Supporting Online Material for "Protein Native State Stabilization by Placing Aromatic Side Chains in N-Glycosylated Reverse Turns"

Elizabeth K. Culyba,[1,3*] Joshua L. Price,[1,3*] Sarah R. Hanson,[1,3] Apratim Dhar,[4] Chi-Huey Wong,[1,3] Martin Gruebele,[4] Evan T. Powers,[1,2,3†] and Jeffery W. Kelly[1,2,3†]

[1]Departments of Chemistry and [2]Molecular and Experimental Medicine and [3]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037; [4]Center for Biophysics and Computational Biology and Departments of Chemistry and Physics, University of Illinois, Urbana, Illinois 61801

Table of Contents

Discussion

Enhanced aromatic sequons: *i*-position residues other than Phe

The enhanced aromatic sequons discussed in the main text each employed a Phe residue two or three positions prior to a glycosylated Asn and involved a stabilizing interaction between Phe and the N-glycan (and possibly Thr). A survey of the PDB suggests that Tyr, Trp, and His might also be able to participate in stabilizing interactions with N-glycans (Fig. S1)

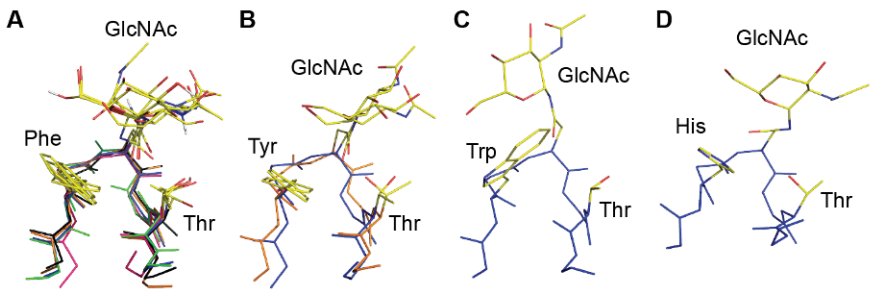


Figure S1. (A-D) Glycosylated type I β-bulge turns from the PDB with an *i*-position (A) Phe (green: HsCD2ad, 1GYA (*1*); blue: 1G82 (*2*); orange: 1FJR (*3*); red: 1ICF (*4*); and black: 3DMK (*5*)), (B) Tyr (blue: 1JS8 (*6*); orange: 3DBX (*7*)), (C) Trp (blue: 1CE7 (*8*)), and (D) His (blue: 2W59 (*9*)). The *i* , *i*+2, and *i*+4 positions are highlighted in yellow. Structures rendered in Pymol.

The thermodynamic consequences of N-glycosylation and the Phe mutation in RnCD2*, AcyP2*, and Pin WW

As noted in the main text, N-glycosylation and the Phe mutation alone do not consistently improve the stability of RnCD2*, AcyP2*, and Pin WW as isolated modifications (see Table 1 in the main text). Without Phe, glycosylation has little impact on the stability of **RnCD2*** ($\Delta\Delta G_f$ = -0.6 ± 0.6 kcal mol^{-1}; cf. **g-RnCD2*** and **RnCD2*** in Table 1). The same is true for **AcyP2***

($\Delta\Delta G_f$ = 0.5 ± 0.8 kcal mol^{-1}; cf. **g-AcyP2*** and **AcyP2*** in Table 1) and for **WW-T** ($\Delta\Delta G_f$ = 0.08 ± 0.08 kcal mol^{-1}; cf. **g-WW-T** and **WW-T** in Table 1).

g-AcyP2*-F is more stable than **g-AcyP2*** ($\Delta\Delta G_f$ = -0.9 ± 0.9 kcal mol^{-1}), suggesting that changing Thr43 to Phe stabilizes glycosylated AcyP2, though the large standard error for this value may indicate that the increase in stability is small. Similarly, **g-WW-F,T** is more stable than **g-WW-T** ($\Delta\Delta G_f$ = -0.29 ± 0.08 kcal mol^{-1}). Adding Phe to **g-RnCD2*** appears not to increase stability significantly ($\Delta\Delta G_f$ = -0.3 ± 0.6 kcal mol^{-1}; cf. **g-RnCD2*-F** and **g-RnCD2*** in Table 1). However, evidence for a stabilizing interaction between Phe and the N-glycan in **g-RnCD2*-F** is strong: glycosylation strongly increases the stability of **RnCD2*-F**, which contains Phe ($\Delta\Delta G_f$ = -1.8 ± 0.4 kcal mol^{-1}; cf. **g-RnCD2*-F** and **RnCD2*-F** in Table 1).

Without glycosylation, changing wild-type residues to Phe destabilizes **RnCD2*** ($\Delta\Delta G_f$ = 0.9 ± 0.5 kcal mol^{-1}; cf. **RnCD2*-F** and **RnCD2*** in Table 1), **AcyP2*** ($\Delta\Delta G_f$ = 1.6 ± 0.9 kcal mol^{-1}; cf. **AcyP2*-F** and **AcyP2*** in Table 1), and **WW-T** ($\Delta\Delta G_f$ = 0.50 ± 0.08 kcal mol^{-1}; cf. **WW-F,T** and **WW-T** in Table 1). The origin of this destabilizing effect is unclear, but it is clearly offset by the stabilizing tripartite interaction between Phe, the Asn(N-glycan), and Thr. As described above, the effect of the Phe mutation changes in the presence of glycosylation: Phe-containing **g-AcyP2*-F** and **g-WW-F,T** are stabilized relative to wild-type-residue-containing **g-AcyP2*** and **g-WW-T**, respectively, and the Phe mutation is much less destabilizing to **g-RnCD2*** than to non-glycosylated **RnCD2***.

Materials and methods

Unless otherwise noted, chemicals and products were purchased from Fisher Scientific or Sigma-Aldrich. Phosphate buffered saline (PBS) was prepared from PBS tablets (SIGMA P-4417) and maintained at pH 7.2 with 0.5 mM TCEP and 0.01% sodium azide. 50 mM acetate

buffer was prepared from a 4× solution made from 4× solutions of acetic acid (Acros Organic 124040025) and sodium acetate trihydrate (SIGMA 236500) to achieve a final pH of 5.5. Acetate buffer was also prepared with 0.5 mM TCEP and 0.01% sodium azide. All buffer solutions were filtered (Millipore 0.2μM). Protein was concentrated using Amicon centrifugation devices, MWCO 3 kDa (Millipore). Final concentrations of RnCD2* and AcyP2* variants were determined by measuring absorbance at 280 nm using calculated extinction coefficients (ExPASy, ProtParam tool). All oligonucleotides for site-directed mutagenesis were purchased from Integrated DNA Technologies (IDT), 25 nmol DNA oligo normalized to 100 μM in IDTE pH 8.0. Wild type RnCD2 and AcyP2 gene constructs were ordered from IDT as miniGenes in pZErO-2 vectors (Kan resistant).

RnCD2* amino acid sequence

The sequence of wild type RnCD2 used for site-directed mutagenesis to produce mutant sequences used:

HHHHHHENLYFQS	DYKDDDDKIEGR	ADCRDSGTVW	GALGHGINLN
IPNFQMTDDI	DEVRWERGSTLV	AEFKRKMKPF	LKSGAFEILA
NGDLKIK<u>N</u>LT	RDDSGTY<u>N</u>VTVY	ST<u>N</u>GTRILDK	ALDLRILEM

The first 6 residues (in blue) are a 6× Histidine-tag, which was included for Nickel affinity chromatography purification. This tag is followed by a 7-residue Tobacco Etch Virus protease cleavage site (TEVs; in red). This tag/protease cleavage site combination is followed by a 9-residue FLAG-tag (in blue), which in turn is followed by the 4-residue Factor Xa cleavage site (Xas; in red) that was included so that all of the tags could be removed from the gene construct (which was done before all experimental measurements were taken). Note that all residues are numbered to correspond to homologous residues in human CD2ad, thus the numbering begins

with 3, i.e. Ala3, and all following residue numbers increase sequentially. It should also be noted that some sequence changes were made to all mutants to ensure that the protein was only glycosylated at the desired position (Asn65) when expressed in Sf9 cells. The wild type RnCD2 sequence contains three glycosylation sequons. The asparagines in these positions, Asn72, Asn82, and Asn89 (underlined), were mutated to glutamine, glutamine, and aspartic acid, respectively. Finally, to confer glycosylation at Asn65 (bold), Asp67 (bold and underlined) was mutated to threonine.

AcyP2* amino acid sequence

The sequence of wild type AcyP2 used for site-directed mutagenesis to produce mutant sequences used:

HHHHHHENLYFQS	DYKDDDDKIEGR	MSTAQSLKSV	DYEVFGRVQG
VCFRMYTEDE	ARKIGVVGWV	KNTS**K**GTVTG	QVQGPEDKVN
SMKSWLSKVG	SPSSRIDRTN	FSNEKTISKL	EYSNFSIRY

The same purification/protease site tag used in the RnCD2* variants was used for AcyP2* variants and as with RnCD2*, the entire tag was removed via Factor Xa cleavage prior to all experiments. Note that the residues are numbered starting with the first residue (Met) after the Factor Xa cleavage site. It should also be noted that some sequence changes were made to all mutants to ensure that the protein was only glycosylated at the desired position (45) when expressed in Sf9 cells. The wild type AcyP2 sequence contains three glycosylation sequons. The serines in these positions, Ser44, Ser82, and Ser96 (underlined), were mutated to alanine. Finally, to confer glycosylation at position 45, Lys45 (bold and underlined) was mutated to Asn.

Visualization of published structural data

RnCD2 structural coordinates were obtained from the PDB (accession code 1HNG). AcyP2 structural coordinates were obtained from the PDB for horse muscle acylphosphatase (accession code 1APS), which shares 94% sequence homology with the human protein. Pin WW coordinates were obtained from the PDB (accession code 1PIN). Structures were visualized and rendered using Pymol.

Molecular biology

All PCR was performed using pfuTurbo polymerase (Stratagene) using recommended conditions. Restriction enzymes were obtained from New England Biolabs and applied as indicated. DNA fragments were ligated with standard conditions supplied for T4 ligase (Roche). Amplified and digested DNA was purified using 1% agarose (molecular biology grade gel prepared in TAE buffer). DNA isolation/purification steps, including genomic isolation, plasmid isolation, restriction digestion clean-up, and PCR purification were performed with Qiagen kits. Clones were transformed, amplified, and maintained in DH5α *E. coli*. All clones were verified for accuracy by sequencing.

Protein purification via FPLC

All FPLC procedures were carried out on an AKTA FPLC from GE Healthcare. HisTrap HP columns (1mL) were run in 25 mM sodium phosphate, 300 mM NaCl, 5-300 mM imidazole, pH 8.0 at a flow rate of 3 mL/min at room temperature. A Superdex 75 10/300 GL column (24 mL) was run in PBS (RnCD2*) or Acetate (AcyP2*) at a flow rate of 0.4 mL/min at room temperature (retention times: RnCD2* with glycan 12.5 min, RnCD2* without glycan 12.75min, AcyP2* with glycan 14.75 min, AcyP2* without glycan 15 min).

Fluorescence spectroscopy

Both RnCD2* and AcyP2* have at least one tryptophan residue buried in the hydrophobic core allowing for an intrinsic fluorescence that depends on the folding status. Fluorescence measurements for RnCD2* and AcyP2 variants were obtained using either a CARY Eclipse (Varian) or an ATF-105 (Aviv) fluorimeter. Measurements were made in quartz cuvettes, at 25 °C, at protein concentrations of 5-30 μg/mL, unless otherwise noted. Fluorescence emission spectra were collected from 315 to 400 nm, following excitation at 280 nm.

Preparation of non-glycosylated RnCD2* and AcyP2* variants

Construction of non-glycosylated variant genes

Genes for non-glycosylated versions of RnCD2* and AcyP2* were subcloned into pT7-7 expression vectors using the PIPES (*10*) method, to create pHisFLAG-RnCD2b and pHisFLAG-AcyP2b sequences with native sequons removed. The total N- to C-protein coding region is: Met-6His-TEVs-FLAGtag-FXas-RnCD2* or AcyP2*.

Site directed mutagenesis:

All mutant variants were engineered from these constructs using quick change site directed mutagenesis.

Expression of non-glycosylated variants in *E. coli* (rich medium)

Bacterial RnCD2* and AcyP2* were expressed as described previously (*11*).

Nickel Affinity Purification

Cells were thawed and re-suspended in an appropriate purification buffer (RnCD2* variants: 25 mM Sodium Phosphate, 300 mM NaCl, 5 mM imidazole, 0.5 mM TCEP, pH8.0; AcyP2* variants: same as above with 25 mM TrisHCl in place of phosphate) in $1/20^{th}$ of the original growth volume. Protease inhibitors (1 tablet/50 mL; Roche EDTA-free) were added. Cells were lysed by sonication. The cell lysate was centrifuged (15,000 rpm, 30 min, 4 °C), the soluble fraction (supernatant) was separated from the insoluble fraction (pellet) and used for Ni-NTA purification. In the case of RnCD2* variants RnCD2*-K and RnCD2*-KF and AcyP2* variant AcyP2*-F, the insoluble fraction was treated with 6 M guanidine hydrochloride (GuHCl) in the appropriate binding buffer and subjected to Ni-NTA purification under denaturing conditions (6M GuHCl).

Superflow Ni-NTA resin was used to affinity purify proteins via the 6×His tag, using conditions described in the Qiagen manual. Denaturing purification was performed similarly with the addition of 6 M GuHCl to all solutions. Eluted fractions were exchanged into Factor Xa cleavage buffer (50mM TrisHCl, 100mM NaCl, pH7.9) and concentrated in Amicon centrifugation devices.

Factor Xa cleavage of N-terminal tags from non-glycosylated proteins

5 mM $CaCl_2$ was added to concentrated protein in 50 mM TrisHCl, 100 mM NaCl, pH 7.9 before Factor Xa (New England Biolabs) treatment (1 μg Factor Xa: 100 μg of RnCD2 or AcyP2). For RnCD2* variants the protease reaction was carried out at 4 °C for 12 h. For AcyP2* variants the protease reaction was carried out at 25 °C for 2h. The cleavage mixture was quenched with 100 μM PMSF and separated and buffer exchanged by FPLC (Superdex 75).

RnCD2* final buffer: PBS, 0.5 mM TCEP, 0.01% sodium azide, pH 7.2. AcyP2* final buffer: 50 mM Acetate, 0.5 mM TCEP, 0.01% sodium azide, pH 5.5. HisFLAG-free RnCD2* ESI MS found: 11578; RnCD2*-K ESI MS found: 11576; RnCD2*-F ESI MS found: 11612; RnCD2*-KF ESI MS found: 11611. HisFLAG-free AcyP2* ESI MS found: 11078; AcyP2*-F ESI MS found: 11124.

Preparation of glycosylated RnCD2* and AcyP2* variants

Cloning for RnCD2 and AcyP2 into an insect shuttle vector

A 5' SacI site (gagctc) and 3' KpnI (ggtacc) site and a preprotrypsin leader sequence (PLS, for secretion into the medium) were designed into both the RnCD2 and AcyP2 genes ordered from IDT. Digestion (SacI and KpnI) and ligation of the products and the insect shuttle vector pFastBac (Invitrogen), yielded clone pPLSHisFLAG-RnCD2i and pPLSHisFLAG-AcyP2i (referred to as RnCD2i and AcyP2i, respectively, from here on out).

Site directed mutagenesis

All mutant variants were engineered from these constructs (in the pFastBac vector) using quick change site directed mutagenesis.

Expression of RnCD2* and AcyP2* in Sf9 (insect) cells

Expression in insect cells was carried out by baculovirus infection as previously described (*11*). After expression, growth media centrifuged (2100×*g*, 20 min) and the conditioned supernatant was collected and 0.2 μM filtered. Protease inhibitors (1 tablet/200 mL; Roche EDTA-free), 0.5 mM TCEP, and 1 mM EDTA were added to the conditioned supernatant.

Ammonium sulfate precipitation of glycosylated variants

The conditioned supernatant was incubated for 1 h with ammonium sulfate (30% wt/vol) at 4 °C with constant stirring and precipitating species were removed. Addition of more ammonium sulfate (80% total wt/vol) to the soluble fraction for 1 h at 4 °C resulted in the precipitation of either RnCD2* or AcyP2* variants. The precipitate was collected by centrifugation (1700 ×*g*, 45 min) followed by vacuum filtration (Whatman Grade 5 qualitative filter paper). Precipitate was stored at -80 °C.

Purification of glycosylated variants by Nickel Affinity Chromotography

Superflow Ni-NTA resin was used to affinity purify proteins via the 6×His tag, using conditions described in the Qiagen manual. Briefly, precipitated protein was resuspended in ¼ of the expression volume of lysis buffer (same as non-glycosylated variants) stirred for 1 h at 4 °C and 0.2 μM filtered. Filtered medium was applied to a gravity Ni-NTA column in appropriate lysis buffer, and washed with 10 column volumes of lysis buffer and 50 column volumes of washing buffer (18 mM imidazole in lysis buffer). Bound protein was removed with 4 column volumes of elution buffer (20 mM TrisHCl, 300 mM imidazole, pH 8.0 for all variants). Alternatively, an FPLC HisTrap HP column (1mL) was used for purification with the same buffer conditions as above. Eluted fractions were exchanged into Concanavilin A (ConA) binding buffer (25 mM TrisHCl, 500 mM NaCl, 1 mM $MnCl_2$, 1 mM $CaCl_2$, pH7.4) and 0.5 mM TCEP and concentrated in Amicon centifugation devices.

Isolation of glycosylated protein by lectin chromotography

Lectin chromatography with Concanavilin A (ConA) was performed on Nickel column eluate with the ConA Glycoprotein Isolation Kit (Pierce) following the protocols described therein. High mannose and paucimannose species were separated from the non-glycosylated protein found in every expression. Elution and wash fractions that contained only glycosylated protein were pooled and exchanged into Factor Xa cleavage buffer (50 mM TrisHCl, 100 mM NaCl, pH7.9).

Factor Xa cleavage of N-terminal tags from glycosylated proteins

5 mM $CaCl_2$ was added to concentrated protein in 50 mM TrisHCl, 100 mM NaCl, pH 7.9 before Factor Xa (New England Biolabs) treatment (1 μg Factor Xa: 100 μg of RnCD2* or AcyP2*). For RnCD2* variants the protease reaction was carried out at 4 °C for 12 h. For AcyP2* variants the protease reaction was carried out at 25 °C for 2 h. The cleavage mixture was quenched with 100 μM PMSF and separated and buffer exchanged by FPLC (Superdex 75). RnCD2* variant final buffer: PBS, 0.5 mM TCEP, 0.01% sodium azide, pH 7.2. AcyP2* variant final buffer: 50 mM acetate, 0.5 mM TCEP, 0.01% sodium azide, pH 5.5. If cleavage was incomplete Nickel-NTA resin was used to remove uncleaved protein.

Western blots of protein expressed in Sf9 cells

Conditioned media from Sf9 insect cells expressing the protein of interest was sampled (20 μL) and the proteins were resolved by SDS-PAGE (10% NUPAGE gel in MES buffer, Invitrogen) and transferred onto a nitrocellulose membrane. The membranes were blocked for 45 min with 4% milk in 0.02% Tween 20/PBS (PBST), probed for 1 h with rabbit-anti-FLAG (Rockland, 1:1000 dilution), and incubated with peroxidase-conjugated goat anti-rabbit IgG

(Thermo Scientific, 1:2800 dilution) for 1 h. Each step was followed by a wash with PBST. Signal was developed with SuperSignal® chemiluminescent substrate (Pierce) and detected by exposure to autoradiography film (Denville Scientific).

ESI-MS characterization.

Liquid chromatography mass spectrometry (LCMS)

LCMS analysis was performed using an Agilent 1100 LC coupled to an Agilent 1100 single quad ESI mass spectrometer. LC was performed with a 4.6mm × 50mm ZORBAX C8 column (Agilent). Table S1 shows observed molecular weights for purified glycosylated RnCD2* and AcyP2* variants (percentages are based on relative abundance).

Table S1. MS characterization of glycosylated RnCD2* and AcyP2* variants.

Variant	$MW_{expected}$	MW_{found}	%	Structure
g-RnCD2*	12956	12956	25	$Man_6GlcNAc_2$
	13119	13118	44	$Man_7GlcNAc_2$
	13282	13282	31	$Man_8GlcNAc_2$
g-RnCD2*-K	12468	12469	6	$Man_3GlcNAc_2$
	12792	12793	13	$Man_5GlcNAc_2$
	12955	12956	22	$Man_6GlcNAc_2$
	13118	13118	43	$Man_7GlcNAc_2$
	13281	13280	16	$Man_8GlcNAc_2$
g-RnCD2*-F	12990	12991	23	$Man_6GlcNAc_2$
	13153	13153	54	$Man_7GlcNAc_2$
	13316	13315	23	$Man_8GlcNAc_2$
g-RnCD2*-KF	12989	12989	21	$Man_6GlcNAc_2$
	13152	13151	33	$Man_7GlcNAc_2$
	13315	13314	46	$Man_8GlcNAc_2$
g-AcyP2*	12117	12116	100	$Man_3GlcNAc_2(Fuc)$
g-AcyP2*-F	12163	12162	100	$Man_3GlcNAc_2(Fuc)$

Characterization of the folding kinetics and thermodynamics of RnCD2* variants

PBS buffer (1×, 0.5 mM TCEP, 0.01% sodium azide, pH 7.2) was made fresh daily from a 10× stock and filtered. Urea and GuHCl solutions were prepared fresh daily in 1×PBS, filtered, and denaturant concentrations were confirmed by index of refraction (IOR). Subsequent dilutions of urea or GuHCl were made with 1×PBS and concentrations were checked by IOR. Constants defined in equations include the universal gas constant (R) and temperature (T). The value of RT at 25 °C was taken to be 0.592 kcal/mol. Data were imported and fit in Mathematica 7 (Wolfram Research). Urea was used as the chaotrope for all RnCD2* variants except for L63F variants. Due to the high thermodynamic stability of the L63F variants and the saturation point of urea at 25 °C, all measurements were also taken in GuHCl solutions for this mutant (variants **g-RnCD2*-F** and **RnCD2*-F**).

Stopped-flow kinetics experiments on RnCD2* variants

Fluorescence measurements related to kinetic experiments were obtained using an AVIV ATF-105 stopped-flow fluorometer for single-mixing experiments. The set-up consisted of two syringes (syringe 1: 1 mL, syringe 2: 2 mL) that allowed up to a 25-fold dilution of the components of syringe 1 with syringe 2, in a minimum of 80 μL, of which the flow cell holds 40 μL. The dead time between start of mixing and acquisition of data was estimated to be 50 ms; in general, only data after the first 200 ms were used for fitting. Excitation was set at 280 nm (bandwidth: 2 nm) and emission was measured at 330 nm (bandwidth: 8 nm). The photomultiplier voltage was set to 1000 V and data was recorded for 20-200 s. For unfolding experiments, the decrease in intensity at 330 nm was monitored after native protein in PBS or low concentrations of urea or GuHCl in syringe 1 was mixed with varying volumes of

concentrated urea or GuHCl solutions in syringe 2. For refolding experiments, the increase in intensity at 330 nm was monitored after denatured protein in a urea or GuHCl solution in syringe 1 was diluted with varying volumes of PBS buffer or low concentrations of urea or GuHCl from syringe 2. All shots of a particular dilution were typically repeated at least 4 times. Continuous irradiation of RnCD2* at 280 nm led to a decrease in fluorescence intensity over time that correlated with the excitation bandwidth, indicating that photobleaching was taking place. The fluorescence intensity at 330 nm (F_{330}) was therefore fit to a double exponential containing a photobleaching (k_{pb}) component and a folding/unfolding component (k_{obs}):

$$F_{330} = e^{-k_{pb}t}(c_1 + c_2 e^{-k_{obs}}) \quad \text{(S1)}$$

where t is time, c_1 is the fluorescence intensity at $t = 0$, and c_2 is the difference in fluorescence between the initial and final states. Note that c_2 was positive in unfolding experiments and negative in refolding experiments (this sign change is expected: only bleaching of the unfolded state contributes to the double exponential kinetics in unfolding experiments, whereas only bleaching of the folded state contributes to double exponential kinetics in folding experiments). There was no indication in any of the kinetic experiments performed that equation S1 was inadequate to describe the observed folding kinetics. Thus, after accounting for photobleaching, folding was a monoexponential process for all variants.

Chaotrope denaturation of RnCD2* variants

All fluorescence measurements for equilibrium chaotrope denaturation experiments were taken on a CARY Eclipse fluorescence spectrophotometer. The temperature at reading was kept constant at 25 °C using a CARY single cell Peltier accessory.

For equilibrium denaturation studies, solutions of RnCD2* variants were prepared in PBS and high concentration of urea or GuHCl (in 1×PBS) at matched protein concentrations (15-

20μg/mL). The solutions were mixed to produce approximately thirty 120 μL samples at regular intervals of urea or GuHCl concentrations. Solutions were allowed to equilibrate for at least 30 min before fluorescence emission spectra were scanned. The average of three scans was used for the analysis described below.

Global fit of kinetic and equilibrium data

Plots of the natural logarithm of the observed rate of equilibration between the folded and unfolded states of a protein, ln (k_{obs}), vs. denaturant concentration have characteristic V-shapes (hence the term "chevron plot"). The quantity k_{obs} is equal to the sum of the unfolding and folding rate constants, k_u and k_f. Chevron plots therefore result from the dependence of ln k_u and ln k_f on urea concentration. The unfolding rate constant dominates k_{obs} at high denaturant concentrations, where the chevron plots for several of the RnCD2* variants are slightly curved. Curvature in the unfolding arm of a chevron plot is often attributed to changes in the structure of the folding transition state. This behavior is accounted for by assuming that ln k_u has a quadratic dependence on denaturant concentration:

$$\ln k_u = \ln k_{u,0} + m_{u1}[D] + m_{u2}[D]^2 \qquad \text{(S2)}$$

Where [D] is denaturant concentration, $k_{u,0}$ is the unfolding rate constant at [D] = 0, and m_{u1} and m_{u2} are the coefficients of the linear and squared terms in the dependence of ln k_u on [D] (it should be noted that k_u and the equilibrium constant K_f are interrelated, so m_{u2} likely reflects a quadratic dependence on denaturant concentration for each of these parameters). The folding rate constant dominates k_{obs} at low denaturant concentrations, where, again, the chevron plots for many of the RnCD2* variants are curved. This has been observed previously by Parker et al (*12*). We have attributed this to the rapid formation of an off-pathway intermediate (but note that this

cannot be kinetically distinguished from an on-pathway intermediate (*3*)). Thus, the effective folding rate constant, k_f^*, depends as follows on denaturant concentration:

$$k_f^* = f_u k_f = \frac{e^{\ln k_{f,0} + m_f[D]}}{1 + e^{\ln K_{i,0} + m_i[D]}} \tag{S3}$$

where f_u is the fraction of not-yet-folded protein that is in the unfolded state (instead of the off-pathway intermediate state; i.e, f_u = [U]/([U]+[I]) = 1/(1+K_i)), k_f is the true folding rate constant at a given denaturant concentration, [D] is denaturant concentration, $k_{f,0}$ is the true folding rate constant at [D] = 0, m_f is the slope of the dependence of ln k_f on [D], $K_{i,0}$ is the equilibrium constant for formation of the off-pathway intermediate at [D] = 0, and m_i is the slope of the dependence of ln K_i on [D]. Summing the expressions for k_f* and k_u yields an equation for k_{obs}:

$$\ln k_{obs} = \ln(k_u + k_f^*) = \ln\left(e^{\ln k_{u,0}} e^{m_{u1}[D] + m_{u2}[D]^2} + \frac{e^{\ln k_{f,0} + m_f[D]}}{1 + e^{\ln K_{i,0} + m_i[D]}} \right) \tag{S4}$$

This equation can be fit to folding kinetics vs. denaturant concentration data to get the parameters of interest (primarily $k_{f,0}$ and $k_{u,0}$). However, the robustness of the fit can be improved by simultaneously fitting kinetics and equilibrium data. The folding equilibrium constant at a given denaturant concentration (K_f) is related to the parameters above as follows:

$$K_f = \frac{k_f}{k_u} = \frac{e^{\ln k_{f,0} + m_f[D]}}{e^{\ln k_{u,0}} e^{m_{u1}[D] + m_{u2}[D]^2}} \tag{S5}$$

This expression can be inserted into the equation for fluorescence-detected equilibrium denaturation:

$$F = F_{f,0} + \varphi_f[D] + \frac{\Delta F + \Delta\varphi[D]}{1 + K_f} = F_{f,0} + \varphi_f[D] + \frac{\Delta F + \Delta\varphi[D]}{1 + \frac{e^{\ln k_{f,0} + m_f[D]}}{e^{\ln k_{u,0}} e^{m_{u1}[D] + m_{u2}[D]^2}}} \tag{S6}$$

Where F is the total fluorescence, $F_{f,0}$ is the fluorescence of the folded protein at [D] = 0, φ_f is the slope of the fluorescence of the folded state vs. [D], ΔF is the difference in fluorescence

between the unfolded and folded states, and $\Delta\varphi$ is the difference between the slopes of the fluorescences of the folded and unfolded states vs [D]. Some of the same parameters occur in the models for the dependence on [D] of the folding kinetics and equilibrium. This circumstance enables the simultaneous fitting of kinetic and equilibrium data mentioned above.

To ensure that the kinetic and equilibrium data had equal influence on the parameter estimates, the equilibrium data were weighted as follows: 1. the equilibrium and kinetic data were fit separately to their models; 2. the root mean squared residuals for the two fits were calculated; 3. the ratio of the kinetic and equilibrium RMS residuals was calculated ($RMS_{kinetic}/RMS_{equilibrium}$); 4. the equilibrium data points were multiplied by this ratio. The combined kinetic and (weighted) equilibrium data sets were then fit simultaneously to the combined kinetic and equilibrium model using Mathematica 7.0 (Wolfram Research). The fit yielded estimates for $k_{f,0}$ and $k_{u,0}$ which were converted to a folding free energy ($\Delta G_{f,0}$) through the relation:

$$\Delta G_{f,0} = -RT \ln K_{f,0} = -RT \ln k_{f,0}/k_{u,0} \quad \text{(S7)}$$

The slope of the dependence of $\Delta G_{f,0}$ on [D] at [D] = 0, $m_{eq,0}$, was determined from the values of m_f and m_{u1} through the relation:

$$m_{eq,0} = -RT(m_f - m_{u1}) \quad \text{(S8)}$$

Before fitting, the values of k_{obs} were corrected for the viscosity of aqueous urea (*13*) as described by Jäger et al (*14*).

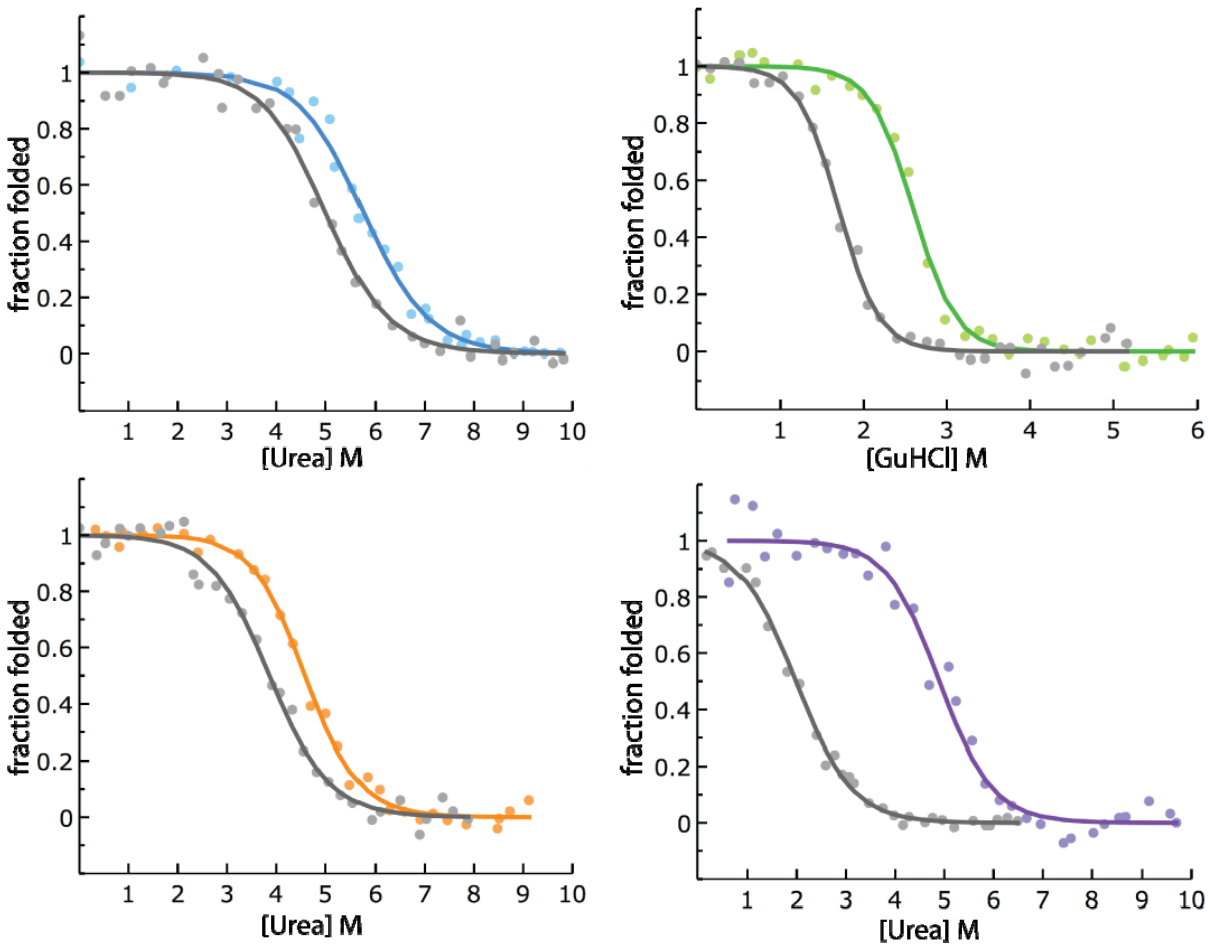
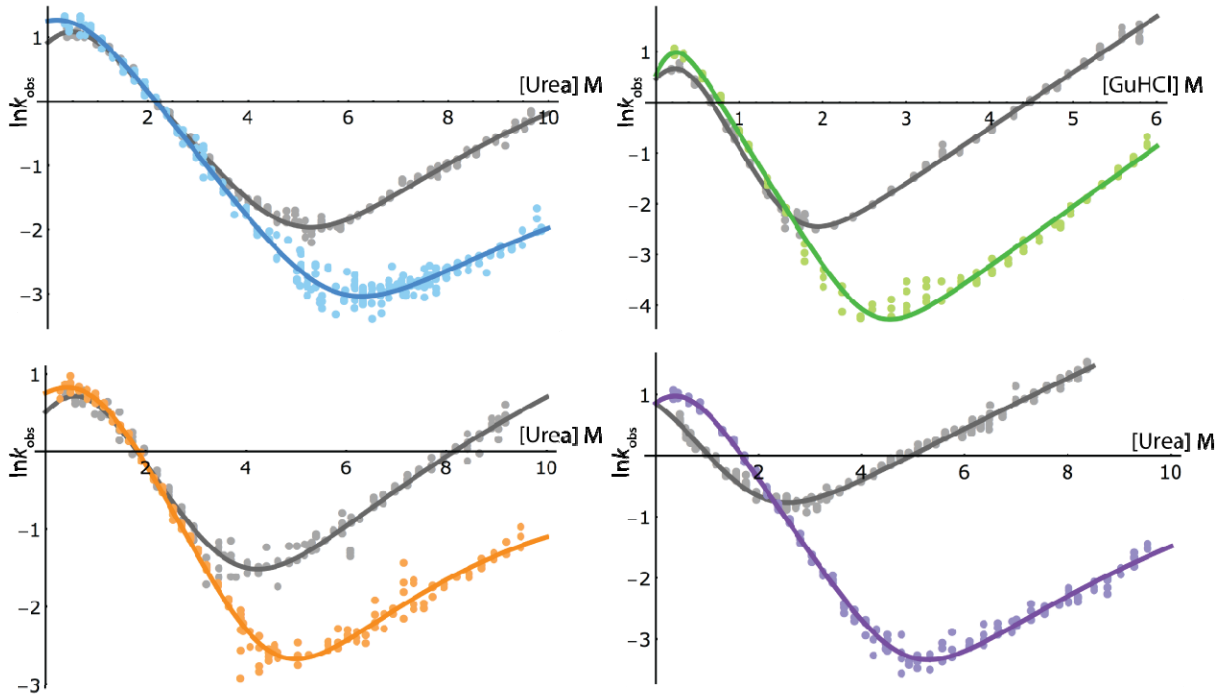


Figure S2. Equilibrium denaturation plots of RnCD2* variants. Equilibrium denaturation curves are depicted for all RnCD2* variants. Equilibrium denaturation and chevron plots were globally fit to determine: $\Delta G_{f,0}$, $k_{f,0}$ and $k_{u,0}$, the free energy of folding, and the folding and unfolding rate constants, respectively, under native conditions; m_f and m_{u1}, the linear slopes of the dependences of folding and unfolding on [chaotrope]; and m_{u2} a small quadratic term accounting for the slight curvature in plots at high chaotrope concentrations. **g-RnCD2*** shown in blue, **g-RnCD2*-K** shown in orange, **g-RnCD2*-F** shown in green, **g-RnCD2*-K,F** shown in purple. All non-glycosylated sequence matched controls shown in gray.



RnCD2* variant	ΔG_f (kcal/mol)	$k_{f,0}$ (s^{-1})	m_f (M^{-1})	$k_{u,0}$ (s^{-1})	m_{u1} (M^{-1})	m_{u2} (M^{-1})	m_{eq} (kcal/mol M)
RnCD2*	-5.1 ± 0.4	7.5±0.5	-0.9±0.02	0.0014±0.0007	0.5±0.1	-0.03±0.007	-1.1±0.2
g-RnCD2*	-5.7 ± 0.5	9.3±0.9	-1.0±0.02	0.0007±0.0007	0.8±0.2	-0.02±0.01	-1.1±0.3
RnCD2*-K	-4.0 ± 0.5	7.3±1.5	-1.1±0.06	0.009±0.004	0.7±0.1	-0.02±0.008	-1.1±0.3
g-RnCD2*-K	-5.5 ± 0.4	11.1±1.5	-1.2±0.04	0.001±0.0005	1.0±0.1	-0.04±0.009	-1.3±0.2
RnCD2*-F	-4.2 ± 0.2	8.0±0.9	-3.0±0.08	0.007±0.0003	1.1±0.01	-0.01±0.0	-2.4±0.2
g-RnCD2*-F	-6.0 ± 0.3	8.0±1.0	-2.7±0.06	0.0004±0.0	1.2±0.03	-0.01±0.0	-2.3±0.2
RnCD2*-KF	-2.1 ± 0.4	3.4±0.9	-1.3±0.1	0.1±0.01	0.5±0.04	-0.005±0.0004	-1.1±0.2
g-RnCD2*-KF	-5.8 ± 0.5	8.3±0.8	-1.2±0.03	0.0005±0.0004	0.9±0.2	-0.03±0.01	-1.3±0.3

Non-glycosylated sequence matched controls in all plots

Figure S3. Chevron plots of RnCD2* variants. Chevron plots are depicted for all variants. Equilibrium denaturation and chevron plots were globally fit to determine: $\Delta G_{f,0}$, $k_{f,0}$ and $k_{u,0}$, the free energy of folding, and the folding and unfolding rate constants, respectively, under native conditions; m_f and m_{u1}, the linear slopes of the dependences of folding and unfolding on [chaotrope]; and m_{u2} a small quadratic term accounting for the slight curvature in plots at high chaotrope concentrations. **g-RnCD2*** shown in blue, **g-RnCD2*-K** shown in orange, **g-RnCD2*-F** shown in green, **g-RnCD2*-K,F** shown in purple. All non-glycosylated sequence matched controls shown in gray.

Glycosylation efficiency of g-RnCD2*-K and g-RnCD2*-F in Sf9 cells

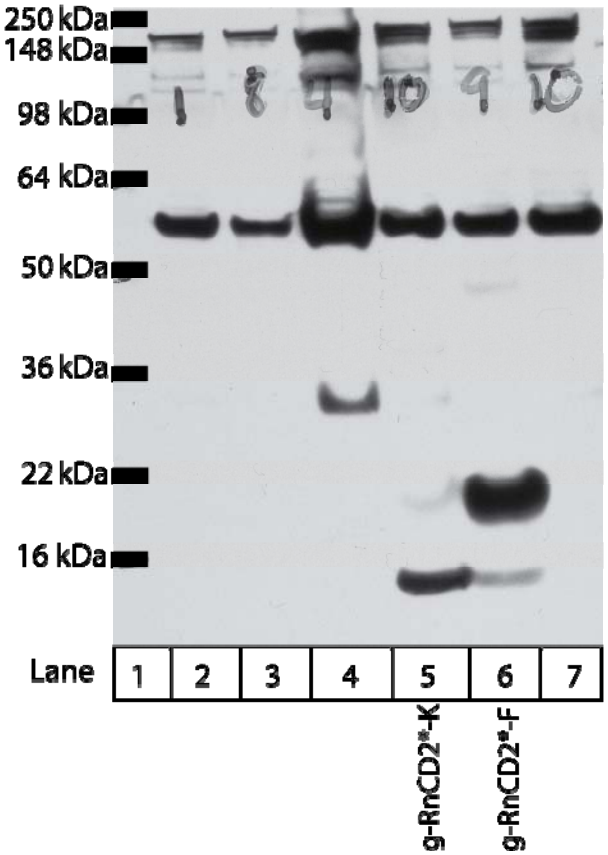


Figure S4. Entire Western blot (anti-FLAG tag) including the selected data presented in Fig. 2C. Lane 1 SeeBlue® Plus2 Pre-Stained Standard (Invitrogen); lanes 2, 3, 4, 7: conditioned media from Sf9 cells infected by baculovirus without a protein expression insert; lane 5: conditioned media from Sf9 cells infected by baculovirus with the **g-RnCD2*-K** expression insert, lane 6: conditioned media from Sf9 cells infected by baculovirus with the **g-RnCD2*-F** expression insert.

Characterization of folding thermodynamics of AcyP2* variants

Acetate buffer (50 mM Acetate, 0.5 mM TCEP, 0.01% sodium azide, pH 5.5) was made fresh daily from a 4× stock and filtered. Urea solutions were prepared fresh daily in 1×acetate, filtered, and concentrations were confirmed by index of refraction (IOR). Subsequent dilutions of urea were made with 1×acetate and concentrations were checked by IOR. Constants defined in equations include the universal gas constant (R) and temperature (T). The value of RT at 25 °C was taken to be 0.592 kcal/mol. Data were imported and fit as described below in Microsoft Excel.

Chaotrope denaturation of AcyP2* variants

All fluorescence measurements for equilibrium chaotrope denaturation experiments were taken on a CARY Eclipse fluorescence spectrophotometer. The temperature at reading was kept constant at 25 °C using a CARY single cell Peltier accessory. Each chaotrope denaturation experiment was repeated at least three times for each variant.

For equilibrium denaturation studies, solutions of AcyP2* variants were prepared in acetate and high concentration of urea (in 1×acetate) at matched protein concentrations (15-30 μg/mL). The solutions were mixed to produce approximately thirty 120 μL samples at regular intervals of urea or GuHCl concentrations. Solutions were allowed to equilibrate for at least 30 min before fluorescence emission spectra were scanned. The average of three scans was used for the analysis described below. Like RnCD2*, AcyP2* unfolding in response to increasing concentrations of urea or GuHCl causes a shift and intensity change in fluorescence spectrum. Thus, plots of fluorescence intensity at single wavelenths (F_λ), versus chaotrope concentration were plotted to demonstrate unfolding. $\Delta G_{f,0}$ and m_{eq} values for AcyP2* were estimated by fitting fluorescence intensity at 330 nm (F_{330}) vs. urea concentration data to:

$$F = F_{f,0} + \varphi_f[D] + \frac{\Delta F + \Delta\varphi[D]}{1 + e^{-(\Delta G_{f,0} + m_{eq}[D]/RT}} \quad \text{(S9)}$$

Where F is the total fluorescence, $F_{f,0}$ is the fluorescence of the folded protein at [D] = 0, φ_f is the slope of the fluorescence of the folded state vs. [D], ΔF is the difference in fluorescence between the unfolded and folded states, and $\Delta\varphi$ is the difference between the slopes of the fluorescences of the folded and unfolded states vs [D].

$\Delta G_{f,0}$ and m_{eq} values derived from single chaotrope denaturation experiments were averaged to give the $\Delta G_{f,0}$ values and fits reported.

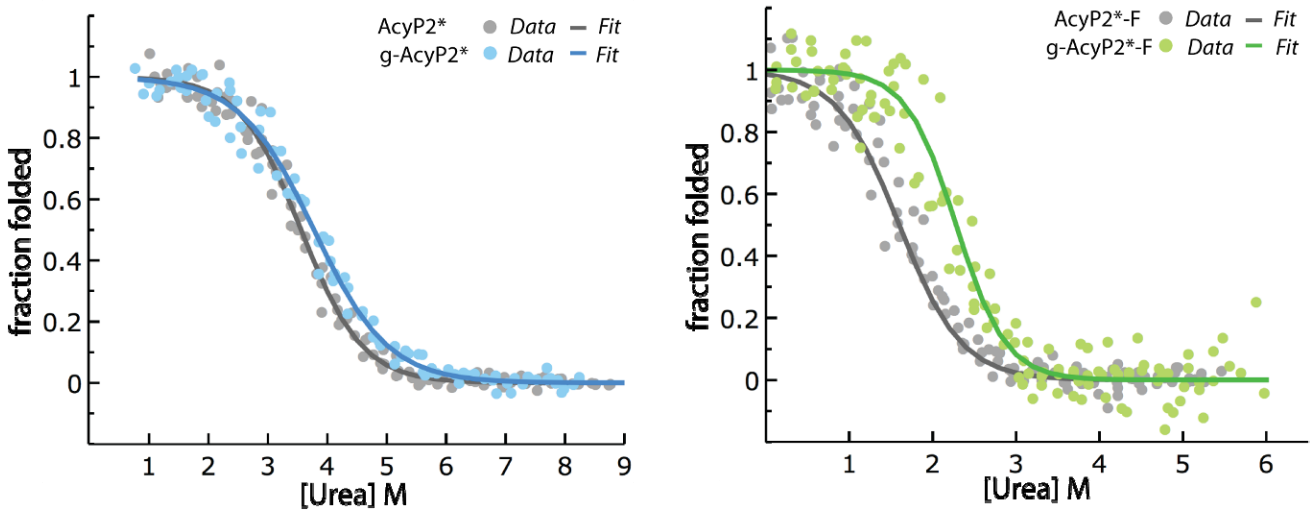


Figure S5. Equilibrium denaturation plots AcyP2* variants. Equilibrium denaturation curves are depicted for all variants. Individual equilibrium denaturation fits were averaged to determine: $\Delta G_{f,0}$, the free energy of folding under native conditions. **g-AcyP2*** shown in blue, **g-AcyP2*-F** shown in green. All non-glycosylated sequence matched controls shown in gray.

Glycosylation efficiency of g-AcyP2*-F and g-AcyP2* in Sf9 cells

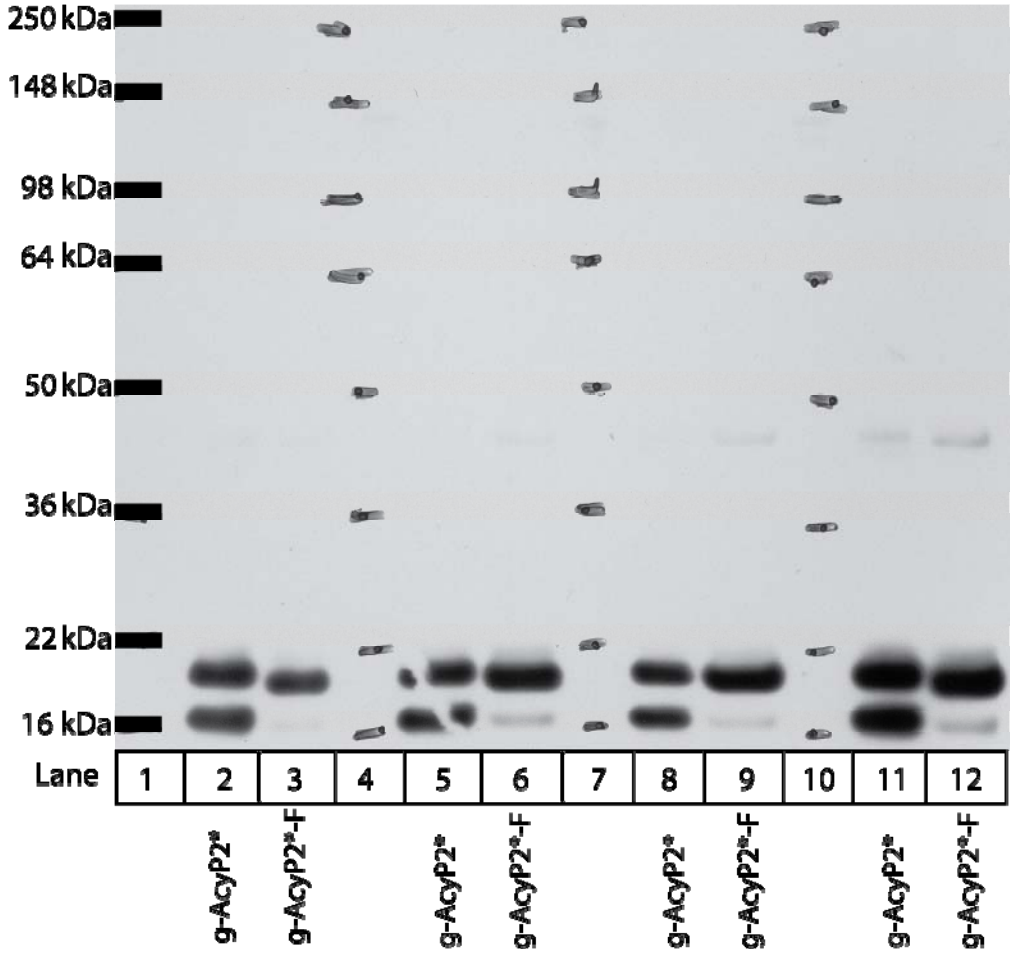


Figure S6. Entire Western blot (anti-FLAG tag) including the selected data presented in Fig. 2F. Replicates of conditioned media from Sf9 cells infected by baculovirus with the **g-AcyP2*** and **g-AcyP2*-F** expression inserts. Lanes 1, 4, 7, 10 SeeBlue® Plus2 Pre-Stained Standard (Invitrogen); lanes 2, 5, 8, 11 **g-AcyP2*** Sf9 conditioned media, lanes 3, 6, 9, 12 **g-AcyP2*-F** Sf9 conditioned media.

Synthesis of Pin WW variants

Pin WW proteins were synthesized as C-terminal acids, employing a solid phase peptide synthesis approach using a standard Fmoc Nα protecting group strategy either manually (protein **WW**) or via a combination of manual and automated methods (proteins **g-WW**, **WW-F**, **g-WW-F**, **WW-T**, **g-WW-T**, **WW-F,T**, and **g-WW-F,T** were synthesized on an Applied Biosystems 433A automated peptide synthesizer except for the manual coupling of Fmoc-Asn(Ac_3GlcNAc)-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(Ac_3GlcNAc)-OH (N-α-Fmoc-N-β-[3,4,6-tri-O-acetyl-2-(acetylamino)-deoxy-2-β-glucopyranosyl]-L-asparagine) (*15, 16*). 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 μmol at 0.23 mmol/g resin loading) was aliquotted into a fritted polypropylene syringe and allowed to swell in CH_2Cl_2 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 (containing the crude protein) was stored at -20 °C until purification.

Deprotecting glycosylated Pin WW proteins

Acetate protecting groups were removed from the 3-, 4-, and 6-hydroxyl groups on the Asn-linked GlcNAc residues in proteins **g-WW**, **g-WW-F**, **g-WW-T**, and **g-WW-F,T** via hydrazinolysis as described previously (*17*). Briefly, the crude protein was dissolved in a solution of 5% hydrazine in 60 mM aqueous dithiothreitol (sometimes containing as much as 50% acetonitrile, to facilitate dissolution of the crude protein) and allowed to stand at room temperature for ~1 h with intermittent agitation. The deprotection reaction was quenched by the addition of ~ 1 mL TFA and ~ 20 mL water. The quenched reaction mixture was frozen and lyophilized to give the crude deprotected protein as a white powder.

HPLC purification and MS characterization of Pin WW variants

Immediately prior to purification, the crude proteins were dissolved in either 1:1 water:acetonitrile, DMSO, or 8 M GuHCl (depending on solubility of the crude protein---8 M GuHCl was frequently required to dissolve the crude glycosylated proteins even though these proteins were readily soluble in water after purification). 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 S2), and purity was established by analytical HPLC.

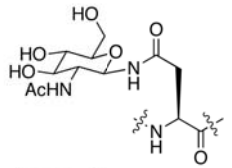
Circular dichroism spectroscopy of Pin WW variants

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

Table S2. Pin WW Domain Protein Sequences and MALDI-TOF Data

Protein	Sequence	Expected $[M+H]^+$ (g/mol)	Observed $[M+H]^+$ (g/mol)
WW	H_2N-KLPPGWEKRM**S**RS**N**G**R**VYYFNHITNASQFERPSG-COOH	4010.0	4010.2
g-WW	H_2N-KLPPGWEKRM**S**RS(**N**)G**R**VYYFNHITNASQFERPSG-COOH	4213.1	4213
WW-F	H_2N-KLPPGWEKRM**F**RS**N**G**R**VYYFNHITNASQFERPSG-COOH	4070.0	4071.1
g-WW-F	H_2N-KLPPGWEKRM**F**RS(**N**)G**R**VYYFNHITNASQFERPSG-COOH	4273.1	4272.5
WW-T	H_2N-KLPPGWEKRM**S**RS**N**G**T**VYYFNHITNASQFERPSG-COOH	3954.9	3955.3
WW-T	H_2N-KLPPGWEKRM**S**RS(**N**)G**T**VYYFNHITNASQFERPSG-COOH	4158.0	4157.2
WW-F,T	H_2N-KLPPGWEKRM**F**RS**N**G**T**VYYFNHITNASQFERPSG-COOH	4015.0	4016.1
WW-F,T	H_2N-KLPPGWEKRM**F**RS(**N**)G**T**VYYFNHITNASQFERPSG-COOH	4218.1	4218.8

6 10 15 20 25 30 35 39

(N) = 

Asn-GlcNAc

tyrosine and tryptophan absorbance at 280 nm in 6 M GuHCl + 20 mM sodium phosphate (ε_{Trp} = 5690 $M^{-1}cm^{-1}$, ε_{Tyr} = 1280 $M^{-1}cm^{-1}$) (*18*). 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:

$$[\theta] = \frac{(D_0 + D_1 \cdot T) + K_f(N_0 + N_1 \cdot T)}{1 + K_f}, \quad \text{(S10)}$$

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:

$$K_f = \exp\left[\frac{-\Delta G_f(T)}{RT}\right], \quad \text{(S11)}$$

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:

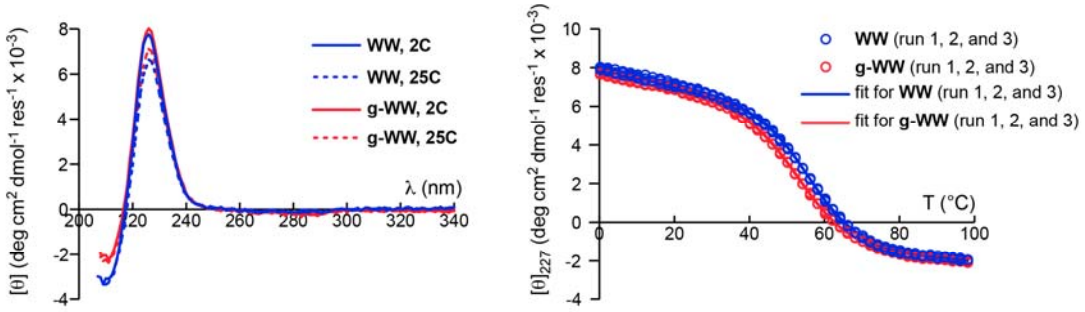
$$\Delta G_f(T) = \frac{\Delta H(T_m)}{T_m}(T_m - T) + \Delta C_p\left[T - T_m - T\ln\left(\frac{T}{T_m}\right)\right], \quad \text{(S12)}$$

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 the fit). The second equation represents $\Delta G_f(T)$ as a Taylor series expansion about the melting temperature:

$$\Delta G_f(T) = \Delta G_0 + \Delta G_1(T - T_m) + \Delta G_2(T - T_m)^2, \quad \text{(S13)}$$

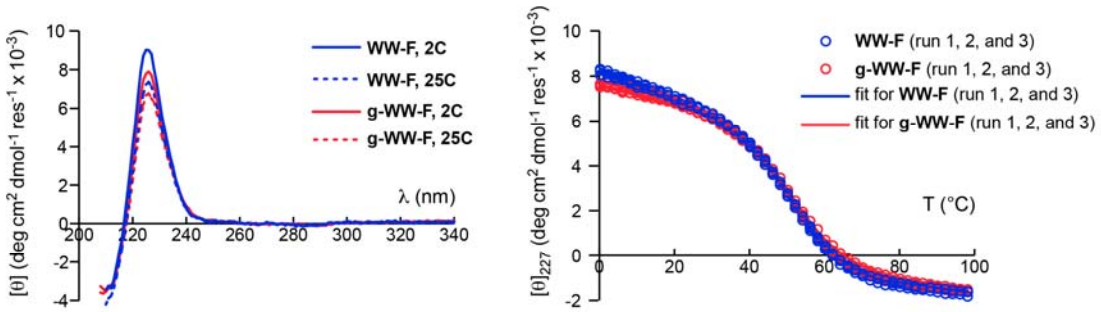
in which ΔG_0, ΔG_1, and ΔG_2 are parameters of the fit and T_m is a constant obtained from the van't Hoff fit (in equation S12). The ΔG_f values displayed in Table 1 of the main text for each Pin WW variant were obtained by averaging the ΔG_f values (calculated at 328.15 K using equation S13) from each of three or more replicate variable temperature CD experiments on the same protein.

CD spectra and variable-temperature CD data for proteins Pin WW proteins **WW**, **g-WW**, **WW-F**, **g-WW-F**, **WW-T**, **g-WW-T**, **WW-F,T**, and **g-WW-F,T** appear in Figures S7-S10, along with parameters from equations S10−S13 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 ΔG_f values, along with the uncertainty in the folding and unfolding rate ratios shown in Table 1 of the main text by propagation of error.



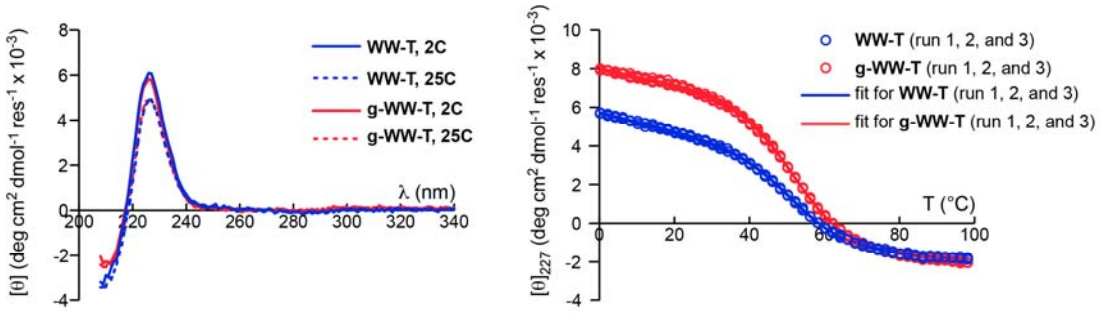
	$\Delta H(T_m)$ (kcal/mol)	ΔC_p (kcal/mol/K)	T_m (K)	ΔG_0 (kcal/mol)	ΔG_1 (kcal/mol/K)	ΔG_2 (kcal/mol/K²)	N_0 ([θ])	N_1 ([θ]/K)	D_0 ([θ])	D_1 ([θ]/K)
protein **WW**										
run 1	-35.0 (1.3)	-0.16 (0.32)	329.6 (0.6)	-0.0005 (0.0632)	0.106 (0.004)	0.0002 (0.0005)	20.1	-0.044	5.8	-0.021
run 2	-32.8 (0.9)	-0.43 (0.23)	329.1 (0.5)	-0.0009 (0.0533)	0.100 (0.002)	0.0007 (0.0003)	19.8	-0.043	6.6	-0.022
run 3	-31.8 (1.0)	-0.45 (0.26)	328.9 (0.7)	-0.0012 (0.0658)	0.096 (0.003)	0.0007 (0.0004)	19.2	-0.041	5.8	-0.021
glycoprotein **g-WW**										
run 1	-30.2 (0.7)	-0.28 (0.20)	326.7 (0.6)	-0.0005 (0.0561)	0.092 (0.002)	0.0004 (0.0003)	18.0	-0.038	4.4	-0.017
run 2	-31.7 (0.7)	-0.21 (0.20)	326.7 (0.5)	-0.0004 (0.0470)	0.097 (0.002)	0.0003 (0.0003)	19.4	-0.042	6.3	-0.023
run 3	-30.6 (0.8)	-0.27 (0.11)	326.2 (0.3)	-0.0005 (0.0202)	0.093 (0.002)	0.0004 (0.0001)	18.1	-0.038	6.5	-0.023

Figure S7. CD spectra (100 μM protein) and variable temperature CD (50 μM protein) data for Pin WW domain protein **WW** (which has Ser at position 16, Asn at position 19, and Arg at position 21) and Pin WW domain glycoprotein **g-WW** (which has Ser at position 16, Asn(GlcNAc) at position 19, and Arg at position 21) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S10-S13) appear in the table, along with parameter standard errors in parentheses.



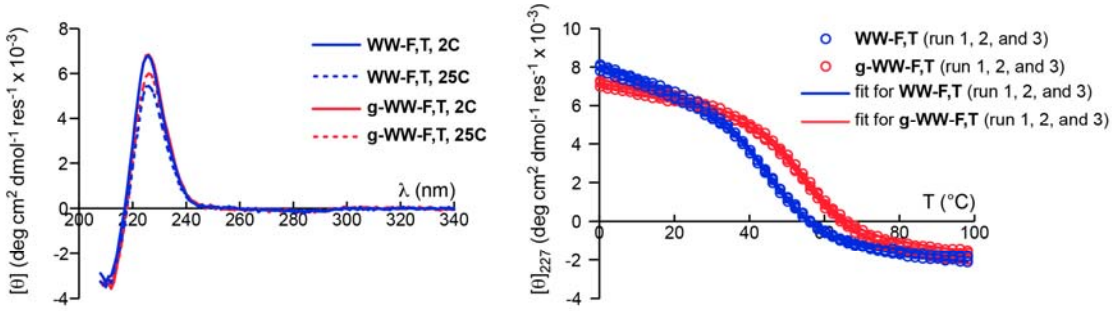
	$\Delta H(T_m)$ (kcal/mol)	ΔC_p (kcal/mol/K)	T_m (K)	ΔG_0 (kcal/mol)	ΔG_1 (kcal/mol/K)	ΔG_2 (kcal/mol/K²)	N_0 ([θ])	N_1 ([θ]/K)	D_0 ([θ])	D_1 ([θ]/K)
protein **WW-F**										
run 1	-32.2 (0.9)	-0.24 (0.26)	324.9 (0.6)	-0.0005 (0.0580)	0.099 (0.003)	0.0004 (0.0004)	23.5	-0.056	6.3	-0.022
run 2	-31.8 (0.7)	-0.23 (0.20)	325.3 (0.5)	-0.0004 (0.0465)	0.098 (0.002)	0.0004 (0.0003)	24.7	-0.060	5.5	-0.019
run 3	-32.5 (0.8)	-0.15 (0.21)	325.6 (0.5)	-0.0003 (0.0458)	0.100 (0.002)	0.0002 (0.0003)	24.7	-0.060	5.4	-0.019
glycoprotein **g-WW-F**										
run 1	-30.3 (0.8)	-0.27 (0.22)	324.2 (0.6)	-0.0005 (0.0578)	0.093 (0.002)	0.0004 (0.0003)	18.4	-0.040	7.0	-0.023
run 2	-30.2 (0.9)	-0.22 (0.24)	324.3 (0.7)	-0.0005 (0.0651)	0.093 (0.003)	0.0004 (0.0004)	18.8	-0.041	8.2	-0.026
run 3	-29.6 (0.9)	-0.21 (0.24)	324.5 (0.8)	-0.0005 (0.0692)	0.091 (0.003)	0.0003 (0.0004)	18.2	-0.039	4.9	-0.017

Figure S8. CD spectra (100 μM protein) and variable temperature CD (50 μM protein) data for Pin WW domain protein **WW-F** (which has Phe at position 16, Asn at position 19, and Arg at position 21) and Pin WW domain glycoprotein **g-WW-F** (which has Phe at position 16, Asn(GlcNAc) at position 19, and Arg at position 21) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S10-S13) 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/K²)	N_0 ([θ])	N_1 ([θ]/K)	D_0 ([θ])	D_1 ([θ]/K)
protein **WW-T**										
run 1	-29.7 (1.0)	-0.11 (0.26)	325.2 (0.8)	-0.0002 (0.0732)	0.091 (0.003)	0.0002 (0.0004)	18.4	-0.046	4.2	-0.016
run 2	-29.3 (1.1)	-0.03 (0.24)	327.0 (0.8)	-0.0002 (0.0731)	0.090 (0.004)	0.00004 (0.00037)	19.6	-0.051	1.3	-0.008
run 3	-30.4 (1.0)	-0.04 (0.25)	325.9 (0.7)	-0.0003 (0.0667)	0.093 (0.003)	0.00006 (0.00039)	19.3	-0.050	3.9	-0.015
glycoprotein **g-WW-T**										
run 1	-29.0 (0.9)	-0.28 (0.23)	324.3 (0.8)	-0.0006 (0.0692)	0.089 (0.002)	0.0004 (0.0003)	16.7	-0.032	5.7	-0.021
run 2	-29.8 (0.8)	-0.19 (0.22)	325.6 (0.7)	-0.0004 (0.0623)	0.091 (0.002)	0.0003 (0.0003)	19.2	-0.041	4.4	-0.017
run 3	-30.4 (0.8)	-0.18 (0.22)	325.5 (0.6)	-0.0004 (0.0583)	0.093 (0.003)	0.0003 (0.0003)	20.5	-0.046	5.8	-0.021

Figure S9. CD spectra (10 μM protein) and variable temperature CD (50 μM protein) data for Pin WW domain protein **WW-T** (which has Ser at position 16, Asn at position 19, and Thr at position 21) and Pin WW domain glycoprotein **g-WW-T** (which has Ser at position 16, Asn(GlcNAc) at position 19, and Thr at position 21) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S10-S13) 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/K²)	N_0 ([θ])	N_1 ([θ]/K)	D_0 ([θ])	D_1 ([θ]/K)
protein **WW-F,T**										
run 1	-30.5 (0.9)	0.03 (0.25)	320.3 (0.6)	-0.0003 (0.0580)	0.095 (0.003)	-0.00003 (0.00039)	27.1	-0.070	4.6	-0.17
run 2	-29.4 (0.9)	0.0008 (0.2192)	321.6 (0.6)	-0.0003 (0.0586)	0.091 (0.003)	0.000004 (0.000342)	27.2	-0.071	3.5	-0.01
run 3	-30.4 (0.8)	0.11 (0.20)	320.4 (0.5)	-0.0003 (0.0471)	0.095 (0.003)	-0.0002 (0.0003)	28.6	-0.075	6.0	-0.02
glycoprotein **g-WW-F,T**										
run 1	-32.5 (1.3)	-0.41 (0.34)	327.8 (0.8)	-0.0009 (0.0777)	0.099 (0.003)	0.0006 (0.0005)	18.7	-0.041	7.6	-0.02
run 2	-32.6 (1.2)	-0.31 (0.33)	327.4 (0.8)	-0.0006 (0.0740)	0.099 (0.003)	0.0005 (0.0005)	17.3	-0.038	9.4	-0.03
run 3	-30.2 (1.0)	-0.15 (0.24)	329.7 (0.8)	-0.0005 (0.0728)	0.091 (0.003)	0.0002 (0.0004)	19.0	-0.043	1.6	-0.01

Figure S10. CD spectra (10 μM protein) and variable temperature CD (50 μM protein) data for Pin WW domain protein **WW-F,T** (which has Phe at position 16, Asn at position 19, and Thr at position 21) and Pin WW domain glycoprotein **g-WW-F,T** (which has Phe at position 16, Asn(GlcNAc) at position 19, and Thr at position 21) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations S10-S13) appear in the table, along with parameter standard errors in parentheses.

Laser temperature jump experiments on Pin WW variants

Relaxation times following a rapid laser-induced temperature jump of ~12 °C were measured by monitoring Trp fluorescence of a 50 μM solution of Pin WW proteins **WW**, **g-WW**, **WW-F**, **g-WW-F**, **WW-T**, **g-WW-T**, **WW-F,T**, and **g-WW-F,T** in 20 mM sodium phosphate (pH 7) using a nanosecond laser temperature jump apparatus, as described previously (*14, 19-21*), to monitor the fluorescence decay of a Trp residue in each protein after a laser-induced temperature jump (see Supplementary Figures S11-S18) at each of several temperatures.

The relaxation traces shown in Supplementary Figures S11-S18 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:

$$f(t) = a_1(t) \cdot f_1 + a_2(t) \cdot f_2, \tag{S14}$$

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 (*14*). Then, the relaxation of the protein to equilibrium at the new temperature following the laser-induced temperature jump can be represented as $\chi_1(t)$:

$$\chi_1(t) = \frac{a_1(t)}{a_1(t) + a_2(t)}, \tag{S15}$$

plotted as a function of time for each protein at several temperatures in Supplementary Figures S11-S18 (*20, 21*).

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

$$\chi(t) = C_1 \cdot \exp\left[\frac{-(t - x_0)}{\tau}\right] + C_2, \tag{S16}$$

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} ($k_{obs} = 1/\tau$). Using the temperature-dependent equilibrium constant K_f for each protein (from the variable temperature CD experiments; see equations S11-S13), folding k_f and unfolding k_u rate constants can be extracted from k_{obs} according to the following equations:

$$k_{obs} = k_f + k_u \tag{S17}$$

$$K_f = \frac{k_f}{k_u} \tag{S18}$$

$$k_f = k_{obs} \cdot \left[1 - \frac{1}{K_f + 1}\right] \tag{S19}$$

The folding rates for each protein can then be fit as a function of temperature to the following Kramers model (*22-24*) equation:

$$k_f(T) = \nu(59\ °C) \cdot \frac{\eta(59\ °C)}{\eta(T)} \exp\left[-\frac{\Delta G_0 + \Delta G_1 \cdot (T - T_m) + \Delta G_2 \cdot (T - T_m)^2}{RT}\right], \tag{S20}$$

in which the temperature-dependent free energy of activation $\Delta G^{\dagger}_f$ is represented as a second order Taylor series expansion about the melting temperature T_m, and $\Delta G^{\dagger}_0$, $\Delta G^{\dagger}_1$, and $\Delta G^{\dagger}_2$ are parameters of the fit (these parameters are given for each protein in Supplementary Figures S11-S18). The pre-exponential term in equation S20 represents the viscosity-corrected frequency ν of the characteristic diffusional folding motion at the barrier (at 59 °C, $\nu = 5 \times 10^5$ s^{-1}) (*25-27*). $\eta(59\ °C)$ is the solvent viscosity at 59 °C and $\eta(T)$ is the solvent viscosity at temperature T, both calculated with equation S21:

$$\eta(T) = A \cdot 10^{\frac{B}{T-C}}, \quad \text{(S21)}$$

where $A = 2.41 \times 10^5$ Pa·s, B = 247.8 K, and C = 140 K (*28*).

The parameters for equations S13 and S20 were used to calculate the folding and unfolding rate ratios at 328.15 K for Pin WW proteins **WW**, **g-WW**, **WW-F**, **g-WW-F**, **WW-T**, **g-WW-T**, **WW-F,T**, and **g-WW-F,T** shown in Table 1 of the main text. Table S3 shows the absolute folding and unfolding rates for each Pin WW variant at 328.15 K that were used to calculate the rate ratios in main text Table 1.

Table S3. Folding and Unfolding Rates for Pin-WW-derived proteins and glycoproteins

Protein	Sequence	k_f (s^{-1})	k_u (s^{-1})
WW	--M**S**RS**N**G**R**-- (15–21)	7636 ± 490	6481 ± 647
g-WW	--M**S**RS**N**G**R**--	5342 ± 250	6764 ± 497
WW-F	--M**F**RS**N**G**R**--	2872 ± 137	4478 ± 359
g-WW-F	--M**F**RS**N**G**R**--	2561 ± 225	4439 ± 535
WW-T	--M**S**RS**N**G**T**--	5673 ± 288	7638 ± 785
g-WW-T	--M**S**RS**N**G**T**--	4854 ± 293	7412 ± 750
WW-F,T	--M**F**RS**N**G**T**--	2119 ± 299	6097 ± 968
g-WW-F,T	--M**F**RS**N**G**T**--	4148 ± 414	4064 ± 558

[a]all tabulated data are at 328.15 K

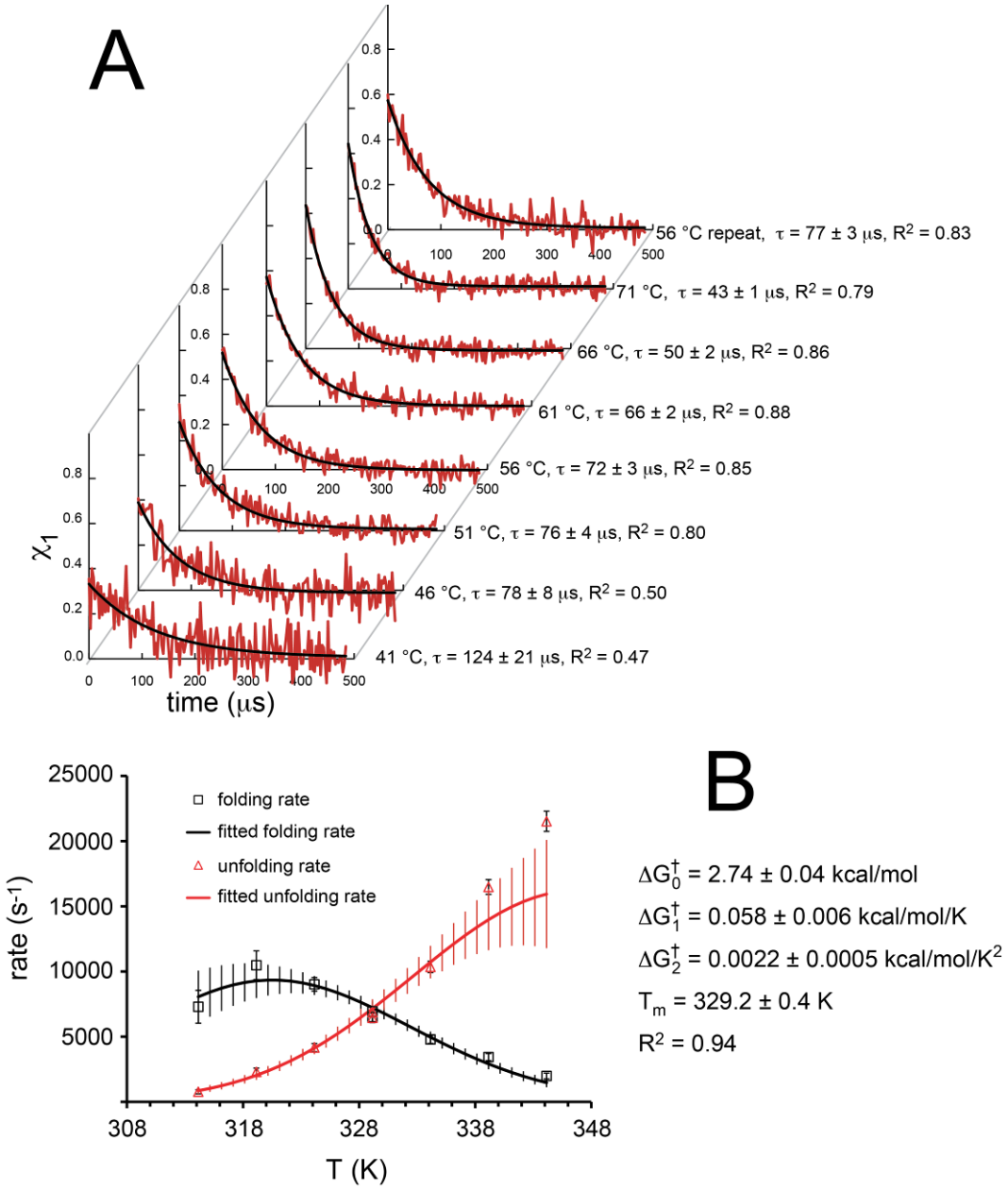


Figure S11. (A) Kinetic relaxation data for Pin WW protein **WW** (which has Ser at position 16, Asn at position 19, and Arg at position 21) at several temperatures following temperature jumps of ~12 °C. Red lines show the change in Trp fluorescence for **WW** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equation S15) 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 S16) with relaxation times as r>1982</year></dates><pub-location>Boca Raton</pub-location><publisher>CRC Press</publisher><urls></urls></rec**or**d></Cite></EndNote>(28).
The parameters for equations S13 and S20 were usede folding rates to equations S20, 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. At high temperature, the fitted uncertainty of the unfolding rate is outside the measured uncertainty, suggesting that the quadratic model for the dependence of the free energy of activation on temperature may not be sufficient. However, the parameters for the fit should still be sufficiently predictive at temperatures close to the melting temperature (the Tm for WW is 329.2 K, and the data in Fig. 4A and Table S3 are calculated at 328.15 K).

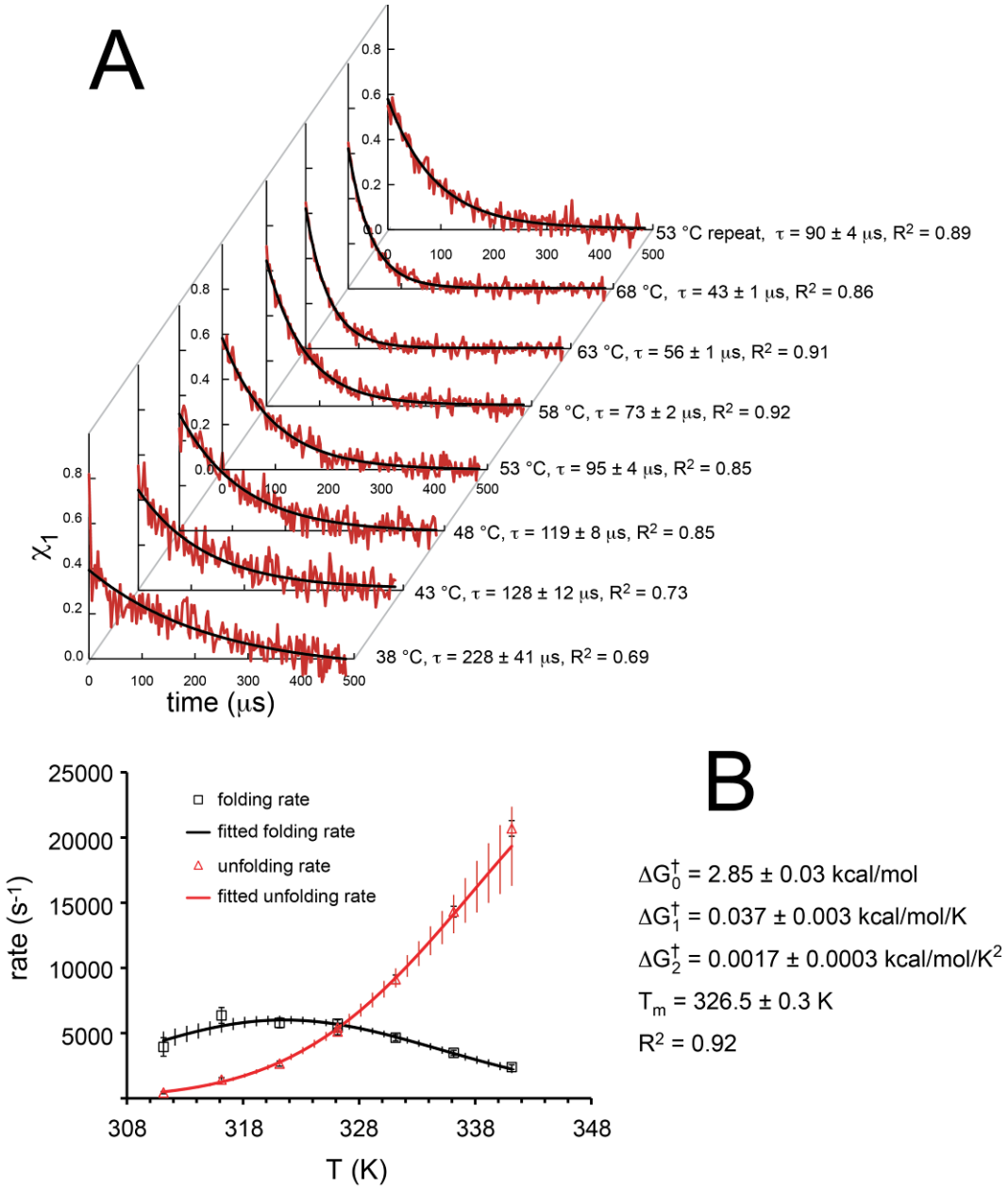


Figure S12. (A) Kinetic relaxation data for Pin WW protein **g-WW** (which has Ser at position 16, Asn(GlcNAc) at position 19, and Arg at position 21) at 50 μM protein concentration in 20 mM sodium phosphate, pH 7, at several temperatures following temperature jumps of ~12 °C. Red lines show the change in Trp fluorescence for **g-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 S15) 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 S16) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW protein **g-WW** as a function of temperature. The black solid line represents the fit of the folding rates to equation S20, 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.

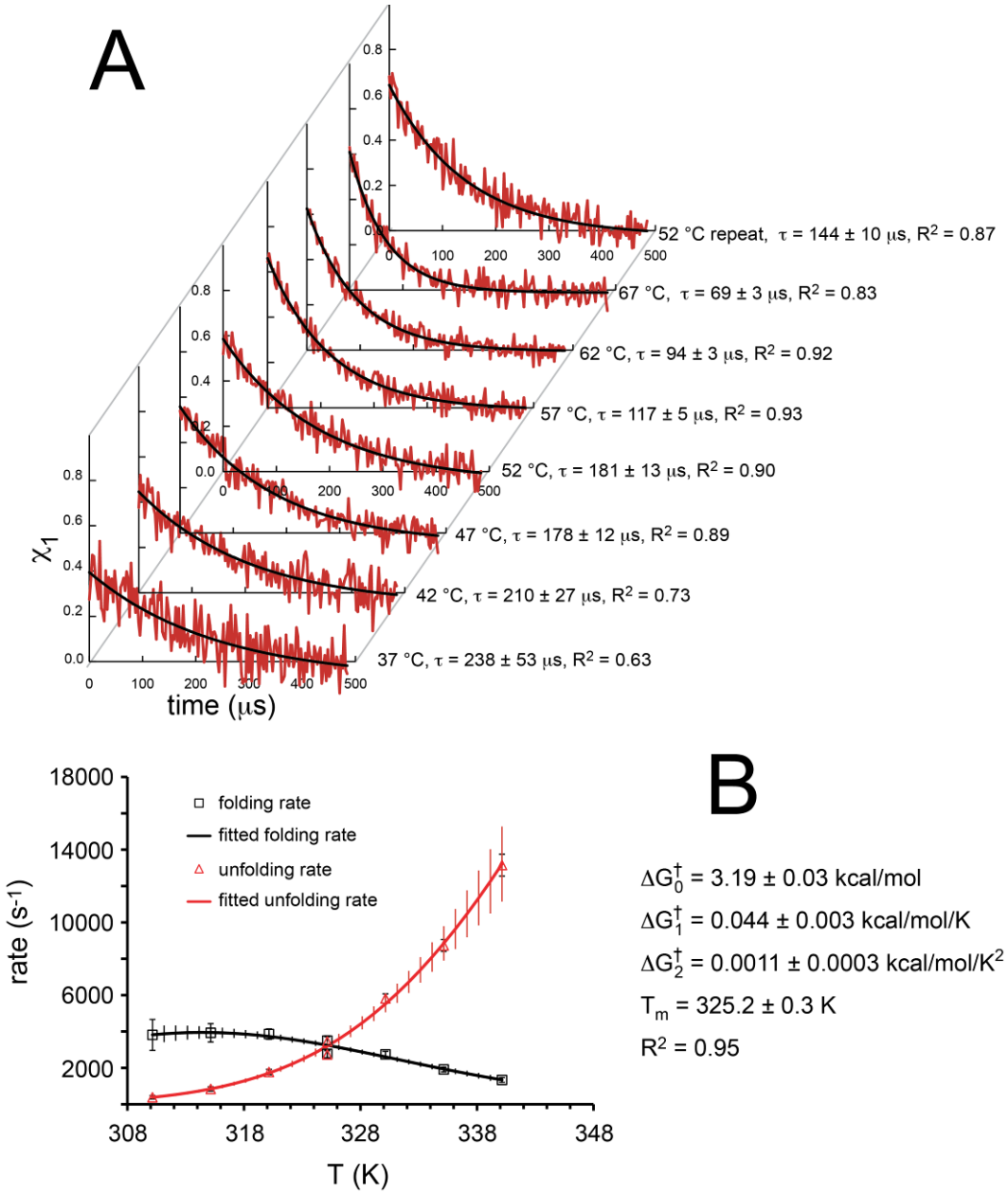


Figure S13. (A) Kinetic relaxation data for Pin WW protein **WW-F** (which has Phe at position 16, Asn at position 19, and Arg at position 21) at 50 μM protein concentration in 20 mM sodium phosphate, pH 7, at several temperatures following temperature jumps of ~12 °C. Red lines show the change in Trp fluorescence for **WW-F** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equation S15) 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 S16) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW protein **WW-F** as a function of temperature. The black solid line represents the fit of the folding rates to equation S20, 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.

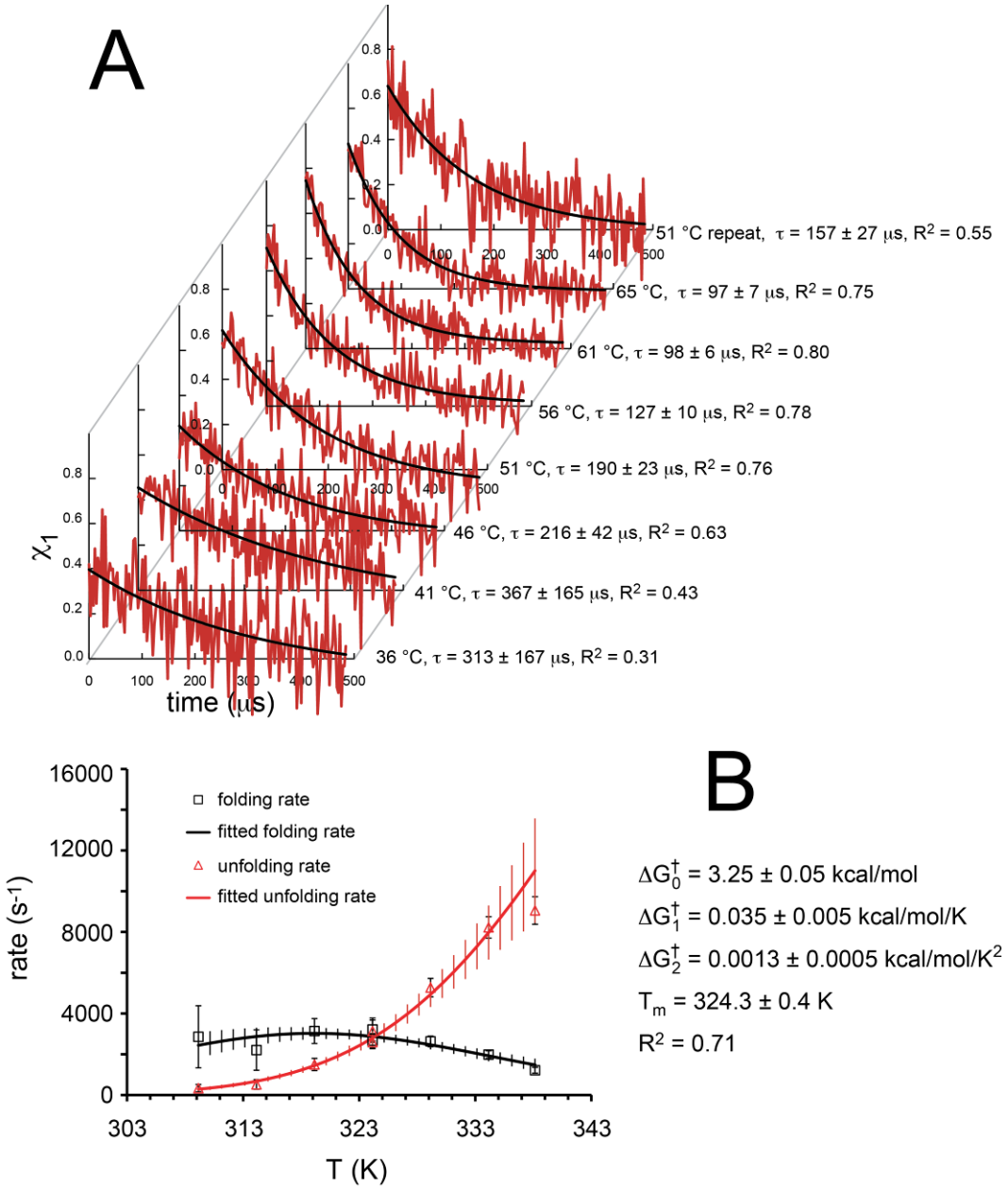


Figure S14. (A) Kinetic relaxation data for Pin WW protein **g-WW-F** (which has Phe at position 16, Asn(GlcNAc) at position 19, and Arg at position 21) at 50 μM protein concentration in 20 mM sodium phosphate, pH 7, at several temperatures following temperature jumps of ~12 °C. Red lines show the change in Trp fluorescence for **g-WW-F** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S15) 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 S16) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW protein **g-WW-F** as a function of temperature. The black solid line represents the fit of the folding rates to equation S20, 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.

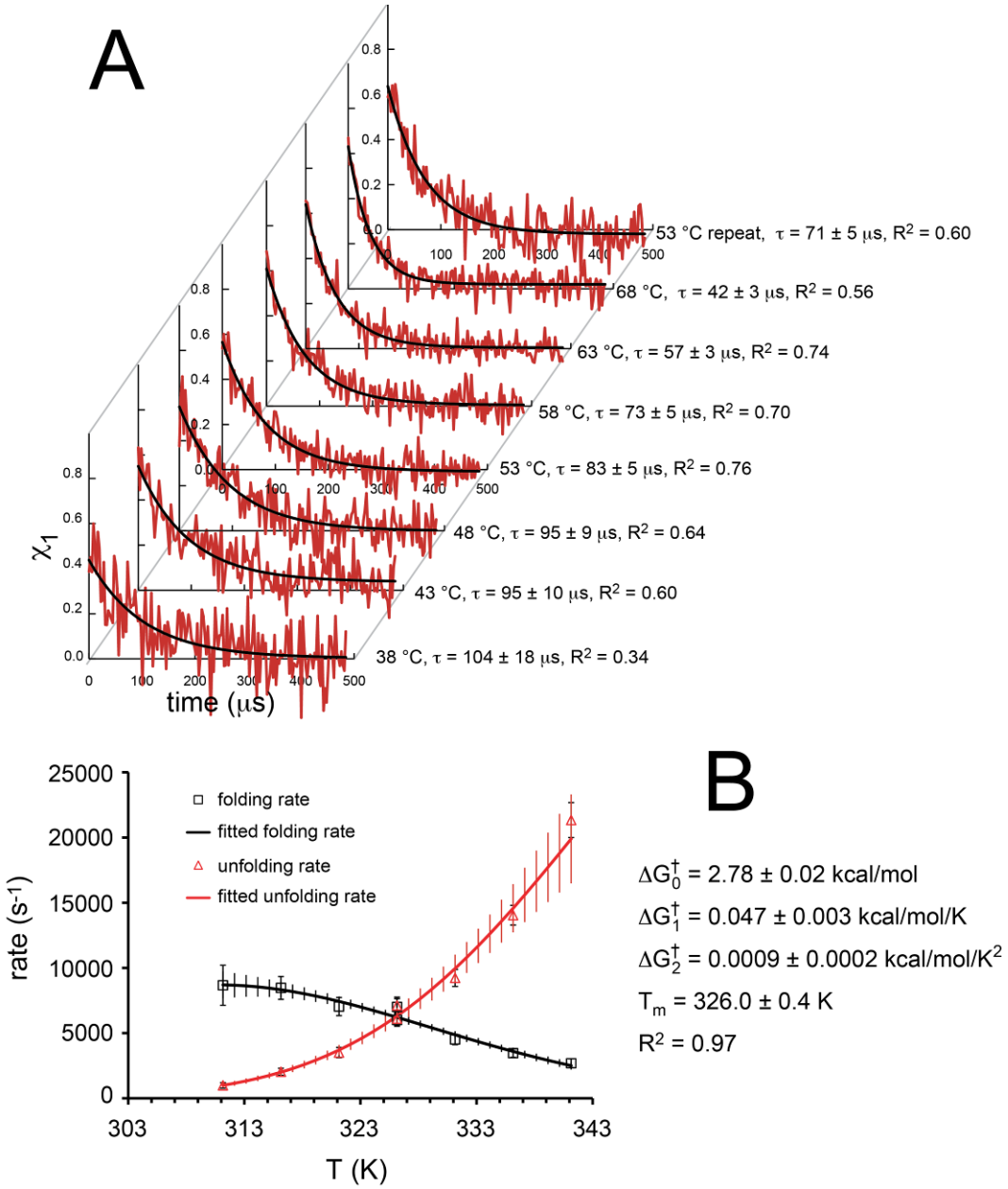


Figure S15. (A) Kinetic relaxation data for Pin WW protein **WW-T** (which has Ser at position 16, Asn at position 19, and Thr at position 21) at 50 μM protein concentration in 20 mM sodium phosphate, pH 7, at several temperatures following temperature jumps of ~12 °C. Red lines show the change in Trp fluorescence for **WW-T** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S15) 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 S16) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW protein **WW-T** as a function of temperature. The black solid line represents the fit of the folding rates to equation S20, 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.

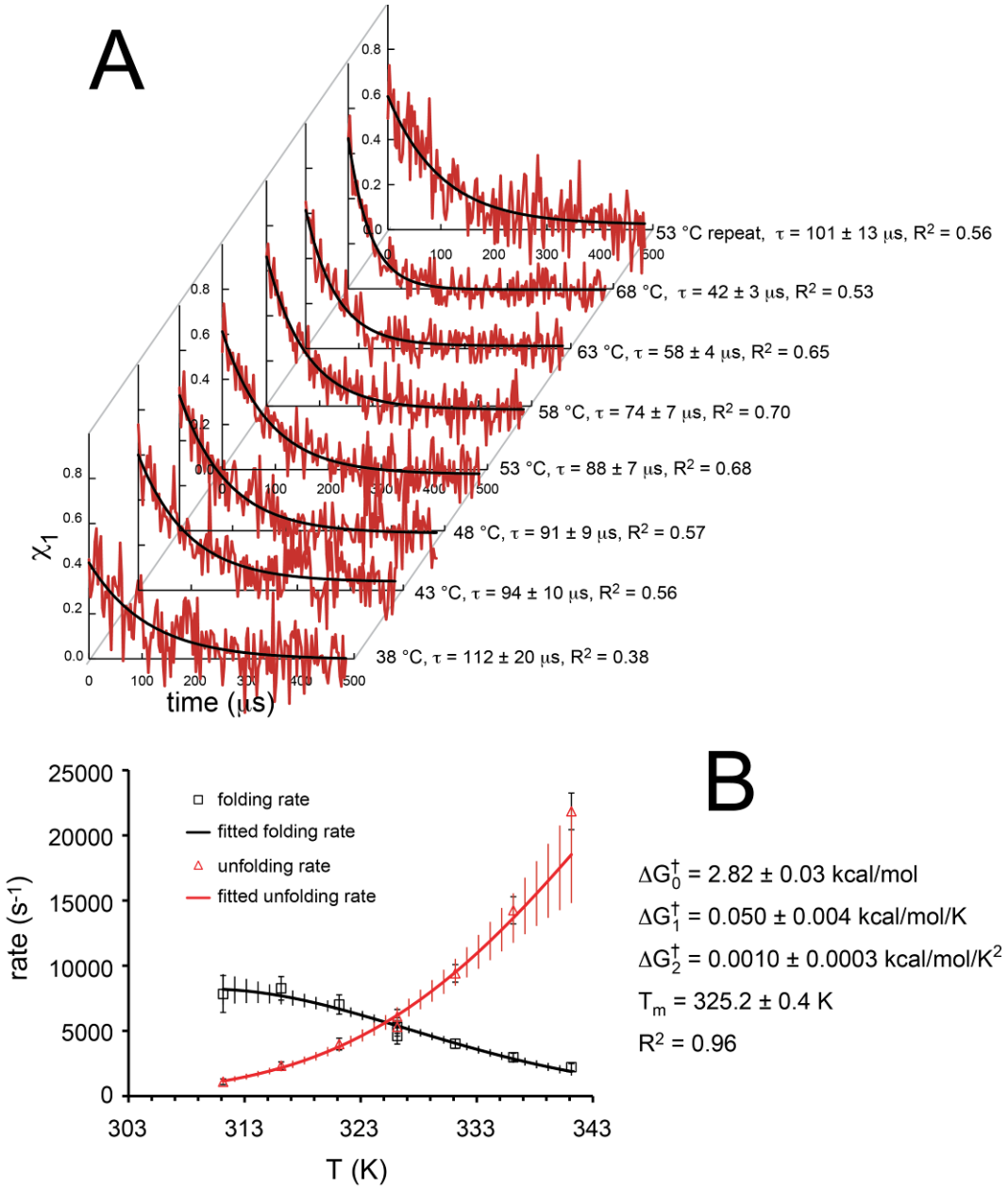


Figure S16. (A) Kinetic relaxation data for Pin WW protein **g-WW-T** (which has Ser at position 16, Asn(GlcNAc) at position 19, and Thr at position 21) at 50 μM protein concentration in 20 mM sodium phosphate, pH 7, at several temperatures following temperature jumps of ~12 °C. Red lines show the change in Trp fluorescence for **g-WW-T** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equation S15) 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 S16) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW protein **g-WW-T** as a function of temperature. The black solid line represents the fit of the folding rates to equation S20, 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.

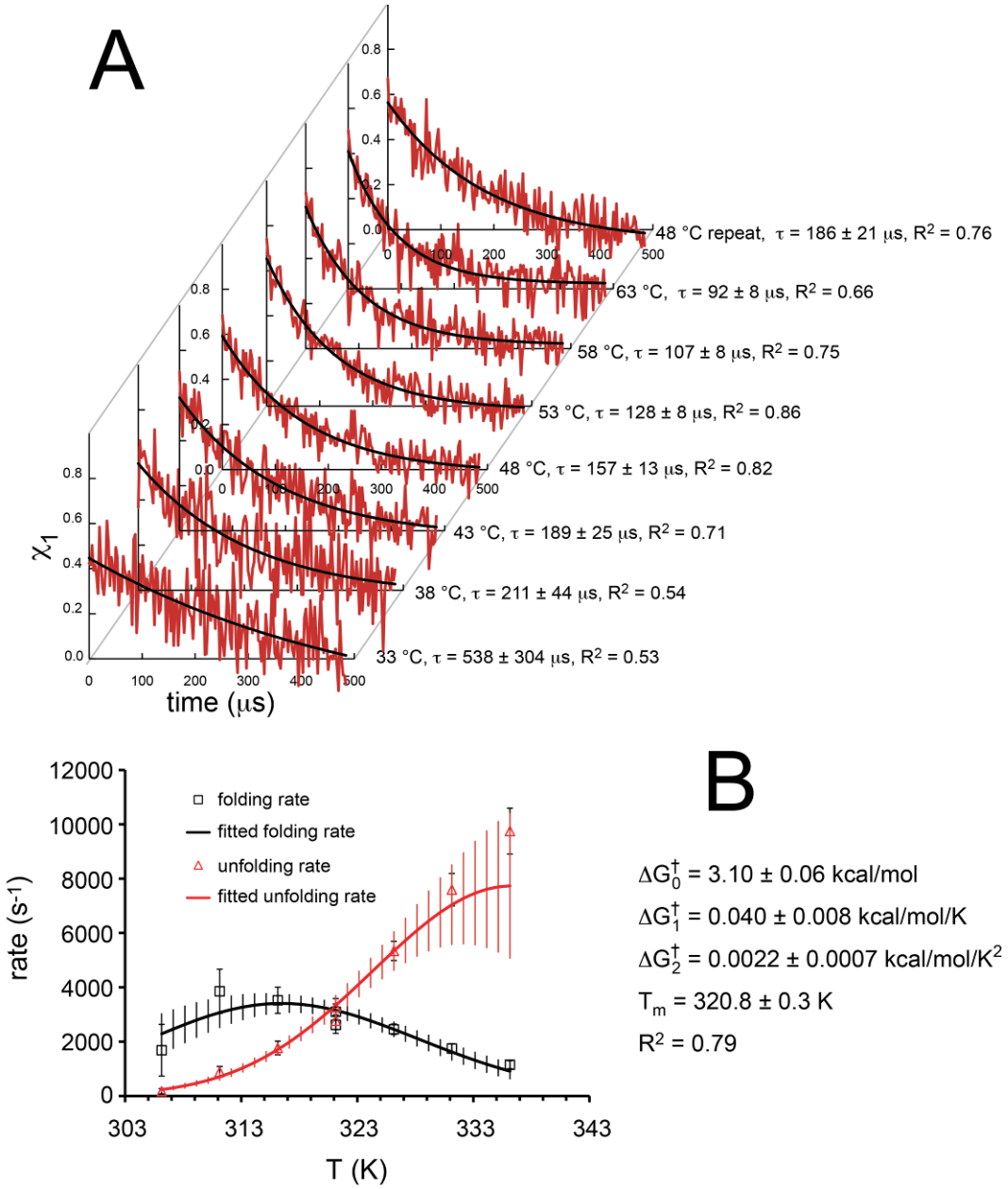


Figure S17. (A) Kinetic relaxation data for Pin WW protein **WW-F,T** (which has Phe at position 16, Asn at position 19, and Thr at position 21) at 50 μM protein concentration in 20 mM sodium phosphate, pH 7, at several temperatures following temperature jumps of ~12 °C . Red lines show the change in Trp fluorescence for **WW-F,T** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equations S15) 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 S16) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW protein **WW-F,T** as a function of temperature. The black solid line represents the fit of the folding rates to equation S20, 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.

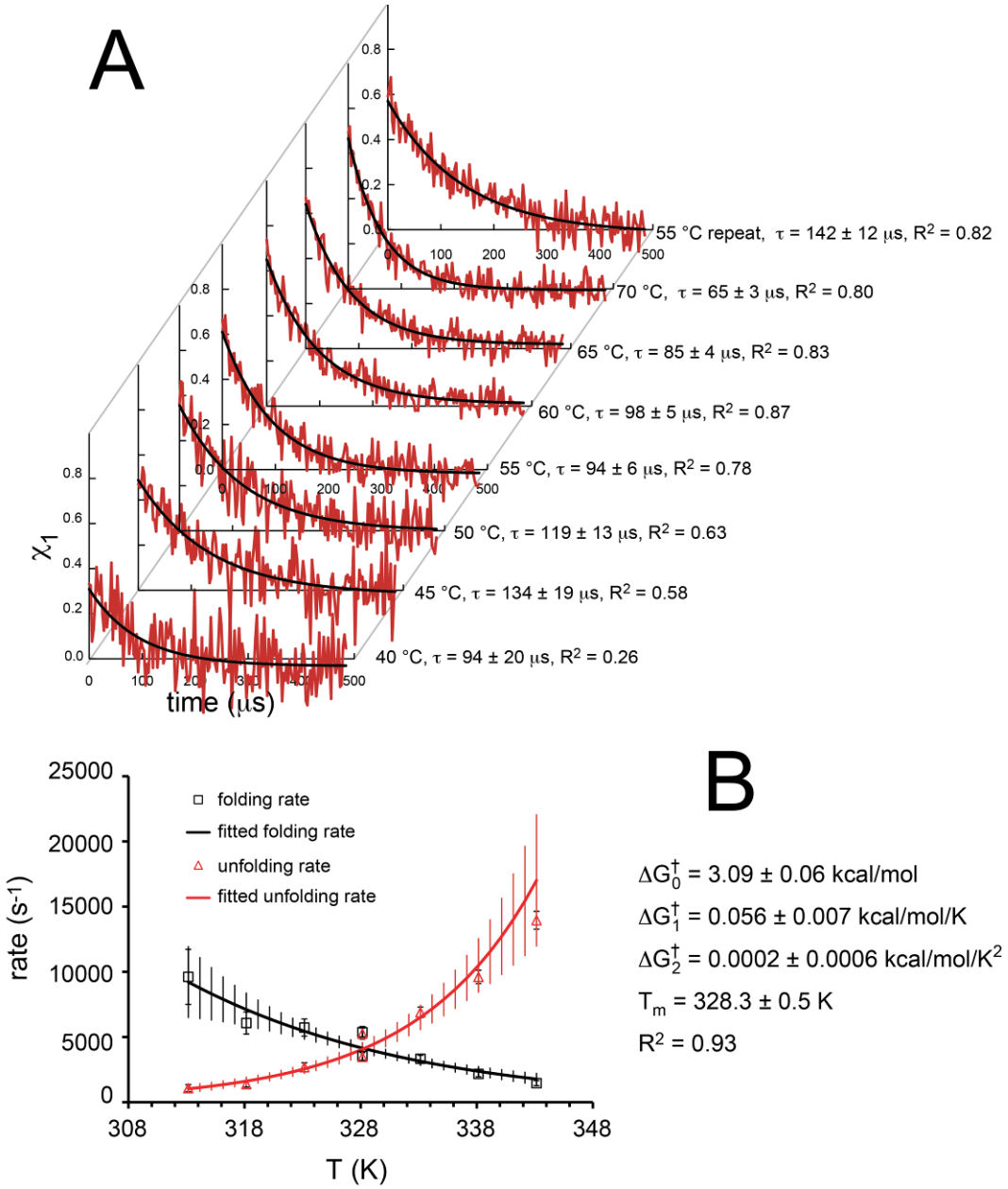


Figure S18. (A) Kinetic relaxation data for Pin WW protein **g-WW-F,T** (which has Phe at position 16, Asn(GlcNAc) at position 19, and Thr at position 21) at 50 μM protein concentration in 20 mM sodium phosphate, pH 7, at several temperatures following temperature jumps of ~12 °C . Red lines show the change in Trp fluorescence for **g-WW-F,T** as the populations of the native and denatured ensembles shift to a new equilibrium at the new temperature, plotted as χ_1 (see equation S15) 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 S16) with relaxation times as indicated. (B) Folding rates (black open squares) and unfolding rates (red open triangles) for Pin WW protein **g-WW-F,T** as a function of temperature. The black solid line represents the fit of the folding rates to equation S20, 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.

References

1. D. F. Wyss *et al.*, *Science* **269**, 1273 (1995).
2. H. J. Hecht *et al.*, *Acta Crystallogr D* **57**, 378 (2001).
3. A. P. West, L. L. LLamas, P. M. Snow, S. Benzer, P. J. Bjorkman, *P Natl Acad Sci USA* **98**, 3744 (2001).
4. G. Guncar, G. Pungercic, I. Klemencic, V. Turk, D. Turk, *Embo J* **18**, 793 (1999).
5. M. R. Sawaya *et al.*, *Cell* **134**, 1007 (2008).
6. M. E. Cuff, K. I. Miller, K. E. van Holde, W. A. Hendrickson, *J Mol Biol* **278**, 855 (1998).
7. D. M. Zajonc, H. Striegl, C. C. Dascher, I. A. Wilson, *P Natl Acad Sci USA* **105**, 17925 (2008).
8. R. Krauspenhaar *et al.*, *BIochem. Biophys. Res. Commun.* **257**, 418 (1999).
9. A. I. Taylor, S. M. Fabiane, B. J. Sutton, R. A. Calvert, *Biochemistry* **48**, 558 (2009).
10. H. E. Klock, S. A. Lesley, *Methods Mol Biol* **498**, 91 (2009).
11. S. R. Hanson *et al.*, *P Natl Acad Sci USA* **106**, 3131 (Mar 3, 2009).
12. M. J. Parker, C. E. Dempsey, M. Lorch, A. R. Clarke, *Biochemistry* **36**, 13396 (Oct 28, 1997).
13. K. Kawahara, C. Tanford, *J Biol Chem* **241**, 3228 (1966).
14. M. Jäger, H. Nguyen, J. C. Crane, J. W. Kelly, M. Gruebele, *J. Mol. Biol.* **311**, 373 (2001).
15. M. Meldal, K. Bock, *Tetrahedron Lett.* **31**, 6987 (1990).
16. L. Otvos, Jr. *et al.*, *Tetrahedron Lett.* **31**, 5889 (1990).
17. S. Ficht, R. J. Payne, R. T. Guy, C.-H. Wong, *Chem. Eur. J.* **14**, 3620 (2008).
18. H. Edelhoch, *Biochemistry* **6**, 1948 (1967).
19. R. M. Ballew, J. Sabelko, C. Reiner, M. Gruebele, *Rev. Sci. Instrum.* **67**, 3694 (1996).
20. R. M. Ballew, J. Sabelko, M. Gruebele, *Proc. Natl. Acad. Sci USA* **93**, 5759 (1996).
21. J. Ervin, J. Sabelko, M. Gruebele, *J. Photochem. Photobiol. sect. B* **54**, 1 (2000).
22. H. A. Kramers, *Physica* **7**, 284 (1940).
23. L. J. Lapidus, W. A. Eaton, J. Hofrichter, *Proc. Natl. Acad. Sci USA* **97**, 7220 (2000).
24. P. Hänggi, P. Talkner, M. Borovec, *Rev. Mod. Phys.* **62**, 251 (1990).
25. O. Bieri *et al.*, *Proc. Natl. Acad. Sci USA* **96**, 9597 (1999).
26. A. Ansari, C. M. Jones, E. R. Henry, J. Hofrichter, W. A. Eaton, *Science* **256**, 1796 (1992).
27. A. A. Fuller *et al.*, *Proc. Natl. Acad. Sci. USA* **106**, 11067 (2009).
28. R. C. Weast, *CRC Handbook of Chemistry and Physics*. (CRC Press, Boca Raton, 1982).