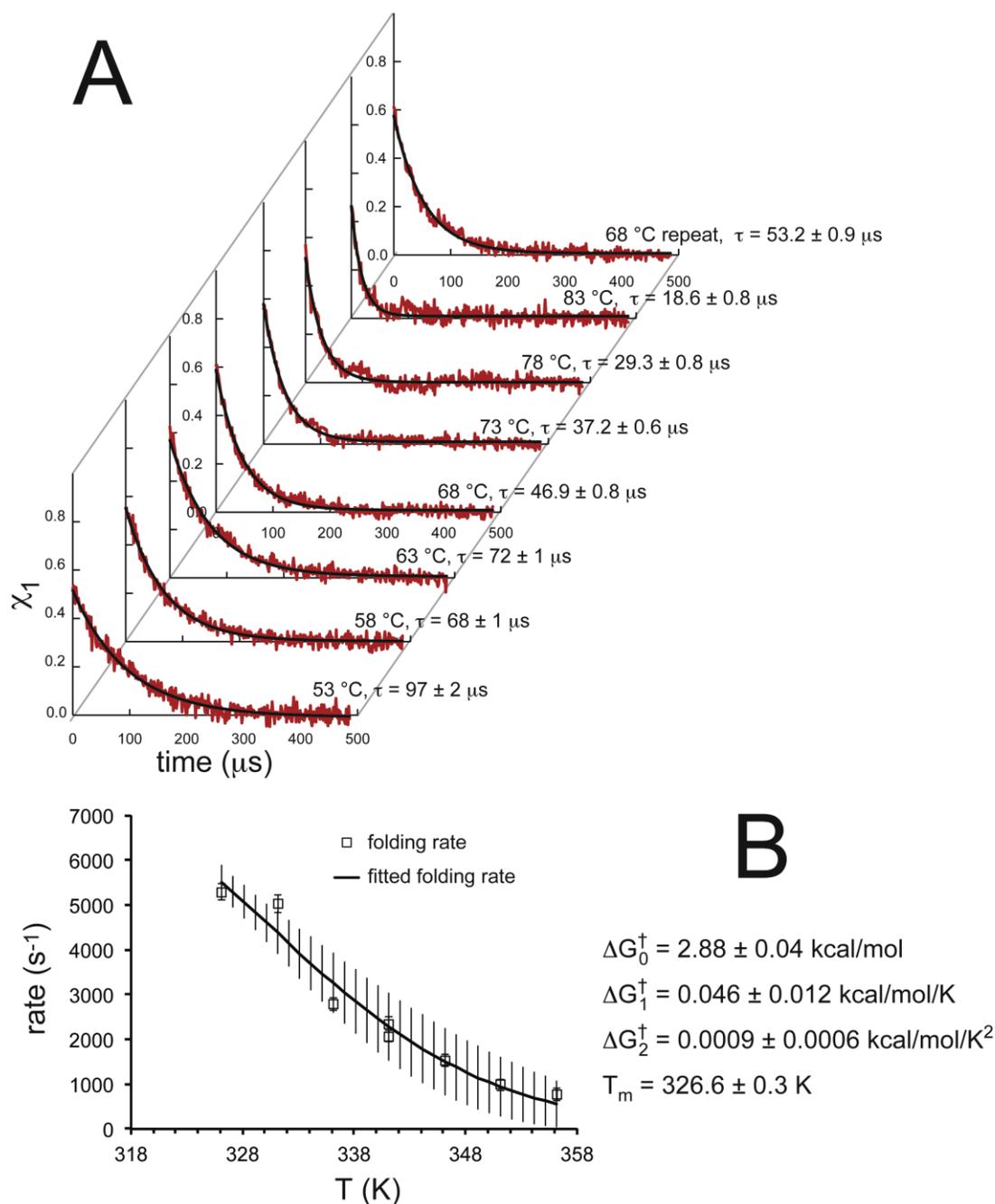


Figure S49. (A) Kinetic relaxation data for Pin WW domain glycoprotein **19g** (which has an Asn-linked GlcNAc at position 19) at several temperatures following temperature jumps of ~ 12 °C. Red lines show the change in Trp fluorescence for **19g** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW domain glycoprotein **19g** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

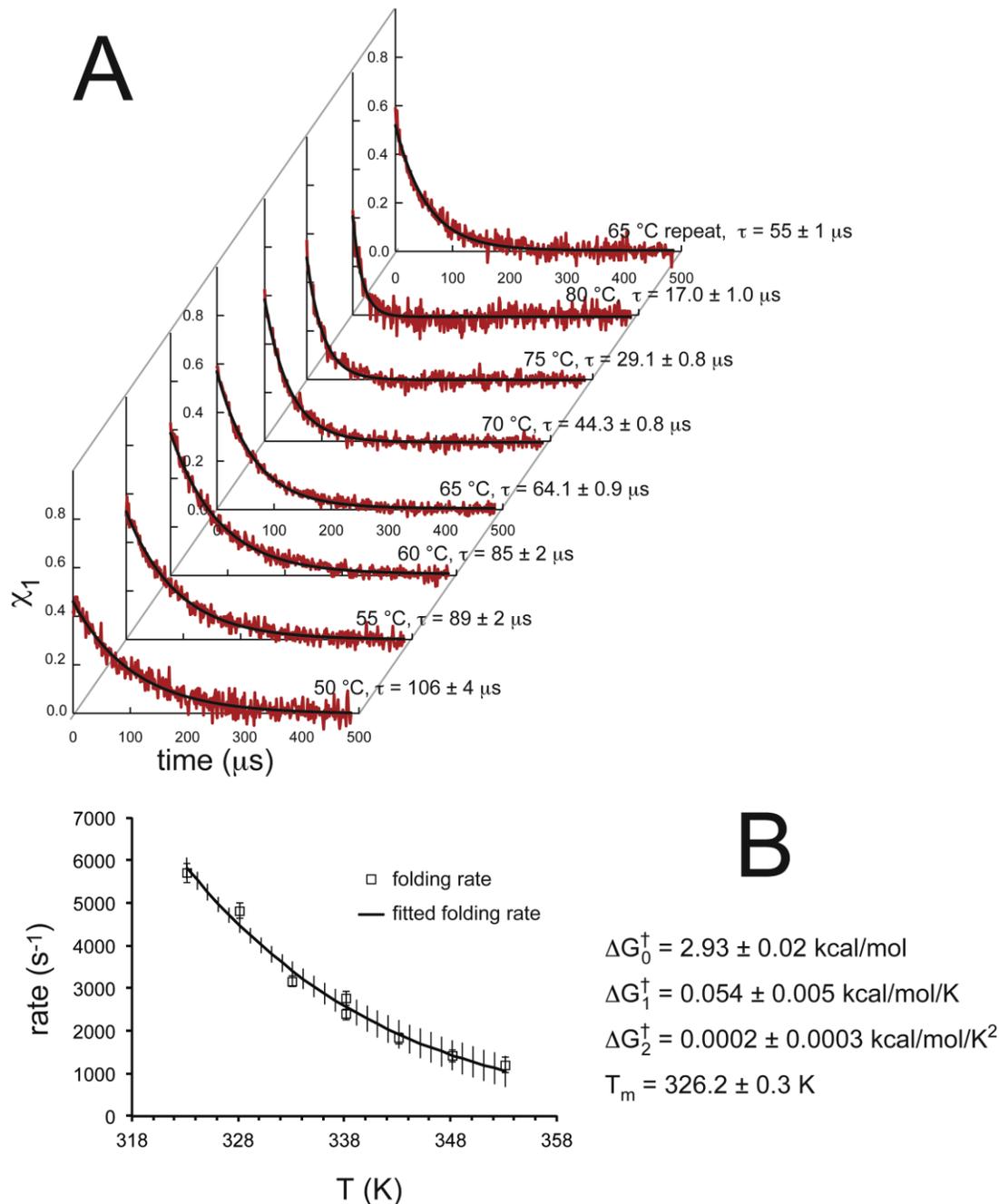


Figure S50. (A) Kinetic relaxation data for Pin WW domain protein **20** (which has Asn at position 20) at several temperatures following temperature jumps of $\sim 12 \text{ }^\circ\text{C}$. Red lines show the change in Trp fluorescence for **20** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW domain protein **20** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

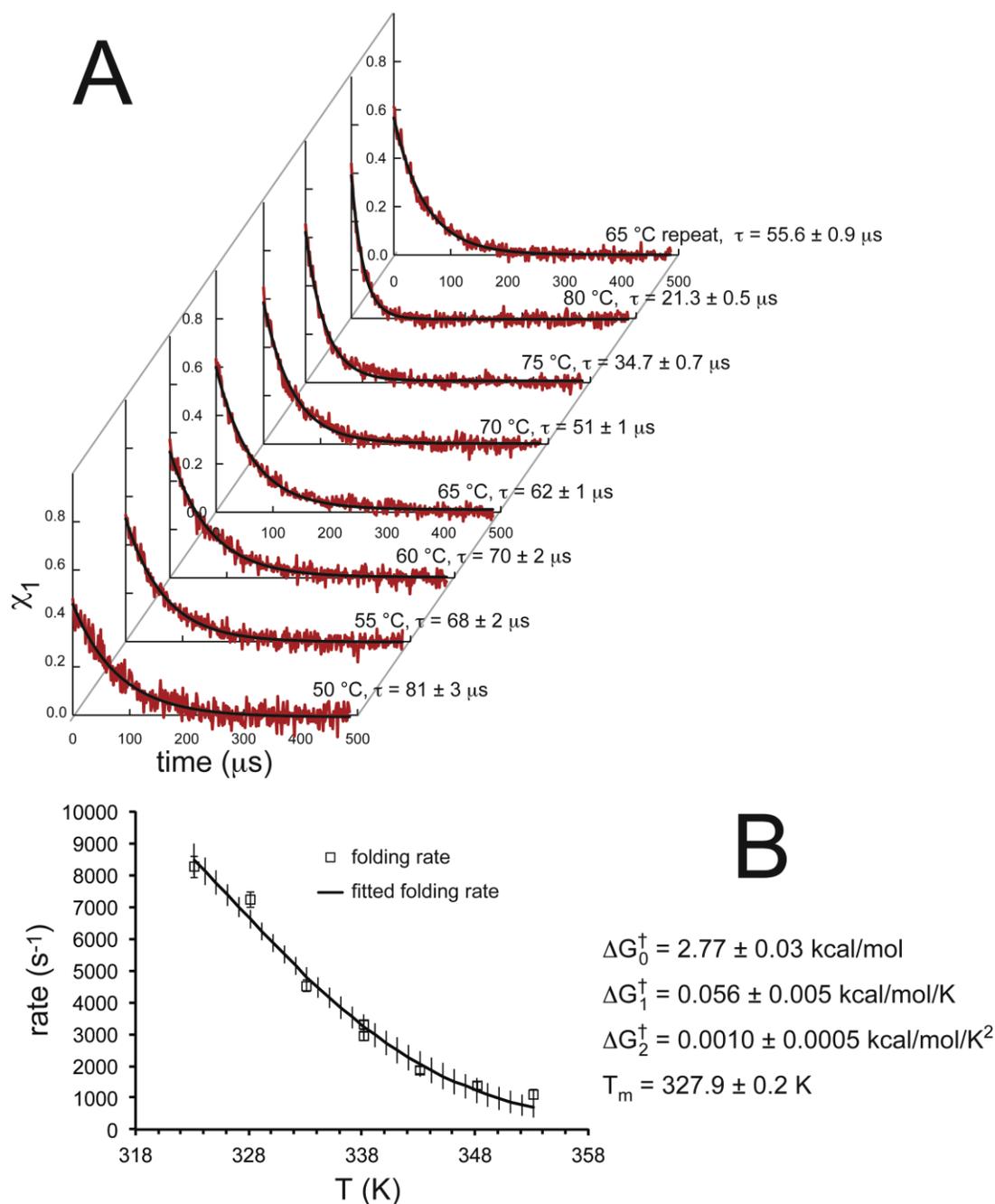


Figure S51. (A) Kinetic relaxation data for Pin WW domain glycoprotein **20g** (which has Asn-linked GlcNAc at position 20) at several temperatures following temperature jumps of $\sim 12^\circ\text{C}$. Red lines show the change in Trp fluorescence for **20g** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW domain glycoprotein **20g** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

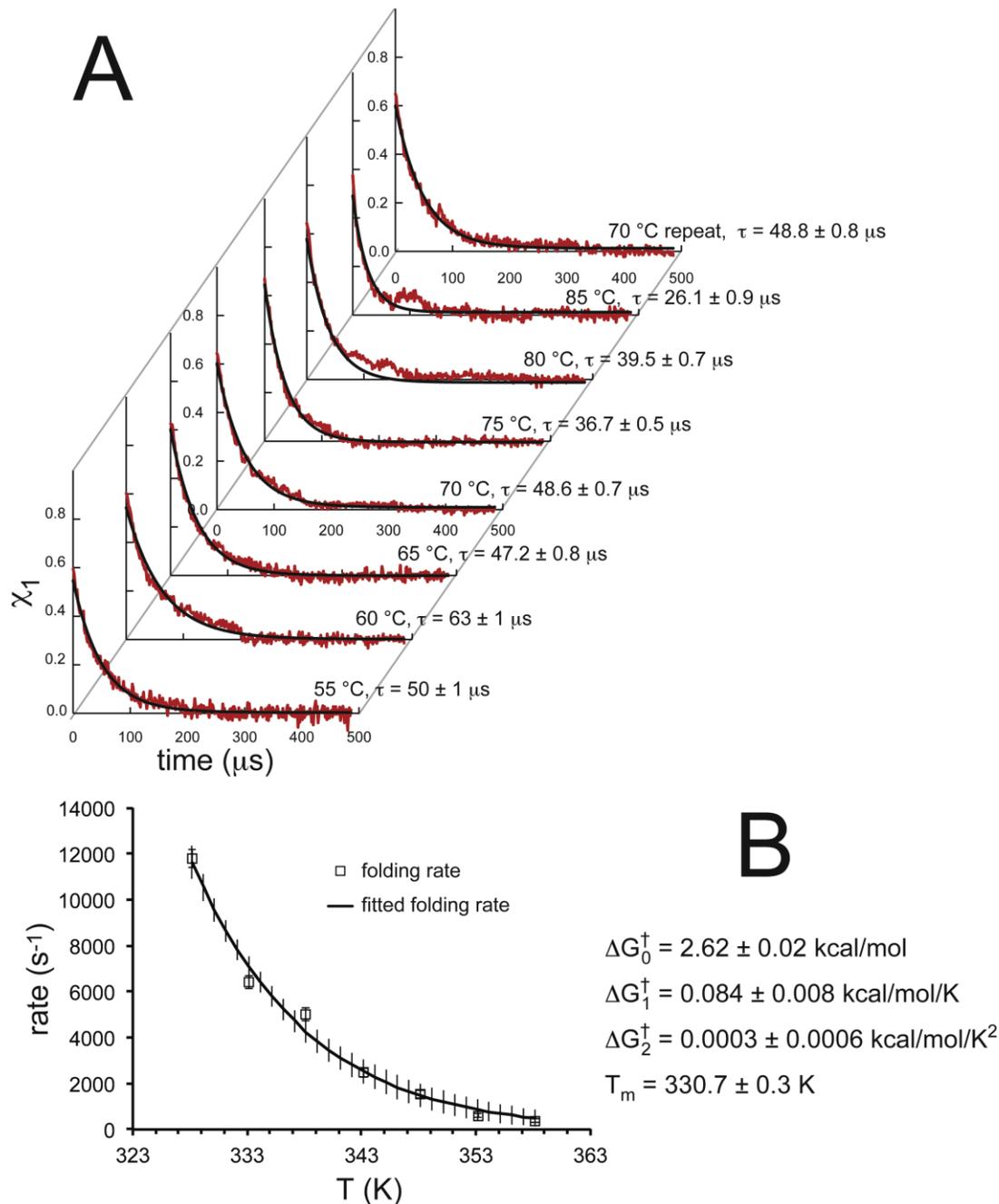


Figure S52. (A) Kinetic relaxation data for **Pin WW** (the wild-type Pin WW domain sequence, which has Asn residues at positions 26 and 30) at several temperatures following temperature jumps of $\sim 12^\circ\text{C}$. Red lines show the change in Trp fluorescence for **Pin WW** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for **Pin WW** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

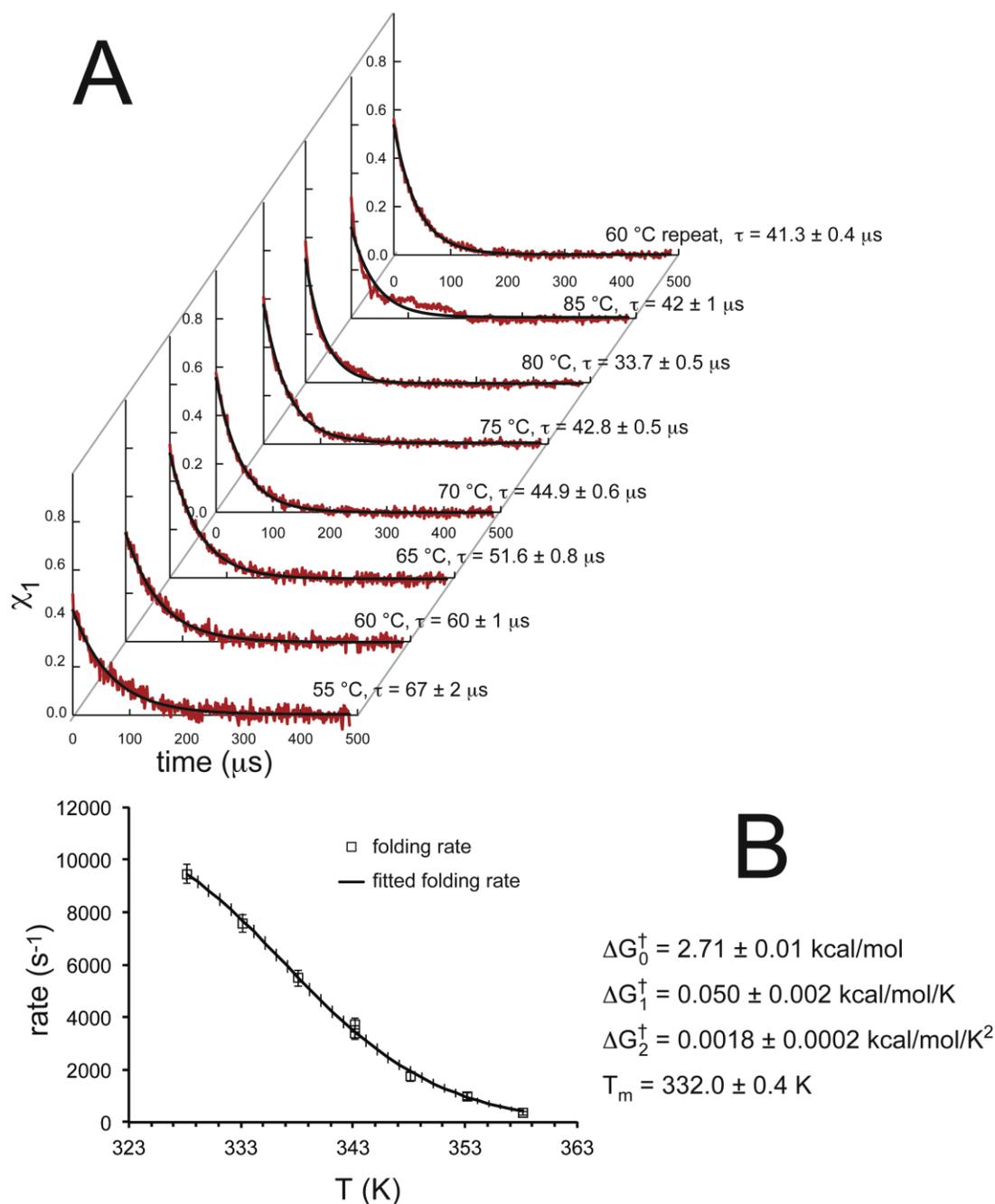


Figure S53. (A) Kinetic relaxation data for Pin WW domain glycoprotein **30g** (which has Asn-linked GlcNAc at position 30) at several temperatures following temperature jumps of $\sim 12^\circ\text{C}$. Red lines show the change in Trp fluorescence for **30g** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW domain glycoprotein **30g** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

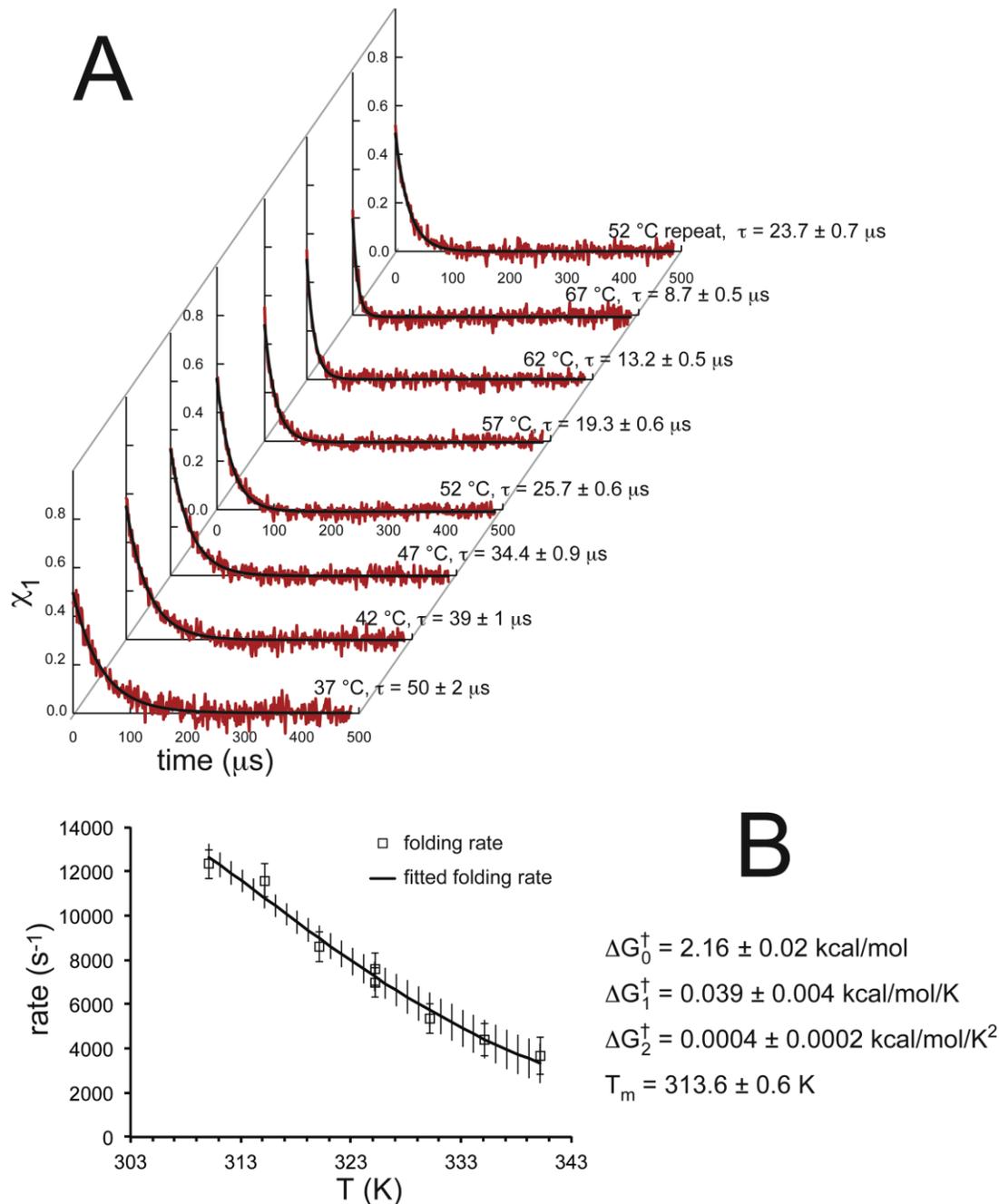


Figure S54. (A) Kinetic relaxation data for Pin WW domain protein **33** (which has Asn at position 33) at several temperatures following temperature jumps of $\sim 12^\circ\text{C}$. Red lines show the change in Trp fluorescence for **33** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW domain protein **33** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

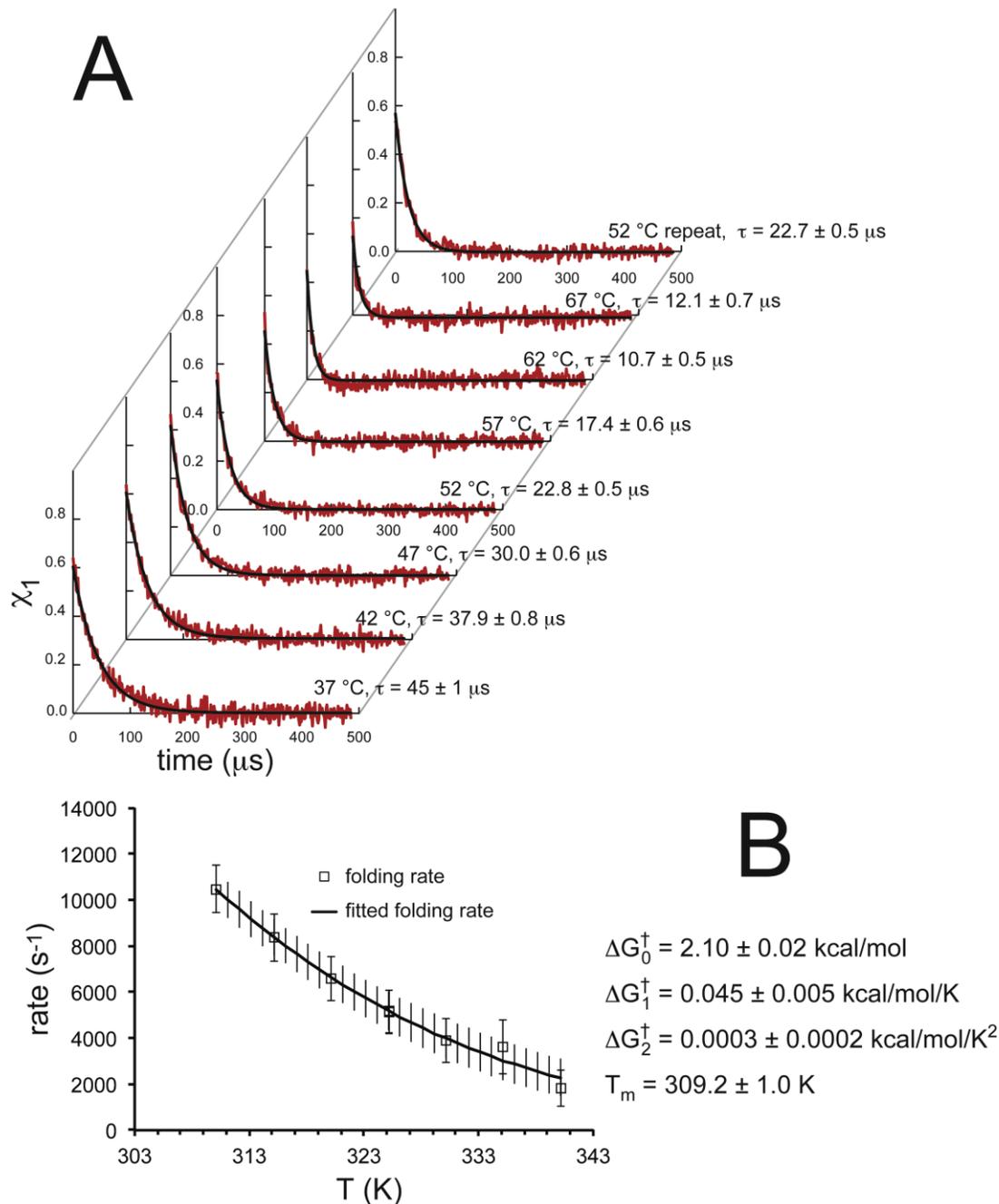


Figure S55. (A) Kinetic relaxation data for Pin WW domain glycoprotein **33g** (which has Asn-linked GlcNAc at position 33) at several temperatures following temperature jumps of $\sim 12^\circ\text{C}$. Red lines show the change in Trp fluorescence for **33g** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S5 and S6) vs. time. Relaxation decay traces at each temperature represent the average of at least ten individual temperature-jump measurements. Black lines show the fit of the data to a monoexponential function (see equation S7) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW domain glycoprotein **33g** as a function of temperature. The black solid line represents the fit of the folding rates to equations S11 and S12, using the indicated parameters. The red solid line represents the fitted unfolding rate (derived from the fit equations for folding rate and thermal denaturation). Capped error bars represent standard error in the folding rate data; uncapped error bars represent standard error in the fits.

Native Topology Model Calculations

In this study, we applied a native topology-based model (Go model) to study the folding of glycoproteins. In this model, each amino acid and each sugar ring is represented by all the heavy atoms. All local, secondary, and tertiary native contacts between amino acids are represented by the Lennard-Jones potential without any discrimination between the various chemical types of the interactions. The Hamiltonian of the system and its parameters can be found elsewhere^{17,18}.

The protein is introduced by five terms (for bonds, angles, torsion angles, the Lennard-Jones term for native interactions, and the excluded volume term for non-native interactions) that define the properties of its folding. The glycan conjugate is similarly introduced, however we neglect possible favorable interactions between the glycans and the proteins.

The simulations were performed using the GROMACS software package¹⁸. Multiple trajectories were simulated using the Langevin equation with friction constant of 0.5 ps⁻¹. Trajectories were collected to look for numerous unfolding/folding transitions at various temperatures. The trajectories were analyzed using the weighted histogram analysis method (WHAM)¹⁹ to study the folding thermodynamics. The fraction of native contacts was taken as a reaction coordinate. Comparing between the calculated thermodynamic features of proteins of different sizes (i.e., with and without conjugated glycan) might be problematic because the energy of the native structure depends on the size of the protein and its exact topology (the terms of bond and dihedral angles). To overcome this limitation, we used only the Lennard-Jones energy component as was calculated by GROMACS, which corresponds only to the pin WW protein moiety without the conjugated glycans (any stabilizing non-bonded interaction between the protein and the glycan were also excluded). The set of native non-bonded contacts and their

distances were calculated using SHADOW algorithm¹⁸. The Asn mutations were built in Pymol using the NMR structure of Pin WW²⁰ and the amino acid substitution was followed by a short energy minimization of the protein with constraints on the heavy atoms where only the atoms of the residue surrounding the mutation site were allowed to move. This procedure was repeated for each of the NMR models and the models with the largest number of native contact were selected for glycosylation and further study. The GlcNAc-conjugated variants were built using the Glyprot web server²¹. We created the GROMACS topology file for both glycosylated and unglycosylated proteins and kept in each file only the intra protein Lennard-Jones interactions and removed the sugar-protein interactions.

Water is not explicitly represented in our calculations, yet the water molecules that are elemental for maintaining the structure are indirectly represented in native topology based models that ensure correct folding. Moreover, the random forces applied by the water molecules on the solute are introduced by the Langevin dynamics in the simulations. The water may additionally affect the thermodynamics of biomolecules through the formation of hydrogen bonds with the protein backbone and with some side chains. In calculating the thermodynamic properties of the protein, the water is also pivotal due to the often large solvation entropy that is released upon folding. In the current study, we treat the glycosylation as a perturbation to the protein and assume that the solvation entropy is identical for the glycosylated and nonglycosylated proteins. Accordingly, in this study we focus on the effect of glycosylation on the configuration entropy.

Potential of Mean Force

The potential of mean force (PMF) diagram of each mutant and its GlcNAc conjugated variant is represented in the equilibrium temperature between folded and unfolded state (Figure S56). The curves were shifted along the y axis to simplify presentation so that the exact value of the PMF is meaningless. The T_F of each variant was extracted from the heat capacity diagrams. The changes in the T_F as resulted by the glycosylation are presented on the structure of the Pin WW domain (Figure S57). The ribbon diagram is color coded such that the red represents stabilization and blue represents destabilization.

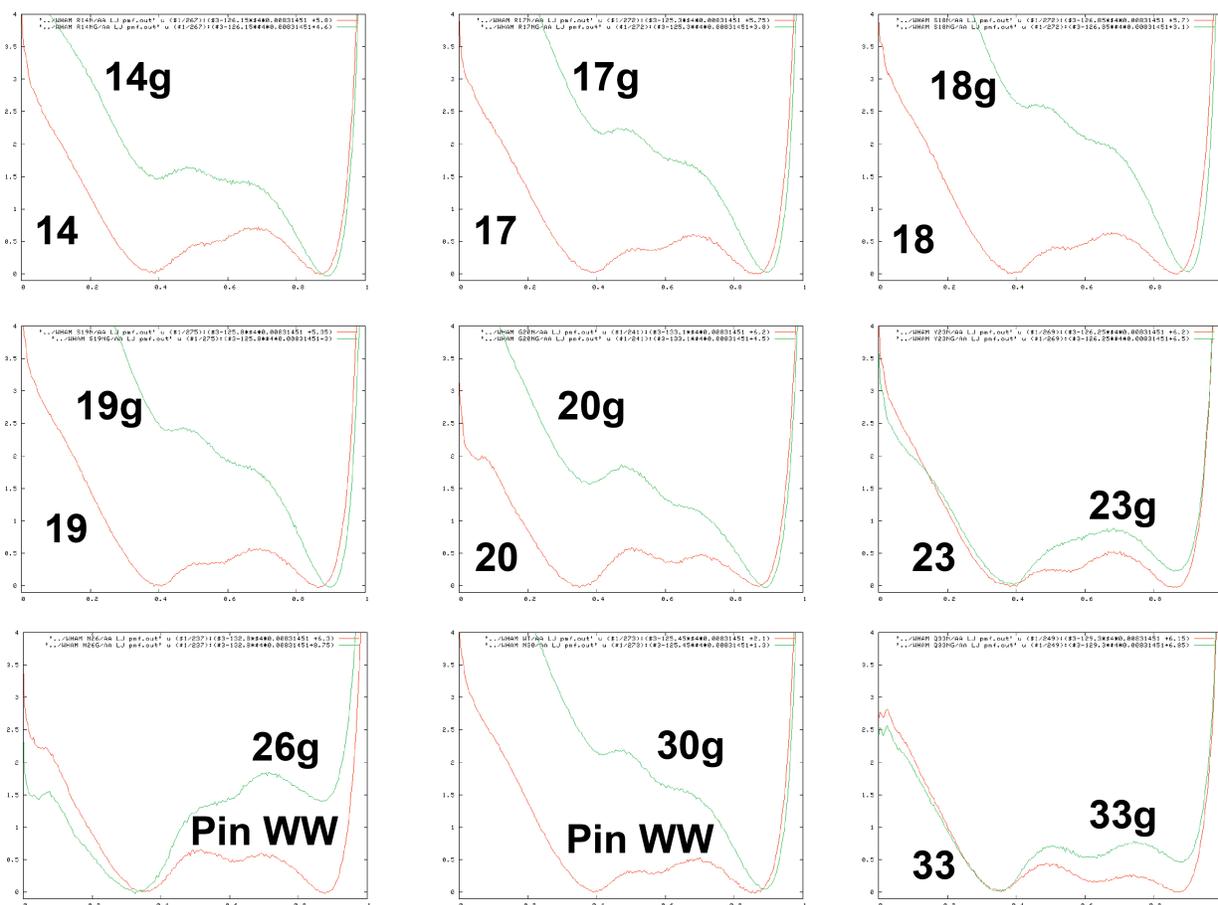


Figure S56. The potential of mean force (PMF) of the Asn (red) and Asn-linked GlcNAc (green) Pin WW variants at each of the nine positions studied experimentally. The PMF data are plotted at the folding temperature of the corresponding Asn variants.

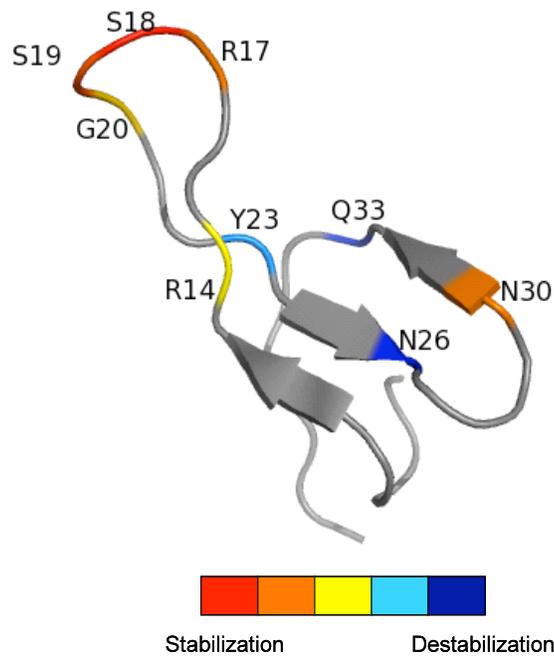


Figure S57. A graphical representation of the effect of glycosylation (at each of the nine selected positions) on the thermodynamic stability of the Pin WW domain, as obtained from computational studies.

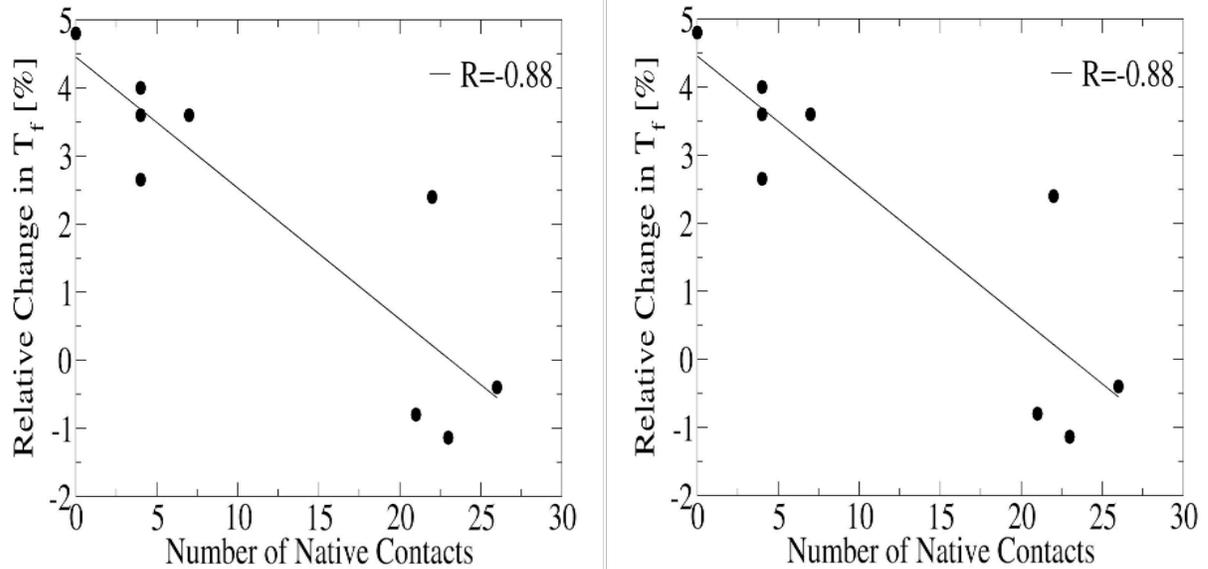


Figure S59. Correlation between the change in the stability of Pin WW due to glycosylation at each of the nine positions and the number of native contacts involving Asn at each position.

Folding/Unfolding rate

In order to determine the folding (and unfolding) rates as seen in our simulations, we calculated the mean passage time (MPT) during which the protein is in its unfolded (and folded) state. This method assumes that the folding rate is inversely correlated to the time during which proteins remain unfolded between two successive transition events. For faster folding proteins, MPT tends to be shorter. This was achieved by undertaking simulations at T_F of each unglycosylated variant as a basis for comparison between the glycosylated and unglycosylated mutants. Since we use a structure based model, the time scale used in our simulations has no direct physical meaning. Nevertheless, the MPT does correlate with the folding rate²², although it does not represent the absolute folding rate. This procedure is similar to an earlier procedure used by other researchers, in which the rates are extracted from many shorter simulations, taking into consideration the first passage time²³.

Ca Model

Preliminary results with the low resolution coarse-grained Ca Go model (which was used to study the effect of glycosylation of Src-SH3 domain^{24,25}) exhibited a very small change in the stability of the protein upon glycosylation. The changes in stability were of the range of $\pm 0.25\%$ from the T_F of the non glycosylated variant. Following this we reduced the graining of the simulation model and transformed to the all-atom Go model¹⁸. The Ca simulations can not capture the influence of the Asn mutation, because each residue is represented only by the Ca bead. In this regard, all the non-glycosylated variants are identical, and this may not capture the effect of glycosylation.

References

- (1) Meldal, M.; Bock, K. *Tetrahedron Lett.* **1990**, *31*, 6987–6990.
- (2) Otvos, L., Jr.; Urge, L.; Hollosi, M.; Wroblewski, K.; Graczyk, G.; Fasman, G. D.; Thurin, J. *Tetrahedron Lett.* **1990**, *31*, 5889–5892.
- (3) Bennett, C. S.; Dean, S. M.; Payne, R. J.; Ficht, S.; Brik, A.; Wong, C.-H. *J. Am. Chem. Soc.* **2008**, *130*, 11945–11952.
- (4) Ficht, S.; Payne, R. J.; Guy, R. T.; Wong, C.-H. *Chem. Eur. J.* **2008**, *14*, 3620–3629.
- (5) Edelhoich, H. *Biochemistry* **1967**, *6*, 1948–1954.
- (6) Ballew, R. M.; Sabelko, J.; Reiner, C.; Gruebele, M. *Rev. Sci. Instrum.* **1996**, *67*, 3694–3699.
- (7) Ballew, R. M.; Sabelko, J.; Gruebele, M. *Proc. Natl. Acad. Sci USA* **1996**, *93*, 5759–5764.
- (8) Ervin, J.; Sabelko, J.; Gruebele, M. *J. Photochem. Photobiol. sect. B* **2000**, *54*, 1–15.
- (9) Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2001**, *311*, 373–393.
- (10) Kramers, H. A. *Physica* **1940**, *7*, 284.
- (11) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci USA* **2000**, *97*, 7220–7225.
- (12) Hänggi, P.; Talkner, P.; Borovec, M. *Rev. Mod. Phys.* **1990**, *62*, 251–341.
- (13) Bieri, O.; Wirz, J.; Hellrung, B.; Schutokowski, M.; Drewello, M.; Kiefhaber, T. *Proc. Natl. Acad. Sci USA* **1999**, *96*, 9597–9601.
- (14) Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Science* **1992**, *256*, 1796–1798.
- (15) Hanson, S. R.; Culyba, E. K.; Hsu, T.-L.; Wong, C.-H.; Kelly, J. W.; Powers, E. T. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 3131–3136.
- (16) Weast, R. C. *CRC Handbook of Chemistry and Physics*; CRC Press: Boca Raton, 1982.
- (17) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Proteins* **2009**, *75*, 430–41.
- (18) Noel, J. K.; Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Nucleic Acids Res, In press*.
- (19) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.
- (20) Kowalski, J. A.; Kiu, K.; Kelly, J. W. *Biopolymers* **2002**, *63*, 111–121.
- (21) Bohne-Lang, A.; von der Lieth, C. W. *Nucleic Acids Research* **2005**, *33*, W214–W219.
- (22) Hagai, T.; Levy, Y. *J Am Chem Soc* **2008**, *130*, 14253–62.
- (23) Koga, N.; Takada, S. *J Mol Biol* **2001**, *313*, 171–80.
- (24) Shental-Bechor, D.; Levy, Y. *Proc Natl Acad Sci U S A* **2008**, *105*, 8256–61.
- (25) Shental-Bechor, D.; Levy, Y. *Curr Opin Struct Biol* **2009**.