Glycosylation of the enhanced aromatic sequon is similarly stabilizing in three distinct reverse turn contexts

Joshua L. Price[a,b], David L. Powers[c], Evan T. Powers[a,b,1], and Jeffery W. Kelly[a,b,d,1]

[a]Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037; [b]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037; [c]Department of Mathematics, Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699; and [d]Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037

Edited by William F. DeGrado, University of Pennsylvania, Philadelphia, PA, and approved July 5, 2011 (received for review April 12, 2011)

Cotranslational N-glycosylation can accelerate protein folding, slow protein unfolding, and increase protein stability, but the molecular basis for these energetic effects is incompletely understood. N-glycosylation of proteins at naïve sites could be a useful strategy for stabilizing proteins in therapeutic and research applications, but without engineering guidelines, often results in unpredictable changes to protein energetics. We recently introduced the enhanced aromatic sequon as a family of portable structural motifs that are stabilized upon glycosylation in specific reverse turn contexts: a five-residue type I β-turn harboring a G1 β-bulge (using a Phe–Yyy–Asn–Xxx–Thr sequon) and a type II β-turn within a six-residue loop (using a Phe–Yyy–Zzz–Asn–Xxx–Thr sequon) [Culyba EK, et al. (2011) *Science* 331:571–575]. Here we show that glycosylating a new enhanced aromatic sequon, Phe–Asn–Xxx–Thr, in a type I′ β-turn stabilizes the Pin 1 WW domain. Comparing the energetic effects of glycosylating these three enhanced aromatic sequons in the same host WW domain revealed that the glycosylation-mediated stabilization is greatest for the enhanced aromatic sequon complementary to the type I β-turn with a G1 β-bulge. However, the portion of the stabilization from the tripartite interaction between Phe, Asn(GlcNAc), and Thr is similar for each enhanced aromatic sequon in its respective reverse turn context. Adding the Phe–Asn–Xxx–Thr motif (in a type I′ β-turn) to the enhanced aromatic sequon family doubles the number of proteins that can be stabilized by glycosylation without having to alter the native reverse turn type.

glycoprotein | biopharmaceutical | conjugate | native state stabilization | β-sheet

Nearly 1/3 of the eukaryotic proteome traverses the cellular secretory pathway (1). Many of these proteins are cotranslationally N-glycosylated at Asn residues within the conserved Asn–Xxx–Thr/Ser sequon. The oligosaccharyl transferase enzyme complex transfers the $Glc_3Man_9GlcNAc_2$ (where Glc is glucose, Man is mannose, and GlcNAc is N-acetylglucosamine) oligosaccharide as one structural unit onto the nascent polypeptide as it is translocated into the endoplasmic reticulum (2–4). The N-linked oligosaccharide, or N-glycan, promotes glycoprotein folding in the endoplasmic reticulum by allowing the glycoprotein to enter the calnexin/calreticulin-assisted folding vs. degradation cycle (5, 6). Glycans can also intrinsically accelerate protein folding (7, 8), promote secondary structure formation (9–24), and increase protein thermodynamic stability (8, 25, 26), but the molecular basis for these effects is incompletely understood (27).

We recently demonstrated that glycosylation of an Asn residue within the sequence Phe–Yyy–Asn–Xxx–Thr (where Yyy can likely be any amino acid and Xxx is any amino acid but Pro) stabilizes the glycosylation-naïve rat CD2 adhesion domain (RnCD2ad) and human muscle acylphosphatase (AcyP2) by approximately −2 kcal mol^{-1}, provided that Asn is located at the $i+2$ position of a type I β-turn with a G1 β-bulge (28, 29), hereafter called a type I β-bulge turn (30). Published structural data (31) from the human ortholog of RnCD2ad (HsCD2ad, Fig. 1*A*) suggest that placement of an N-glycan at $i+2$ in the type I β-bulge turn context allows the α-face of GlcNAc1 of the N-glycan to engage in stabilizing hydrophobic interactions with the aromatic ring of Phe at the i position (a stabilizing C-H/π interaction may also play a role), and the side-chain methyl group of Thr at the $i+4$ position (32). We call the Phe–Yyy–Asn–Xxx–Thr sequence an "enhanced aromatic sequon" because of its increased propensity to form a stabilizing compact structure upon glycosylation, relative to the canonical Asn–Xxx–Thr sequon, and because it is more efficiently glycosylated by the cell than the Asn–Xxx–Thr sequon (30).

Nearly 9% of the reverse turns in the Protein Data Bank (PDB) are type I β-bulge turns (28, 33), so installing the Phe–Yyy–Asn–Xxx–Thr enhanced aromatic sequon could be an attractive strategy for increasing the stability of the many proteins that harbor type I β-bulge turns. Identifying other suitable reverse turn types that could position Phe, GlcNAc1, and Thr close enough to facilitate a tripartite interaction would further expand the number of proteins that could benefit from the increased stability and possibly the increased glycosylation efficiency afforded by the enhanced aromatic sequon.

Our previous efforts to apply the enhanced aromatic sequon to other types of reverse turns employed the WW domain of the human protein Pin1 (WW) (30), a glycosylation-naïve protein in which three antiparallel β-strands are connected by two reverse turns (34). In wild-type WW, loop 1 adopts an unusual six-residue hydrogen-bonded loop harboring an internal type II β-turn (Fig. 1*B*); 0.1% of the reverse turns in the PDB have this conformation (33). The side-chain beta carbons (Cβ's) at the $i, i+3$, and $i+5$ positions of this loop are within 4.0–4.6 Å of each other; close enough to facilitate a stabilizing interaction between Phe, GlcNAc1 and Thr, similar to the interactions observed in the glycosylated type I β-bulge turn of HsCD2ad (Fig. 1*A*). Grafting these residues onto the i, $i+3$, and $i+5$ positions of the six-residue loop generated a modified enhanced aromatic sequon with the sequence Phe–Yyy–Zzz–Asn–Xxx–Thr (where Yyy and Zzz can likely be any amino acid).

Chemical glycosylation of the Phe–Yyy–Zzz–Asn–Xxx–Thr sequon (with a single GlcNAc, GlcNAc1) in the six-residue loop of WW increased the stability of the resulting WW variant by −0.7 kcal mol^{-1} (30), a smaller effect than observed for the Phe–Yyy–Asn–Xxx–Thr sequon in the type I β-bulge turns of

Author contributions: J.L.P., E.T.P., and J.W.K. designed research; J.L.P. performed research; J.L.P., D.L.P., E.T.P., and J.W.K. analyzed data; J.L.P., E.T.P., and J.W.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: jkelly@scripps.edu or epowers@scripps.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1105880108/-/DCSupplemental.

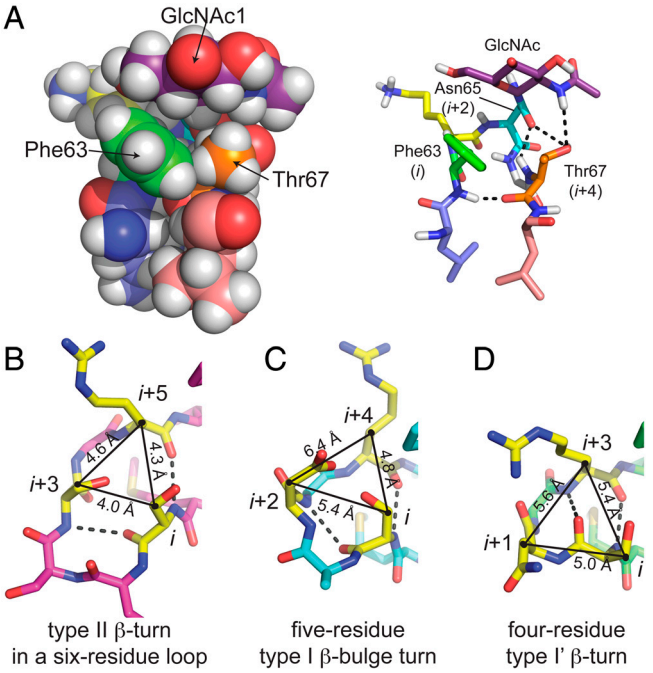


Fig. 1. Matching enhanced aromatic sequons with reverse turn conformations that can facilitate stabilizing interactions among Phe, Asn(GlcNAc1), and Thr. (*A*) Glycosylated five-residue type I β-bulge turn from the adhesion domain of the human protein CD2 (PDB accession code 1GYA; ref. 31). (*Left*) CPK representation; (*Right*) stick representation. (*B*) Type II β-turn in a six-residue loop (PDB accession code 1PIN; ref. 34) (*C*) five-residue type I β-bulge turn (PDB accession code 2F21; ref. 36), and (*D*) four-residue type I′ β-turn (PDB accession code 1ZCN; ref. 36) from variants of the WW domain of human protein Pin1. Structures are rendered in Pymol, with dotted lines depicting hydrogen bonds. Amino acid positions within the WW domain where we incorporated components of the enhanced aromatic sequon are highlighted in yellow. Interatomic distances between the side-chain beta carbons (Cβ's) at these positions are indicated with black lines, the distances in Å are depicted in black font.

RnCD2 and AcyP2 ($\Delta\Delta G_f = -1.8$ kcal mol^{-1}, -2.5 kcal mol^{-1}, respectively). One possible interpretation of these results is that the type II β-turn within the six-residue loop does not promote the stabilizing tripartite interaction between Phe, GlcNAc, and Thr as effectively as does the five-residue type I β-bulge turn. However, key host context differences between WW, RnCD2, and AcyP2 could also be partially responsible for these observations, including differences in folding topology and mechanism (35), and differences in the amino acids that flank the glycosylated reverse turns (30). Moreover, because the WW domain is synthesized chemically via a solid-phase strategy, the N-glycan in WW (GlcNAc) is much smaller than the N-glycans in RnCD2 (oligomannose) and AcyP2 (fucosylated paucimannose). Interactions between the host sequences and these extended glycans could also contribute to the stabilization associated with glycosylating the Phe–Yyy–Asn–Xxx–Thr sequon in the type I β-bulge turns within RnCD2 and AcyP2.

The studies reported herein have two aims: To add to the number of reverse turn types into which stabilizing enhanced aromatic sequons can be placed and to enable direct stability comparisons of glycosylated enhanced aromatic sequons in different reverse turn types within the same host protein. To meet these goals, we sought a single protein into which several types of enhanced aromatic sequons and their corresponding reverse turn types could be inserted without changing the overall structure or the flanking sequences. The WW domain is ideal for these requirements: Many WW variants harboring different reverse turn types in loop 1 have been structurally (34, 36, 37) and biophysically (36–39) characterized. Crystal structures exist for WW domains harboring a type II β-turn in a six-residue loop (Fig. 1*B*), a five-residue type I β-bulge turn (Fig. 1*C*), and a four-residue type I′ β-turn (Fig. 1*D*) as loop 1 (34, 36). We hypothesized that a type I′ β-turn, which makes up 11% of the reverse turns in the PDB (33), would also be an additional conformational host for a complementary enhanced aromatic sequon: the Cβ's of the side chains at the $i, i+1$, and $i+3$-positions are close enough (<5.6 Å; see Fig 1*D*) to support a stabilizing tripartite interaction among Phe, Asn (GlcNAc1), and Thr. Importantly, the chemical synthesis of homogeneously glycosylated (40) WW domains is efficient (30, 41) enabling numerous analogs to be prepared, each having an identical N-glycan (in this case, GlcNAc1). Our data show that type I′ β-turns are suitable conformational hosts for a stabilizing enhanced aromatic sequon. This result significantly expands the scope of protein stabilization by glycosylating enhanced aromatic sequons. Furthermore, our data show that the order of stabilization by glycosylating enhanced aromatic sequons in the different turn types is: type I β-bulge turns > type II β-bulge turns in a six-residue loop > type I′ β-turns.

Results

Strategy and Nomenclature. The four-, five-, and six-residue reverse turns comprising loop 1 of WW were converted to their corresponding enhanced aromatic sequons by replacing the amino acid at position 16 (Ser in all cases) with Phe, replacing the amino acid at position 19 (Asn, Asp, or Ser, respectively) with Asn (GlcNAc1), and replacing the amino acid at position 21 (Arg in all cases) with Thr (36, 38). Note that we are using the same number to indicate amino acids in analogous positions in WW variants with different loop 1 lengths (36, 37). Thus, the sequences of the enhanced aromatic sequons in the four-, five-, and six-residue reverse turns comprising loop 1 are Phe16–Asn(GlcNAc1)19–Gly20–Thr21, Phe16–Ala18–Asn(GlcNAc1)19–Gly20–Thr21, and Phe16–Arg17–Ser18-–Asn(GlcNAc1)19–Gly20–Thr21, respectively (see Table 1).

The stabilizing effect of glycosylating enhanced aromatic sequons can be quantified by comparing the stabilities of WW variants with glycosylated enhanced aromatic sequons to the stabilities of their nonglycosylated counterparts. The contributions of two- and three-way interactions amongst the Phe16, Asn(GlcNAc1)19, and Thr21 side chains to the overall stabilizing effect of glycosylation can be estimated using triple mutant cycle analyses, as we have done previously (30). This parsing of stabilization energies through energetic comparisons was accomplished by replacing Phe16, Asn(GlcNAc1)19, and Thr21 with Ser16, Asn19, and Arg21, respectively, in every possible combination, for a total of eight proteins in each of the three correlated enhanced aromatic sequon-reverse turn contexts. The results of these analyses are described in the sections below.

The WW variants are named by the number of amino acids in the loop 1 reverse turn, followed by the letter "**g**" if the variant is N-glycosylated on Asn19, the letter "**F**" if it has Phe at position 16, and the letter "**T**" if it has Thr at position 21. The lack of the letters **g**, **F**, and/or **T** indicates that the variant is not N-glycosylated on Asn19, that position 16 is Ser, and/or that position 21 is Arg, respectively. For example, variant **4g-F,T** has a 4-residue loop 1 type I′ β-turn, with Asn(GlcNAc1) at position 19, Phe at position 16, and Thr at position 21. Variant **4** has a 4-residue loop 1 type I′ β-turn, with Asn at position 19, Ser at position 16, and Arg at position 21 (see Table 1 for the names of the WW variants studied).

Stabilization from Glycosylating Enhanced Aromatic Sequons. To quantify the stabilizing effect of glycosylating enhanced aromatic sequons in loop 1 of the corresponding four-, five-, and six-residue reverse turns, we used variable temperature circular dichroism (CD) spectropolarimetry to analyze the thermodynamic stability of WW variants **4-F,T**, **4g-F,T**, **5-F,T**, **5g-F,T**, **6-F,T**, and

Table 1. Sequences, melting temperatures, and folding free energies of Pin WW-derived glycoproteins and their nonglycosylated counterparts*

Protein	Sequence†	T_m (°C)	ΔT_m (°C)	ΔG_f (kcal/mol)	$\Delta\Delta G_f$ (kcal/mol)
	15 21				
4	MS–NGR	64.4 ± 0.4		0.06 ± 0.04	
			2.2 ± 0.6		-0.23 ± 0.06
4g	MS–N̲GR	66.6 ± 0.4		-0.17 ± 0.04	
4-F	MF–NGR	66.7 ± 0.5		-0.18 ± 0.08	
			1.5 ± 0.7		-0.18 ± 0.08
4g-F	MF–N̲GR	68.2 ± 0.5		-0.36 ± 0.05	
4-T	MS–NGT	62.2 ± 0.4		0.30 ± 0.04	
			-0.8 ± 0.6		0.07 ± 0.07
4g-T	MS–N̲GT	61.4 ± 0.5		0.37 ± 0.05	
4-F,T	MF–NGT	63.5 ± 0.3		0.18 ± 0.03	
			3.2 ± 0.7		-0.39 ± 0.09
4g-F,T	MF–N̲GT	66.7 ± 0.6		-0.21 ± 0.08	
	15 21				
5	MS-ANGR	68.7 ± 0.2		-0.38 ± 0.02	
			0.6 ± 0.3		-0.07 ± 0.04
5g	MS-AN̲GR	69.3 ± 0.2		-0.46 ± 0.03	
5-F	MF-ANGR	65.2 ± 0.3		-0.02 ± 0.03	
			5.0 ± 0.4		-0.55 ± 0.04
5g-F	MF-AN̲GR	70.3 ± 0.2		-0.58 ± 0.02	
5-T	MS-ANGT	68.9 ± 0.2		-0.42 ± 0.02	
			2.4 ± 0.3		-0.23 ± 0.03
5g-T	MS-AN̲GT	71.3 ± 0.3		-0.65 ± 0.03	
5-F,T	MF-ANGT	66.0 ± 0.2		-0.11 ± 0.02	
			9.2 ± 0.2		-0.94 ± 0.03
5g-F,T	MF-AN̲GT	75.2 ± 0.2		-1.05 ± 0.02	
	15 21				
6	MSRSNGR	56.2 ± 0.3		0.95 ± 0.04	
			-2.6 ± 0.4		0.21 ± 0.06
6g	MSRSN̲GR	53.6 ± 0.3		1.16 ± 0.04	
6-F	MFRSNGR	51.0 ± 0.3		1.45 ± 0.06	
			0.7 ± 0.4		-0.17 ± 0.08
6g-F	MFRSN̲GR	51.7 ± 0.3		1.28 ± 0.04	
6-T	MSRSNGT	52.5 ± 0.3		1.22 ± 0.05	
			-0.2 ± 0.5		0.04 ± 0.07
6g-T	MSRSN̲GT	52.3 ± 0.3		1.26 ± 0.05	
6-F,T	MFRSNGT	47.4 ± 0.4		1.72 ± 0.09	
			7.6 ± 0.5		-0.70 ± 0.10
6g-F,T	MFRSN̲GT	55.0 ± 0.3		1.02 ± 0.04	

*Tabulated data are given as mean ± standard error at 65 °C for WW variants at 10 μM in 20 mM aqueous sodium phosphate, pH 7.
†N̲ = Asn(glycan).

6g-F,T. CD data for **6-F,T** and **6g-F,T** and their derivatives (described below) have been published previously at a protein concentration of 50 μM (30), but were restudied herein at a protein concentration of 10 μM (the energetic data are comparable at both concentrations) to facilitate direct comparisons with **4-F,T**, **4g-F,T**, **5-F,T**, and **5g-F,T** and their derivatives (some of which were not completely soluble at 50 μM). Table 1 shows the melting temperature T_m and free energy of folding ΔG_f (at 65 °C) for each protein and corresponding glycoprotein, along with the effect of glycosylation on the T_m and ΔG_f (at 65 °C) for each protein (complete CD spectra and variable temperature CD data for **4**, **5**, and **6** and their derivatives are shown in Figs. S1–S3, respectively). We use 65 °C as a reference temperature because it is within the transition regions of all the variants studied herein. Extrapolating ΔG_f to temperatures outside the transition region using thermodynamic parameter estimates from fits to variable temperature CD data is unreliable (because errors in ΔC_p, the least-well defined parameter from such fits, become magnified outside the transition region; see *SI Text* and Figs. S1–S3). For sets of proteins with similar ΔC_p values, the differences between their T_m values should reflect the differences between their ΔG_f values both at 65 °C and at lower temperatures.

The T_m of glycoprotein **4g-F,T** is 3.2 ± 0.7°C higher than that of protein **4-F,T** ($\Delta\Delta G_f = -0.39 \pm 0.09$ kcal mol^{-1} at 65 °C), indicating that glycosylating the Phe–Asn–Xxx–Thr enhanced aromatic sequon in the context of a four-residue type I′ β-turn stabilizes WW. Glycosylating the Phe–Yyy–Asn–Xxx–Thr sequon in the context of the five-residue type I β-turn also stabilizes WW ($\Delta T_m = 9.2 \pm 0.2$°C, $\Delta\Delta G_f = -0.94 \pm 0.03$ kcal mol^{-1} at 65 °C), as does glycosylating the Phe–Yyy–Zzz–Asn–Xxx–Thr sequon in the type II β-turn in a six-residue loop ($\Delta T_m = 7.6 \pm 0.5$°C, $\Delta\Delta G_f = -0.70 \pm 0.10$ kcal mol^{-1} at 65 °C). These data indicate that the Phe–Yyy–Asn–Xxx–Thr enhanced aromatic sequon corresponding to the five-residue type I β-bulge turn is, overall, the best for stabilizing WW amongst those studied here.

Interaction Energies in Enhanced Aromatic Sequons from Triple Mutant Cycle Analysis. To determine whether the Phe, Asn (GlcNAc1) and Thr side chains interact similarly in each correlated enhanced aromatic sequon/reverse turn context, we measured the thermodynamic stabilities of each WW variant in the four-, five-, and six-residue reverse turn groups in Table 1. The data from each group of eight WW variants make a triple mutant cycle (Fig. 2). Triple mutant cycles contain more information than conventional double mutant cycles, because each of the six "faces" of a triple mutant cycle "cube" is itself a double mutant cycle (42). Whereas double mutant cycles provide information about the energetic impact of an interaction between two residues, a triple mutant cycle provides information about the energetic impact of the two- and three-way interactions.

Extracting this information from a triple mutant cycle is straightforward, and begins with analyzing the double mutant cycle faces of the triple mutant cycle cube (Fig. 2). The double mutant cycle formed by proteins **4** and **4-F** and glycoproteins **4g** and **4g-F** (the front face of the triple mutant cycle cube in Fig. 2*A*), reveals that glycosylation of Asn19 (in the presence of Arg21) stabilizes glycoprotein **4g** relative to protein **4** ($\Delta\Delta G_{f,1} = -0.23 \pm 0.06$ kcal mol^{-1} at 65 °C). Similarly, glycosylation of Asn19 (in the presence of Arg 21) stabilizes **4g-F** relative to **4g** ($\Delta\Delta G_{f,2} = -0.18 \pm 0.08$ kcal mol^{-1} at 65 °C). The difference between $\Delta\Delta G_{f,2}$ and $\Delta\Delta G_{f,1}$ ($\Delta\Delta\Delta G_{f,\text{front}} = 0.05 \pm 0.10$ kcal mol^{-1} at 65 °C) indicates that changing Ser16 to Phe16 (while keeping Arg21 constant) does not significantly change the effect of glycosylating Asn19 in the four-residue type I′ β-turn. In other words, Phe16 and Asn(GlcNAc1)19 do not interact favorably in **4g-F**.

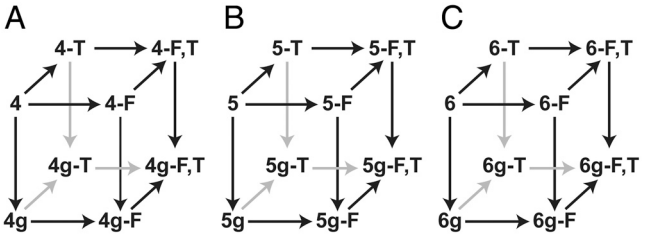


Fig. 2. Triple mutant cycle cubes formed by protein **4**, glycoprotein **4g**, and their derivatives (*A*); protein **5**, glycoprotein **5g**, and their derivatives (*B*); and protein **6**, glycoprotein **6g**, and their derivatives (*C*).

Changing Arg21 to Thr21 changes this trend. The double mutant cycle formed by proteins **4-T**, **4g-T**, **4-F,T**, and **4g-F,T** (the back face of the triple mutant cycle cube shown in Fig. 2*A*) reveals that in the presence of Thr21 (instead of Arg21), Phe16 and Asn(GlcNAc1) interact favorably ($\Delta\Delta\Delta G_{f,\mathrm{back}} = -0.46 \pm 0.11$ kcal mol^{-1} at 65 °C). The difference between the front and back double mutant cycles is an estimate of the energy of the three-way interaction between Phe16, Asn(GlcNAc1)19, and Thr21. The large difference between $\Delta\Delta\Delta G_{f,\mathrm{back}}$ and $\Delta\Delta\Delta G_{f,\mathrm{front}}$ for the four-residue type I′ β-turn ($\Delta\Delta\Delta\Delta G_f = -0.51 \pm 0.15$ kcal mol^{-1} at 65 °C) indicates that Phe16, Asn(GlcNAc1) 19 and Thr21 engage in a favorable three-way interaction in **4g-F,T**.

Similar analyses of the triple mutant cycles formed by proteins **5** and **6** and their derivatives (Fig. 2 *B* and *C*) reveal a favorable interaction between Phe16, Asn(GlcNAc1)19, and Thr21 in the five-residue type I β-bulge turn ($\Delta\Delta\Delta\Delta G_f = -0.23 \pm 0.07$ kcal mol^{-1} at 65 °C) and in the type II β-turn in a six-residue loop ($\Delta\Delta\Delta\Delta G_f = -0.36 \pm 0.15$ kcal mol^{-1} at 65 °C). This three-way interaction between Phe16, Asn(GlcNAc1)19, and Thr21 is similarly favorable in each reverse turn context (perhaps more favorable in the type I′ β-turn than in the type I β-bulge turn, but recall that this is only part of the overall stabilizing effect of N-glycosylation).

The attribution of $\Delta\Delta\Delta G_{f,\mathrm{front}}$ and $\Delta\Delta\Delta G_{f,\mathrm{back}}$ values to the interaction between Phe16 and Asn(GlcNAc1)19, and of $\Delta\Delta\Delta\Delta G_f$ to the tripartite interaction among Phe16, Asn(GlcNAc1)19, and Thr21, assumes that the Ser16 side chain does not interact with the side chains at positions 19 or 21, and that the Arg21 side chain does not interact with the side chains at positions 16 or 19, in any variant. This assumption is, to a first approximation, consistent with the available structural data. Crystal structures of WW in the context of the full-length Pin1 protein (34, 36) show that the side chains at positions 16, 19, and 21 generally interact only with solvent or the main chain (see Fig. 1 *B–D*). The lone exception is an interaction between the side-chain hydroxyl of Ser16 and the side-chain carboxylate of Asp19 in the type I β-bulge turn (Fig. 1*C*). However, the equivalent interaction (between the Ser16 hydroxyl and the Asn19 side-chain carbonyl) in the variants of **5** that have Ser at position 16 (**5**, **5g**, **5-T**, and **5g-T**) should be the same whether Asn19 is N-glycosylated or not, and thus should not affect our analysis. We also note that the reverse turn structures are likely to depend primarily on loop length and the identities of a few key residues (e.g., Asn19 and Gly20 in the variants of **4** and Gly20 in the variants of **5**, because these amino acids are strongly favored in these positions of type I′ β-turns and type I β-bulge turns, respectively) (28, 36). Because these factors are kept constant within the variants that make up each triple mutant cycle, the corresponding reverse turn structures should remain roughly constant as well.

Estimates for Interaction Energies from Linear Regression on Triple Mutant Cycle Data. We used least-squares regression to extract additional information about interactions amongst Phe, Asn (GlcNAc1), and Thr from the triple mutant cycles formed by WW variant groups **4**, **5**, and **6**. We fit the folding free energy data (at 65 °C) from the triple mutant cycle formed by **4** and its derivatives to the following equation:

$$\begin{aligned}\Delta G_f = \Delta G_f^o &+ C_{\mathrm{F}} \cdot W_{\mathrm{F}} + C_{\underline{\mathbf{N}}} \cdot W_{\underline{\mathbf{N}}} + C_{\mathrm{T}} \cdot W_{\mathrm{T}} \\ &+ C_{\mathrm{F},\underline{\mathbf{N}}} \cdot W_{\mathrm{F}} \cdot W_{\underline{\mathbf{N}}} + C_{\mathrm{F,T}} \cdot W_{\mathrm{F}} \cdot W_{\mathrm{T}} \\ &+ C_{\underline{\mathbf{N}},\mathrm{T}} \cdot W_{\underline{\mathbf{N}}} \cdot W_{\mathrm{T}} + C_{\mathrm{F},\underline{\mathbf{N}},\mathrm{T}} \cdot W_{\mathrm{F}} \cdot W_{\underline{\mathbf{N}}} \cdot W_{\mathrm{T}}. \quad [1]\end{aligned}$$

Eq. **1** shows how the ΔG_f of a given variant of **4** is related to the average $\Delta G_f{}^o$ of **4**, plus a series of correction terms that account for the interactions amongst the amino acids at positions 16, 19, and 21. Each correction term is a product of one or more indicator variables W (which reflect whether a mutation is present in the given variant) and a free energy contribution factor C. W_{F} is 0 when position 16 is Ser or 1 when it is Phe; $W_{\underline{\mathbf{N}}}$ is 0 when position 19 is Asn or 1 when it is Asn(GlcNAc1); W_{T} is 0 when position 21 is Arg or 1 when it is Thr. C_{F}, $C_{\underline{\mathbf{N}}}$, and C_{T} describe the energetic consequences of the Ser16 to Phe16, Asn19 to Asn (GlcNAc1)19, and Arg21 to Thr21 mutations, respectively. These energies probably primarily reflect the difference in conformational preferences between Ser and Phe at position 16, Asn and Asn(GlcNAc1) at position 19, and Arg and Thr at position 21. $C_{\mathrm{F},\underline{\mathbf{N}}}$, $C_{\mathrm{F,T}}$, and $C_{\underline{\mathbf{N}},\mathrm{T}}$ describe the free energies of the two-way interactions between Phe16 and Asn(GlcNAc1)19, between Phe16 and Thr21, and between Asn(GlcNAc1)19 and Thr21, respectively. $C_{\mathrm{F},\underline{\mathbf{N}},\mathrm{T}}$ describes the energetic impact of the three-way interaction between Phe16, Asn(GlcNAc1)19, and Thr21. $C_{\mathrm{F},\underline{\mathbf{N}}}$, $C_{\mathrm{F,T}}$, $C_{\underline{\mathbf{N}},\mathrm{T}}$, and $C_{\mathrm{F},\underline{\mathbf{N}},\mathrm{T}}$ are essentially equivalent to the two- and three-way interaction energies ($\Delta\Delta\Delta G_f$ and $\Delta\Delta\Delta\Delta G_f$ values) that could be calculated by a conventional analysis (e.g., as in the preceding section) of the triple mutant cycle data (42), but obtaining them by regression is more convenient, and automatically gives their standard errors in the regression output. Similar analyses were performed for **5** and **6** and their derivatives, and the results are shown in Table 2. Note that the caveats to the conventional analysis of triple mutant cycle data mentioned in the preceding section apply to this analysis as well.

According to Eq. **1**, the stabilizing effect of glycosylating the enhanced aromatic sequon in **4-F,T** [$\Delta\Delta G_f = \Delta G_f$(**4g-F,T**)-$\Delta G_f$(**4-F,T**)] is equal to the sum of the corresponding values of $C_{\underline{\mathbf{N}}}$, $C_{\mathrm{F},\underline{\mathbf{N}}}$, $C_{\underline{\mathbf{N}},\mathrm{T}}$, and $C_{\mathrm{F},\underline{\mathbf{N}},\mathrm{T}}$. The same is true for **5-F,T** and **6-F,T**. Thus, by comparing $C_{\underline{\mathbf{N}}}$, $C_{\mathrm{F},\underline{\mathbf{N}}}$, $C_{\underline{\mathbf{N}},\mathrm{T}}$, and $C_{\mathrm{F},\underline{\mathbf{N}},\mathrm{T}}$ values we can trace the origins of the stabilizing effect of glycosylating the enhanced aromatic sequon in each reverse turn context (Fig. 3).

Changing Asn19 to Asn(GlcNAc1)19 affects each turn type differently: it stabilizes the four-residue type I′ β-turn ($C_{\underline{\mathbf{N}}} = -0.23 \pm 0.08$ kcal mol^{-1}), does not affect the five-residue type I β-bulge turn substantially ($C_{\underline{\mathbf{N}}} = -0.07 \pm 0.06$ kcal mol^{-1}), and destabilizes the type II β-turn within a six-residue loop ($C_{\underline{\mathbf{N}}} = 0.21 \pm 0.06$ kcal mol^{-1}). It is possible that Asn(GlcNAc1) has backbone dihedral angle preferences that are more compatible with the $i+1$ position of a type I′ β-turn than with the $i+2$ position of a type I β-bulge turn or with the $i+3$ position of a type II β-turn. If so, such preferences would differ substantially from those of Asn itself (43), which is favored at $i+1$ in a type I′ β-turn, and at $i+2$ in a type I β-turn, but not at $i+3$ in a type II β-turn (44).

The two-way interaction between Phe16 and Asn(GlcNAc1)19 stabilizes the five-residue type I β-bulge turn ($C_{\mathrm{F},\underline{\mathbf{N}}} = -0.48 \pm 0.09$ kcal mol^{-1}) and the type II β-turn within a the six-residue loop ($C_{\mathrm{F},\underline{\mathbf{N}}} = -0.38 \pm 0.08$ kcal mol^{-1}), but does not substantially change the stability of the four-residue type I′ β-turn ($C_{\mathrm{F},\underline{\mathbf{N}}} = 0.05 \pm 0.11$ kcal mol^{-1}). These differences appear not to correlate with differences among the Cβ-Cβ distances between positions 16 and 19 in the four-, five-, and six-residue turns (Fig. 1 *B–D*),

Table 2. Triple mutant cycle analysis of folding free energy data at 65 °C (338.15 K) for glycosylated and nonglycosylated WW variants harboring either a four-, five-, or six-residue reverse turn in loop 1*

	Type I′ β-turn	Type I β-bulge turn	Type II β-turn in six-residue loop
ΔG_f^o	0.06 ± 0.06 (0.287)	-0.38 ± 0.04 (0.000)	0.95 ± 0.04 (0.000)
C_{F}	-0.24 ± 0.08 (0.005)	0.36 ± 0.06 (0.000)	0.50 ± 0.06 (0.000)
$C_{\underline{\mathbf{N}}}$	-0.23 ± 0.08 (0.009)	-0.07 ± 0.06 (0.248)	0.21 ± 0.06 (0.000)
C_{T}	0.23 ± 0.08 (0.010)	-0.04 ± 0.06 (0.557)	0.27 ± 0.06 (0.000)
$C_{\mathrm{F},\underline{\mathbf{N}}}$	0.05 ± 0.11 (0.661)	-0.48 ± 0.09 (0.000)	-0.38 ± 0.08 (0.000)
$C_{\mathrm{F,T}}$	0.15 ± 0.11 (0.168)	-0.05 ± 0.09 (0.562)	0.00 ± 0.08 (0.983)
$C_{\underline{\mathbf{N}},\mathrm{T}}$	0.31 ± 0.12 (0.015)	-0.16 ± 0.09 (0.088)	-0.17 ± 0.08 (0.051)
$C_{\mathrm{F},\underline{\mathbf{N}},\mathrm{T}}$	-0.54 ± 0.15 (0.001)	-0.23 ± 0.12 (0.078)	-0.36 ± 0.11 (0.006)

*Parameters are given as mean ± standard error. *P* values given in parentheses indicate the probability that random sampling error accounts for the difference between zero and the observed value of the parameter.

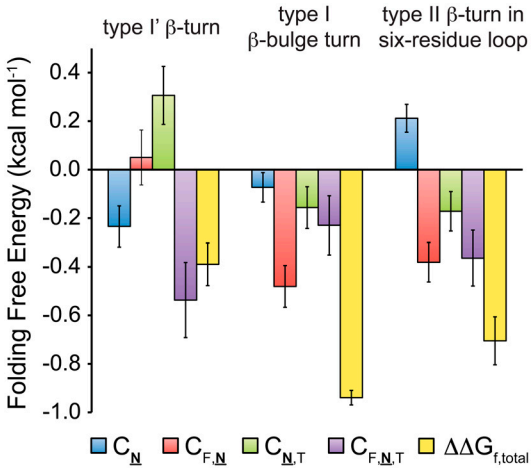


Fig. 3. Origin of the increase in stability of **4-F,T**, **5-F,T**, and **6-F,T** upon glycosylation. $\Delta\Delta G_{f,\text{total}}$ (yellow bars) is the sum of the energetic effects of (1) the Asn19 to Asn(GlcNAc)19 mutation ($C_{\underline{N}}$, blue bars); (2) the two-way interaction between Phe16 and Asn(GlcNAc)19 ($C_{F,\underline{N}}$, red bars); (3) the two-way interaction between Asn(GlcNAc)19 and Thr21 ($C_{\underline{N},T}$, green bars); and (4) the three-way interaction between Phe16, Asn(GlcNAc)19, and Thr21 ($C_{F,\underline{N},T}$, purple bars). $C_{\underline{N}}$, $C_{F,\underline{N}}$, $C_{\underline{N},T}$, and $C_{F,\underline{N},T}$, are parameters obtained from least-squares regression of Eq. **1**; error bars represent the corresponding standard errors.

though it is possible that the backbone flexibility and/or direction of the Cα–Cβ bond vectors in the five- and six-residue turns allow for better two-way interactions between Phe16 and Asn(GlcNAc)19 than are possible in the four-residue turn.

The two-way interaction between Asn(GlcNAc)19 and Thr21 stabilizes the five- and six-residue turns ($C_{\underline{N},T} = -0.16 \pm 0.09$ kcal mol^{-1} and -0.17 ± 0.08 kcal mol^{-1}, respectively), but substantially destabilizes the four-residue turn ($C_{\underline{N},T} = 0.31 \pm 0.12$ kcal mol^{-1}). Published structural data (31) indicate that the glycosylated enhanced aromatic sequon in an analogous type I β-bulge turn in HsCD2ad involves three hydrogen bonds between Thr and Asn(GlcNAc1): one between the Thr side-chain oxygen and the amide proton of the 2-acetamido group of GlcNAc, and two between the Asn side-chain amide carbonyl oxygen and the backbone amide and side-chain hydroxyl protons of Thr (Fig. 1*A*). The differences we observe between the $C_{\underline{N},T}$ values in the four-, five-, and six-residue turn contexts could reflect the presence of analogous hydrogen bonds in the type I β-bulge turn of **5g-F,T** and in six-residue loop of **6g-F,T**, but not in the type I′ β-turn of **4g-F,T**.

The $C_{F,\underline{N},T}$ values for the four-residue type I′ β-turn ($C_{F,\underline{N},T} = -0.54 \pm 0.15$ kcal mol^{-1}), the five-residue type I β-bulge turn ($C_{F,\underline{N},T} = -0.23 \pm 0.12$ kcal mol^{-1}), and the type II β-turn within a six-residue loop ($C_{F,\underline{N},T} = -0.36 \pm 0.11$ kcal mol^{-1}) mirror the $\Delta\Delta\Delta\Delta G_f$ values obtained by comparison of the front and back double mutant cycles in each triple mutant cube in Fig. 2, confirming that the three-way interaction between Phe16, Asn(GlcNAc1)19, and Thr21 stabilizes each reverse turn type by similar amounts.

Discussion

Glycosylating an enhanced aromatic sequon in its correlated reverse turn context is stabilizing, though the origins of this stabilizing effect differ amongst the enhanced aromatic sequon/reverse turn pairs (Fig. 3). In the type I′ β-turn, this effect comes predominantly from the three-way interaction between Phe16, Asn(GlcNAc1)19, and Thr21 ($C_{F,\underline{N},T}$) and from the Asn19 to Asn(GlcNAc1)19 mutation ($C_{\underline{N}}$), offset by an unfavorable two-way interaction between Asn(GlcNAc1)19 and Thr21 ($C_{\underline{N},T}$). In the type I β-bulge turn, the two-way interaction between Phe16 and Asn(GlcNAc1)19 ($C_{F,\underline{N}}$) contributes more than does the three-way interaction between Phe16, Asn(GlcNAc1)19, and Thr21 ($C_{F,\underline{N},T}$). In the type II β-turn within a six-residue loop, the two-way interaction between Phe16 and Asn(GlcNAc1)19 ($C_{F,\underline{N}}$) and the three-way interaction between Phe16, Asn(GlcNAc1)19 and Thr21 ($C_{F,\underline{N},T}$) contribute similar amounts, offset by the unfavorable effect of the Asn19 to Asn(GlcNAc1)19 mutation ($C_{\underline{N}}$). Despite these differences, our results show that each reverse turn type is a suitable host for its corresponding enhanced aromatic sequon.

Adding N-glycans to naïve sites in proteins can be an attractive strategy for increasing their stability. This approach has been used in the development of protein drugs (45–48), where new N-glycans can extend serum half-life (49–51) and shelf-life, owing in part to increased protease resistance (52), decreased aggregation propensity, and compensation for the destabilizing effect of methionine oxidation (53). Historically, efforts to increase protein stability via N-glycosylation have depended on a trial-and-error approach (51, 54), which often results in unpredictable energetic consequences (41, 55, 56). By matching each enhanced aromatic sequon to an appropriate reverse turn conformation, we have provided engineering guidelines by which N-glycosylation can reliably stabilize proteins. These matches include Phe–Asn–Xxx–Thr for type I′ β-turns, Phe–Yyy–Asn–Xxx–Thr for type I β-bulge turns, and Phe–Yyy–Zzz–Asn–Xxx–Thr for type II β-turns within a six-residue loop. Each appears to facilitate native-state stabilizing interactions between Phe, Asn(GlcNAc) and Thr in glycosylation-naïve proteins that have not evolved to optimize protein-carbohydrate interactions (30). The structure-stability relationships unveiled by this work should also enable investigators to better predict which glycans can be removed from a glycoprotein to increase crystallization propensity, without yielding an unfolded or destabilized protein.

As noted earlier, the type I β-bulge turn and the type II β-turn in a six-residue loop (in which we previously applied the Phe–Yyy–Asn–Xxx–Thr and Phe–Yyy–Zzz–Asn–Xxx–Thr sequons, respectively) comprise less than 9% of all reverse turns in the PDB (28, 33). By successfully applying the Phe–Asn–Xxx–Thr enhanced aromatic sequon to the type I′ β-turn (which comprises nearly 11% of all reverse turns in the PDB), we have doubled the number of candidate proteins in which enhanced aromatic sequons can be employed without altering the conformation or the number of residues comprising the native reverse turn (57, 58).

Materials and Methods

Protein Synthesis. WW domain variants were synthesized as C-terminal acids via solid-phase peptide synthesis, using the standard Fmoc protecting group strategy as described previously (41), and in *SI Text*. Amino acids were activated by 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (Advanced ChemTech) and N-hydroxybenzotriazole hydrate (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$_3$GlcNAc)–OH (N-α-Fmoc-N-β-[3,4,6-tri-O-acetyl-2-(acetylamino)-deoxy-2-β-glucopyranosyl]-L-asparagine). Piperidine and N,N-diisopropylethylamine were procured from Aldrich, and N,N-dimethylformamide was obtained from Fisher.

Side-chain protecting groups were globally removed and proteins were cleaved from the resin by stirring the resin for approximately 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), and proteins were precipitated from the TFA solution by addition of diethyl ether (approximately 45 mL). Acetate protecting groups were subsequently removed from the 3-, 4-, and 6-hydroxyl groups of GlcNAc in Asn(GlcNAc)-containing proteins by hydrazinolysis, as described previously (41, 59) and in *SI Text*. The WW variants were purified by reverse-phase HPLC on a C18 column using a linear gradient of water in acetonitrile with 0.2% v/v TFA. The identity of each WW variant was confirmed by matrix-assisted laser desorption/ionization time-of-flight spectrometry (MALDI-TOF; see Table S1), and purity was evaluated by analytical HPLC.

CD Measurements. CD measurements were made using an Aviv 62A DS spectropolarimeter, using quartz cuvettes with path lengths of 0.1 or 1 cm. WW domain solutions were prepared in 20 mM sodium phosphate buffer, pH 7;

protein solution concentrations were determined spectroscopically from tyrosine and tryptophan absorbance at 280 nm in 6 M guanidine hydrochloride +20 mM sodium phosphate ($\epsilon_{Trp} = 5690\ M^{-1}\ cm^{-1}$, $\epsilon_{Tyr} = 1280\ M^{-1}\ cm^{-1}$) as described previously (41, 60). CD spectra were obtained by monitoring molar ellipticity from 340 to 200 nm in 1 nm increments, with five-second averaging times. Variable temperature CD data were obtained by monitoring molar ellipticity at 227 nm from 0.2 °C to 98.2 °C or 108.2 °C at 2 °C intervals, with 90 s equilibration time between data points and 30 s averaging times. We fit the variable temperature CD data to obtain T_m and ΔG_f values for each protein, as described previously (41), and in *SI Text*. The CD data for **4**, **5**, and **6**, and their derivatives are shown in Figs. S1–S3.

ACKNOWLEDGMENTS. We thank R. A. Dwek for helpful prepublication discussions. This work was supported in part by National Institutes of Health Grant GM051105 (J.W.K. and E.T.P), the Skaggs Institute for Chemical Biology, and the Lita Annenberg Hazen Foundation. J.L.P was supported in part by a National Research Service Award National Institutes of Health Postdoctoral Fellowship F32 GM086039. The Scripps Research Institute has filed a provisional patent application covering the results described in this paper.

1. Imperiali B (1997) Protein glycosylation: The clash of the titans. *Acc Chem Res* 30:452–459.
2. Yan A, Lennarz WJ (2005) Unraveling the mechanism of protein N-glycosylation. *J Biol Chem* 280:3121–3124.
3. Kornfeld R, Kornfeld S (1985) Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* 54:631–664.
4. Kelleher DJ, Gilmore R (2006) An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* 16:47R–62R.
5. Helenius A, Aebi M (2001) Intracellular functions of N-linked glycans. *Science* 291:2364–2369.
6. Molinari M (2007) N-glycan structure dictates extension of protein folding or onset of disposal. *Nat Chem Biol* 3:313–320.
7. Yamaguchi H (2002) Chaperone-like functions of N-glycans in the formation and stabilization of protein conformation. *Trends Glycosci Glyc* 14:139–151.
8. Hanson SR, et al. (2009) The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc Natl Acad Sci USA* 106:3131–3136.
9. Wormald MR, et al. (1991) The conformational effects of N-glycosylation on the tailpiece from serum IgM. *Eur J Biochem* 198:131–139.
10. Laszlo O, Jr, et al. (1991) Glycosylation of synthetic peptides breaks helices. *Int J Pept Protein Res* 38:476–482.
11. Laczko I, et al. (1992) Synthesis and conformational studies of N-glycosylated analogs of the HIV-1 principal neutralizing determinant. *Biochemistry* 31:4282–4288.
12. Urge L, Gorbics L, Otvos L, Jr (1992) Chemical glycosylation of peptide T at natural and artificial glycosylation sites stabilizes or rearranges the dominant reverse turn structure. *Biochem Biophys Res Commun* 184:1125–1132.
13. Andreotti AH, Kahne D (1993) Effects of glycosylation on peptide backbone conformation. *J Am Chem Soc* 115:3352–3353.
14. Liang R, Andreotti AH, Kahne D (1995) Sensitivity of glycopeptide conformation to carbohydrate chain length. *J Am Chem Soc* 117:10395–10396.
15. Rickert KW, Imperiali B (1995) Analysis of the conserved glycosylation site in the nicotinic acetylcholine receptor: Potential roles in complex assembly. *Chem Biol* 2:751–759.
16. Imperiali B, Rickert KW (1995) Conformational implications of asparagine-linked glycosylation. *Proc Natl Acad Sci USA* 92:97–101.
17. Live DH, Kumar RA, Beebe X, Danishefsky SJ (1996) Conformational influences of glycosylation of a peptide: A possible model for the effect of glycosylation on the rate of protein folding. *Proc Natl Acad Sci USA* 93:12759–12761.
18. O'Connor SE, Imperiali B (1996) Modulation of protein structure and function by asparagine-linked glycosylation. *Chem Biol* 3:803–812.
19. O'Connor SE, Imperiali B (1997) Conformational switching by asparagine-linked glycosylation. *J Am Chem Soc* 119:2295–2296.
20. O'Connor SE, Imperiali B (1998) A molecular basis for glycosylation-induced conformational switching. *Chem Biol* 5:427–437.
21. Imperiali B, O'Connor SE (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol* 3:643–649.
22. O'Connor SE, Pohlmann J, Imperiali B, Saskiawan I, Yamamoto K (2001) Probing the effect of the outer saccharide residues of N-linked glycans on peptide conformation. *J Am Chem Soc* 123:6187–6188.
23. Wormald MR, et al. (2002) Conformational studies of oligosaccharides and glycopeptides: Complementarity of NMR, X-ray crystallography, and molecular modelling. *Chem Rev* 102:371–386.
24. Bosques CJ, Tschampel SM, Woods RJ, Imperiali B (2004) Effects of glycosylation on peptide conformation: A synergistic experimental and computational study. *J Am Chem Soc* 126:8421–8425.
25. Wormald MR, Dwek RA (1999) Glycoproteins: Glycan presentation and protein-fold stability. *Struct Fold Des* 7:R155–R160.
26. Shental-Bechor D, Levy Y (2008) Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proc Natl Acad Sci USA* 105:8256–8261.
27. Barb AW, Borgert AJ, Liu MA, Barany G, Live D (2010) Intramolecular glycan–protein interactions in glycoproteins. *Methods Enzymol* 478:365–388.
28. Sibanda BL, Blundell TL, Thornton JM (1989) Conformation of beta-hairpins in protein structures—a systematic classification with applications to modeling by homology, electron-density fitting and protein engineering. *J Mol Biol* 206:759–777.
29. Richardson JS (1981) The anatomy and taxonomy of protein structure. *Adv Protein Chem* 34:167–339.
30. Culyba EK, et al. (2011) Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns. *Science* 331:571–575.
31. Wyss DF, et al. (1995) Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science* 269:1273–1278.
32. Laughrey ZR, Kiehna SE, Riemen AJ, Waters ML (2008) Carbohydrate-pi interactions: What are they worth? *J Am Chem Soc* 130:14625–14633.
33. Oliva B, Bates PA, Querol E, Aviles FX, Sternberg MJE (1997) An automated classification of the structure of protein loops. *J Mol Biol* 266:814–830.
34. Ranganathan R, Lu KP, Hunter T, Noel JP (1997) Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* 89:875–886.
35. Nickson AA, Clarke J (2010) What lessons can be learned from studying the folding of homologous proteins? *Methods* 52:38–50.
36. Jäger M, et al. (2006) Structure-function-folding relationship in a WW domain. *Proc Natl Acad Sci USA* 103:10648–10653.
37. Fuller AA, et al. (2009) Evaluating β-turn mimics as β-sheet folding nucleators. *Proc Natl Acad Sci USA* 106:11067–11072.
38. Jäger M, Nguyen H, Crane JC, Kelly JW, Gruebele M (2001) The folding mechanism of a β-sheet: The WW domain. *J Mol Biol* 311:373–393.
39. Kaul R, Angeles AR, Jäger M, Powers ET, Kelly JW (2001) Incorporating β-turns and a turn mimetic out of context in loop 1 of the WW domain affords cooperatively folded β-sheets. *J Am Chem Soc* 123:5206–5212.
40. Bertozzi CR, Kiessling LL (2001) Chemical glycobiology. *Science* 291:2357–2364.
41. Price JL, et al. (2010) Context-dependent effects of asparagine glycosylation on Pin WW folding kinetics and thermodynamics. *J Am Chem Soc* 132:15359–15367.
42. Horovitz A, Fersht AR (1992) Cooperative interactions during protein folding. *J Mol Biol* 224:733–740.
43. Hovmoller S, Zhou T, Ohlson T (2002) Conformations of amino acids in proteins. *Acta Crystallogr D* 58:768–776.
44. Hutchinson EG, Thornton JM (1994) A revised set of potentials for beta-turn formation in proteins. *Protein Sci* 3:2207–2216.
45. Walsh G, Jefferis R (2006) Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* 24:1241–1252.
46. Sinclair AM, Elliott S (2005) Glycoengineering: The effect of glycosylation on the properties of therapeutic proteins. *J Pharm Sci-US* 94:1626–1635.
47. Li HJ, d'Anjou M (2009) Pharmacological significance of glycosylation in therapeutic proteins. *Curr Opin Biotech* 20:678–684.
48. Sola RJ, Griebenow K (2010) Glycosylation of therapeutic proteins an effective strategy to optimize efficacy. *BioDrugs* 24:9–21.
49. Egrie JC, Dwyer E, Browne JK, Hitz A, Lykos MA (2003) Darbepoetin alfa has a longer circulating half-life and greater in vivo potency than recombinant human erythropoietin. *Exp Hematol* 31:290–299.
50. Su DM, Zhao HL, Xia HZ (2010) Glycosylation-modified erythropoietin with improved half-life and biological activity. *Int J Hematol* 91:238–244.
51. Ceaglio N, Etcheverrigaray M, Kratje R, Oggero A (2008) Novel long-lasting interferon alpha derivatives designed by glycoengineering. *Biochimie* 90:437–449.
52. Raju TS, Scallon BJ (2006) Glycosylation in the Fc domain of IgG increases resistance to proteolytic cleavage by papain. *Biochem Biophys Res Commun* 341:797–803.
53. Liu D, et al. (2008) Structure and stability changes of human IgG1 Fc as a consequence of methionine oxidation. *Biochemistry* 47:5088–5100.
54. Elliott S, Chang D, Delorme E, Eris T, Lorenzini T (2004) Structural requirements for additional *N*-linked carbohydrate on recombinant human erythropoietin. *J Biol Chem* 279:16854–16862.
55. Hackenberger CPR, Friel CT, Radford SE, Imperiali B (2005) Semisynthesis of a glycosylated Im7 analogue for protein folding studies. *J Am Chem Soc* 127:12882–12889.
56. Chen MM, et al. (2010) Perturbing the folding energy landscape of the bacterial immunity protein Im7 by site-specific N-linked glycosylation. *Proc Natl Acad Sci USA* 107:22528–22533.
57. DeGrado WF, Summa CM, Pavone V, Nastri F, Lombardi A (1999) De novo design and structural characterization of proteins and metalloproteins. *Annu Rev Biochem* 68:779–819.
58. Gellman SH (1998) Minimal model systems for beta sheet secondary structure in proteins. *Curr Opin Chem Biol* 2:717–725.
59. Ficht S, Payne RJ, Guy RT, Wong C-H (2008) Solid-phase synthesis of peptide and glycopeptide thioesters through side-chain-anchoring strategies. *Chem-Eur J* 14:3620–3629.
60. Edelhoch H (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* 6:1948–1954.

Supporting Information

Price et al. 10.1073/pnas.1105880108

SI Materials and Methods

Synthesis of Pin WW Variants. Proteins **6**, **6-F**, **6-T**, and **6-F,T**, and glycoproteins **6g**, **6g-F**, **6g-T**, and **6g-F,T** were synthesized previously (1). Proteins **4**, **4-F**, **4-T**, **4-F,T**, **5**, **5-F**, **5-T**, and **5-F,T** and glycoproteins **4g**, **4g-F**, **4g-T**, **4g-F,T**, **5g**, **5g-F**, **5g-T**, and **5g-F,T** were synthesized as C-terminal acids, employing a solid phase peptide synthesis approach using a standard Fmoc Nα protecting group strategy either manually or via a combination of manual and automated peptide synthesis (Applied Biosystems 433A automated peptide synthesizer). The complete amino acid sequences of **4** and **5** and their derivatives are given in Table S1. 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) (2, 3). Piperidine and N,N-diisopropylethylamine (DIEA) were obtained from Aldrich, N,N,dimethylformamide was obtained from Fisher, and N-methyl pyrrolidinone (NMP) was procured 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 min. 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 mins. 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 μL DIEA (500 μmol, 10 eq). (Only 1.5 eq of amino acid were used during the coupling of the expensive Fmoc–Asn(Ac_3GlcNAc)–OH monomer, and the 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 (including Fmoc–Asn(Ac_3GlcNAc)–OH) 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 approximately 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, and the resin was rinsed with additional TFA. Proteins were precipitated from the concentrated TFA solution by addition of diethyl ether (approximately 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 Domain Proteins. Acetate protecting groups were removed from the 3-, 4-, and 6-hydroxyl groups on the Asn-linked GlcNAc residues in glycoproteins **4g**, **4g-F**, **4g-T**, **4g-F,T**, **5g**, **5g-F**, **5g-T**, **5g-F,T**, **6g**, **6g-F**, **6g-T**, and **6g-F,T** via hydrazinolysis as described previously (4). Briefly, the crude protein was dissolved in a solution of 5% hydrazine solution in 60 mM aqueous dithiothreitol (containing as much as 50% acetonitrile, to facilitate dissolution of the crude protein) and allowed to stand at room temperature for approximately 1 h with intermittent agitation. The deprotection reaction was quenched by the addition of approximately 1 mL TFA and approximately 20 mL water. The quenched reaction mixture was frozen and lyophilized to give the crude deprotected protein (often an oily white precipitate).

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 GdnHCl (8 M GdnHCl was frequently required to dissolve the crude glycosylated proteins even though these proteins were readily soluble in water after purification; DMSO is preferable, as injecting 8 M GdnHCl onto C18 columns appears to significantly shorten their useful lifetime). Proteins were purified by 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), and purity was established by analytical HPLC.

Circular Dichroism Spectropolarimetry. Measurements were made with an Aviv 62A DS Circular Dichroism Spectropolarimeter, using quartz cuvettes with a path length of either 0.1 or 1 cm. Protein solutions were prepared in 20 mM sodium phosphate buffer, pH 7, and protein concentrations were determined spectroscopically based on tyrosine and tryptophan absorbance at 280 nm in 6 M GdnHCl + 20 mM sodium phosphate (ϵ_{Trp} = 5690 $M^{-1}\,cm^{-1}$, ϵ_{Tyr} = 1280 $M^{-1}\,cm^{-1}$) (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 or 108.2 °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:

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

where T is temperature in Kelvin, D_0 is the y-intercept and D_1 is the slope of the posttransition baseline; N_0 is the y-intercept and N_1 is the slope of the pretransition 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:

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

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:

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

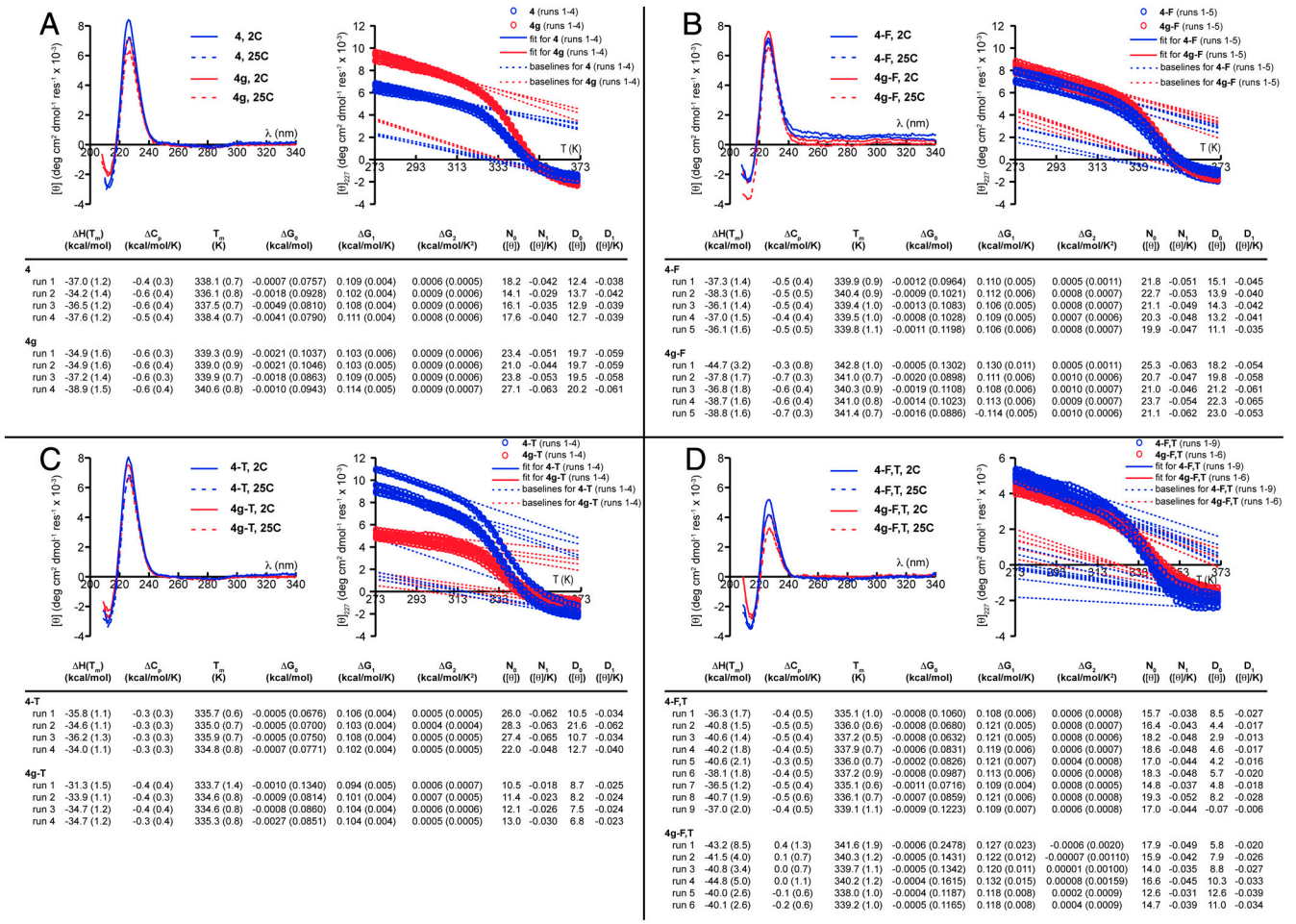
where $\Delta H(T_m)$ is the enthalpy of folding at the melting temperature and ΔC_p is the heat capacity of folding ($\Delta H(T_m)$, ΔC_p, and T_m are parameters of the fit). The second equation represents $\Delta G_f(T)$ as a Taylor series expansion about the melting temperature:

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

where 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 Eq. **S3**). In this series expansion, ΔG_0 is the free energy of folding at T_m (and is therefore very close to 0), ΔG_1 is equal to $d\Delta G/dT$, or $-\Delta S_f(T_m)$ (the entropy at T_m) and ΔG_2 is $(1/2) \times (d^2\Delta G/2dT^2) = -(1/2) \times d\Delta S_f/dT = -(1/2) \times \Delta C_p/T$. Fits to the two equations give overlapping information, and indeed the data in Figs. S1–S3 show that the parameter estimates obtained from the fits to the two equations are consistent with each other. We prefer to use the results from fitting to Eq. **S4** because fits to quadratic functions are generally more robust than fits to functions with logarithms, and to be consistent with our previous work (1). The ΔG_f values displayed in Table 1 of the main text for each Pin WW protein were obtained by averaging the ΔG_f values (calculated at 338.15 K using Eq. 3) from each of three or more replicate variable-temperature CD experiments on the same protein.

CD spectra and variable-temperature CD data for Pin WW proteins **4**, **4-F**, **4-T**, **4-F,T**, **4g**, **4g-F**, **4g-T**, and **4g-F,T** appear in Fig. S1. CD spectra and variable-temperature CD data for proteins **5**, **5-F**, **5-T**, **5-F,T**, **5g**, **5g-F**, **5g-T**, and **5g-F,T** appear in Fig. S2. CD spectra [reported previously (1), but included here for comparison with CD spectra from **4** and **5** and their derivatives] and variable-temperature CD data for proteins **6**, **6-F**, **6-T**, **6-F,T** **6g**, **6g-F**, **6g-T** and **6g-F,T** appear in Fig. S3. Figs. S1–S3 show values for the parameters from Eqs. **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 ΔG_f values (in Table 1 of the main text) by propagation of error.

1. Culyba EK, et al. (2011) Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns. *Science* 331:571–575.
2. Meldal M, Bock K (1990) Pentafluorophenyl esters for temporary carboxyl group protection in solid phase synthesis of N-linked glycopeptides. *Tetrahedron Lett* 31:6987–6990.
3. Otvos L Jr, et al. (1990) Automated solid-phase synthesis of glycopeptides. Incorporation of unprotected mono- and disaccharide units of N-glycoprotein antennae into T cell epitopic peptides. *Tetrahedron Lett* 31:5889–5892.
4. Ficht S, Payne RJ, Guy RT, Wong C-H (2008) Solid-phase synthesis of peptide and glycopeptide thioesters through side-chain-anchoring strategies. *Chem-Eur J* 14:3620–3629.
5. Edelhoch H (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* 6:1948–1954.



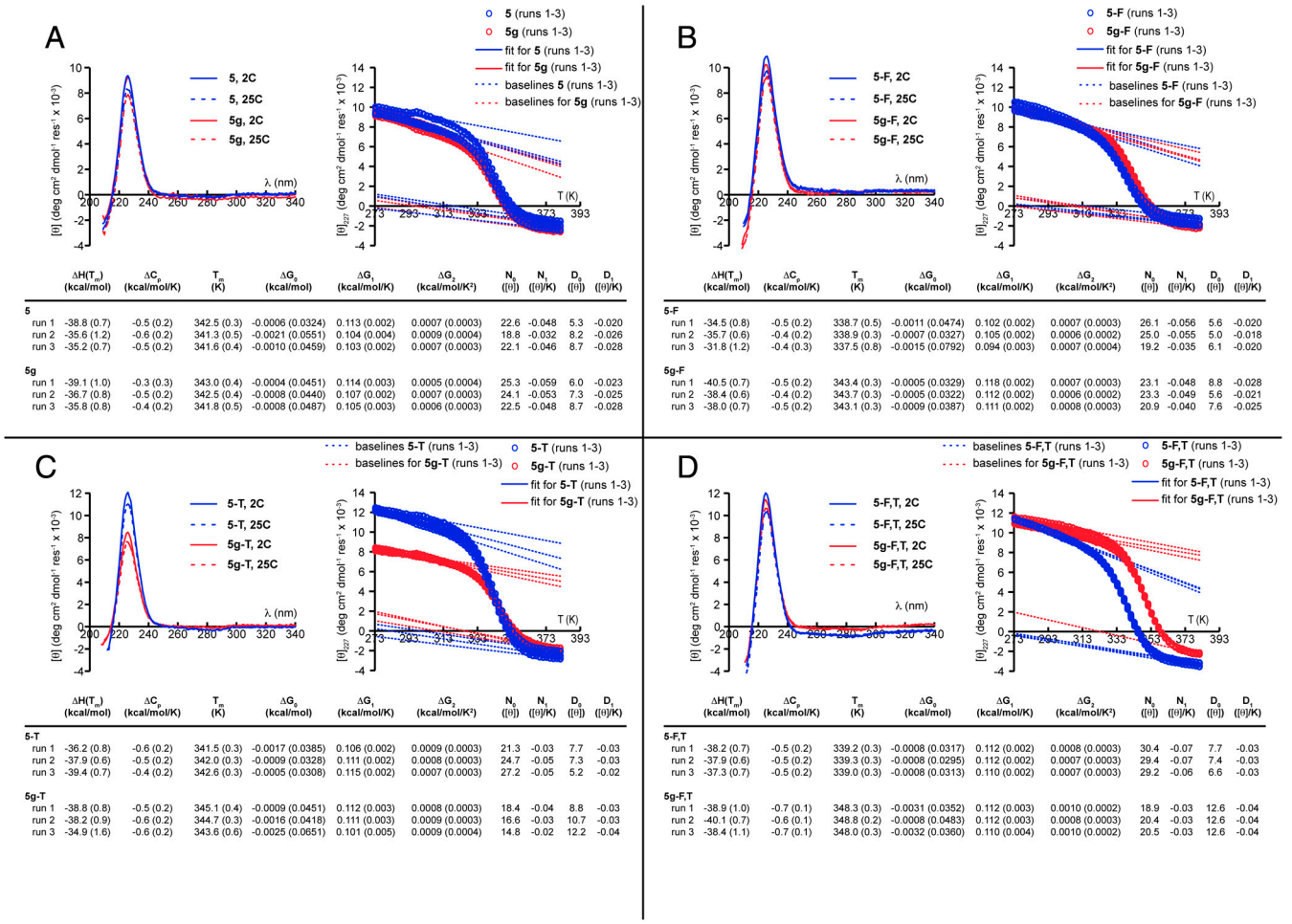
	$\Delta H(T_m)$ (kcal/mol)	ΔC_p (kcal/mol/K)	T_m (K)	ΔG_0 (kcal/mol)	ΔG_1 (kcal/mol/K)	ΔG_2 (kcal/mol/K²)	N_0 ([θ])	N_1 ([θ]/K)	D_0 ([θ])	D_1 ([θ]/K)
4										
run 1	-37.0 (1.2)	-0.4 (0.3)	338.1 (0.7)	-0.0007 (0.0757)	0.109 (0.004)	0.0006 (0.0005)	18.2	-0.042	12.4	-0.038
run 2	-34.2 (1.4)	-0.6 (0.4)	336.1 (0.8)	-0.0018 (0.0928)	0.102 (0.004)	0.0009 (0.0006)	14.1	-0.029	13.7	-0.042
run 3	-36.5 (1.2)	-0.6 (0.4)	337.5 (0.7)	-0.0049 (0.0810)	0.108 (0.004)	0.0009 (0.0006)	16.1	-0.035	12.9	-0.039
run 4	-37.6 (1.2)	-0.5 (0.4)	338.4 (0.7)	-0.0041 (0.0790)	0.111 (0.004)	0.0008 (0.0006)	17.6	-0.040	12.7	-0.039
4g										
run 1	-34.9 (1.6)	-0.6 (0.3)	339.3 (0.9)	-0.0021 (0.1037)	0.103 (0.006)	0.0009 (0.0006)	23.4	-0.051	19.7	-0.059
run 2	-34.9 (1.6)	-0.6 (0.4)	339.0 (0.9)	-0.0021 (0.1046)	0.103 (0.005)	0.0009 (0.0006)	21.0	-0.044	19.7	-0.059
run 3	-37.2 (1.4)	-0.6 (0.3)	339.9 (0.7)	-0.0018 (0.0863)	0.109 (0.005)	0.0009 (0.0006)	23.8	-0.053	19.5	-0.058
run 4	-38.9 (1.5)	-0.6 (0.4)	340.6 (0.8)	-0.0010 (0.0943)	0.114 (0.005)	0.0009 (0.0007)	27.1	-0.063	20.2	-0.061

	$\Delta H(T_m)$ (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)
4-F										
run 1	-37.3 (1.4)	-0.5 (0.4)	339.9 (0.9)	-0.0012 (0.0964)	0.110 (0.005)	0.0005 (0.0011)	21.8	-0.051	15.1	-0.045
run 2	-38.3 (1.6)	-0.5 (0.5)	340.4 (0.9)	-0.0009 (0.1021)	0.112 (0.006)	0.0008 (0.0007)	22.7	-0.053	13.9	-0.040
run 3	-36.1 (1.4)	-0.5 (0.4)	339.4 (1.0)	-0.0013 (0.1083)	0.106 (0.005)	0.0008 (0.0007)	21.1	-0.049	14.3	-0.042
run 4	-37.0 (1.5)	-0.4 (0.4)	339.5 (1.0)	-0.0008 (0.1028)	0.109 (0.005)	0.0007 (0.0006)	20.3	-0.048	13.2	-0.041
run 5	-36.1 (1.6)	-0.5 (0.5)	339.8 (1.1)	-0.0011 (0.1198)	0.106 (0.006)	0.0008 (0.0007)	19.9	-0.047	11.1	-0.035
4g-F										
run 1	-44.7 (3.2)	-0.3 (0.8)	342.8 (1.0)	-0.0005 (0.1302)	0.130 (0.011)	0.0005 (0.0011)	25.3	-0.063	18.2	-0.054
run 2	-37.8 (1.7)	-0.7 (0.3)	341.0 (0.7)	-0.0020 (0.0898)	0.111 (0.006)	0.0010 (0.0006)	20.7	-0.047	19.8	-0.058
run 3	-36.8 (1.8)	-0.6 (0.4)	340.3 (0.9)	-0.0019 (0.1108)	0.108 (0.006)	0.0010 (0.0007)	21.0	-0.046	21.2	-0.061
run 4	-38.7 (1.6)	-0.6 (0.4)	341.0 (0.8)	-0.0014 (0.1023)	0.113 (0.006)	0.0009 (0.0007)	23.7	-0.054	22.3	-0.065
run 5	-38.8 (1.6)	-0.7 (0.3)	341.4 (0.7)	-0.0016 (0.0886)	-0.114 (0.005)	0.0010 (0.0006)	21.1	-0.062	23.0	-0.053

	$\Delta H(T_m)$ (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)
4-T										
run 1	-35.8 (1.1)	-0.3 (0.3)	335.7 (0.6)	-0.0005 (0.0676)	0.106 (0.004)	0.0005 (0.0005)	26.0	-0.062	10.5	-0.034
run 2	-34.6 (1.1)	-0.3 (0.3)	335.0 (0.7)	-0.0005 (0.0700)	0.103 (0.004)	0.0004 (0.0004)	28.3	-0.063	21.6	-0.062
run 3	-36.2 (1.3)	-0.3 (0.3)	335.9 (0.7)	-0.0005 (0.0750)	0.108 (0.004)	0.0005 (0.0005)	27.4	-0.065	10.7	-0.034
run 4	-34.0 (1.1)	-0.3 (0.3)	334.8 (0.8)	-0.0007 (0.0771)	0.102 (0.004)	0.0005 (0.0005)	22.0	-0.048	12.7	-0.040
4g-T										
run 1	-31.3 (1.5)	-0.4 (0.4)	333.7 (1.4)	-0.0010 (0.1340)	0.094 (0.005)	0.0006 (0.0007)	10.5	-0.018	8.7	-0.025
run 2	-33.9 (1.1)	-0.4 (0.3)	334.6 (0.8)	-0.0009 (0.0814)	0.101 (0.004)	0.0007 (0.0005)	11.4	-0.023	8.2	-0.024
run 3	-34.7 (1.2)	-0.4 (0.4)	334.6 (0.8)	-0.0008 (0.0860)	0.104 (0.004)	0.0006 (0.0006)	12.1	-0.026	7.5	-0.024
run 4	-34.7 (1.2)	-0.3 (0.4)	335.3 (0.8)	-0.0027 (0.0851)	0.104 (0.004)	0.0005 (0.0005)	13.0	-0.030	6.8	-0.023

	$\Delta H(T_m)$ (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)
4-F,T										
run 1	-36.3 (1.7)	-0.4 (0.5)	335.1 (1.0)	-0.0008 (0.1060)	0.108 (0.006)	0.0006 (0.0008)	15.7	-0.038	8.5	-0.027
run 2	-40.8 (1.5)	-0.5 (0.5)	336.0 (0.6)	-0.0008 (0.0680)	0.121 (0.005)	0.0008 (0.0007)	16.4	-0.043	4.4	-0.017
run 3	-40.6 (1.4)	-0.5 (0.4)	337.2 (0.5)	-0.0008 (0.0632)	0.121 (0.005)	0.0008 (0.0006)	18.2	-0.048	2.9	-0.013
run 4	-40.2 (1.8)	-0.4 (0.5)	337.9 (0.7)	-0.0006 (0.0831)	0.119 (0.006)	0.0006 (0.0007)	18.6	-0.048	4.6	-0.017
run 5	-40.6 (2.1)	-0.3 (0.5)	336.0 (0.7)	-0.0002 (0.0826)	0.121 (0.007)	0.0004 (0.0008)	17.0	-0.044	4.2	-0.016
run 6	-38.1 (1.8)	-0.4 (0.5)	337.2 (0.9)	-0.0008 (0.0987)	0.113 (0.006)	0.0006 (0.0008)	18.3	-0.048	5.7	-0.020
run 7	-36.5 (1.2)	-0.5 (0.4)	335.1 (0.6)	-0.0011 (0.0716)	0.109 (0.004)	0.0008 (0.0005)	14.8	-0.037	4.8	-0.018
run 8	-40.7 (1.9)	-0.5 (0.6)	336.1 (0.7)	-0.0007 (0.0859)	0.121 (0.006)	0.0008 (0.0008)	19.3	-0.052	8.2	-0.028
run 9	-37.0 (2.0)	-0.4 (0.5)	339.1 (1.1)	-0.0009 (0.1223)	0.109 (0.007)	0.0006 (0.0008)	17.0	-0.044	-0.07	-0.006
4g-F,T										
run 1	-43.2 (8.5)	0.4 (1.3)	341.6 (1.9)	-0.0006 (0.2478)	0.127 (0.023)	-0.0006 (0.0020)	17.9	-0.049	5.8	-0.020
run 2	-41.5 (4.0)	0.1 (0.7)	340.3 (1.2)	-0.0005 (0.1431)	0.122 (0.012)	-0.00007 (0.00110)	15.9	-0.042	7.9	-0.026
run 3	-40.8 (3.4)	0.0 (0.7)	339.7 (1.1)	-0.0005 (0.1342)	0.120 (0.011)	0.00001 (0.00100)	14.0	-0.035	8.8	-0.027
run 4	-44.8 (5.0)	0.0 (1.1)	340.2 (1.2)	-0.0004 (0.1615)	0.132 (0.015)	0.00008 (0.00159)	16.6	-0.045	10.3	-0.033
run 5	-40.0 (2.6)	-0.1 (0.6)	338.0 (1.0)	-0.0004 (0.1187)	0.118 (0.008)	0.0002 (0.0009)	12.6	-0.031	12.6	-0.039
run 6	-40.1 (2.6)	-0.2 (0.6)	339.2 (1.0)	-0.0005 (0.1165)	0.118 (0.008)	0.0004 (0.0009)	14.7	-0.039	11.0	-0.034

Fig. S1. CD spectra, variable-temperature CD data, and fit parameters (see Eqs. **S1**–**S4**) for (*A*) protein **4** and glycoprotein **4g**; (*B*) protein **4-F** and glycoprotein **4g-F**; (*C*) protein **4-T** and glycoprotein **4g-T**; and (*D*) protein **4-F,T**, and glycoprotein **4g-F,T** in 20 mM sodium phosphate buffer, pH 7. Variable-temperature CD experiments were conducted at a protein concentration of 10 μM. CD spectra for **4**, **4g**, **4-T**, and **4g-T** were obtained at 100 μM protein concentration. CD spectra for **4-F**, **4g-F**, **4-F,T**, and **4g-F,T** were obtained at 10 μM protein concentration.



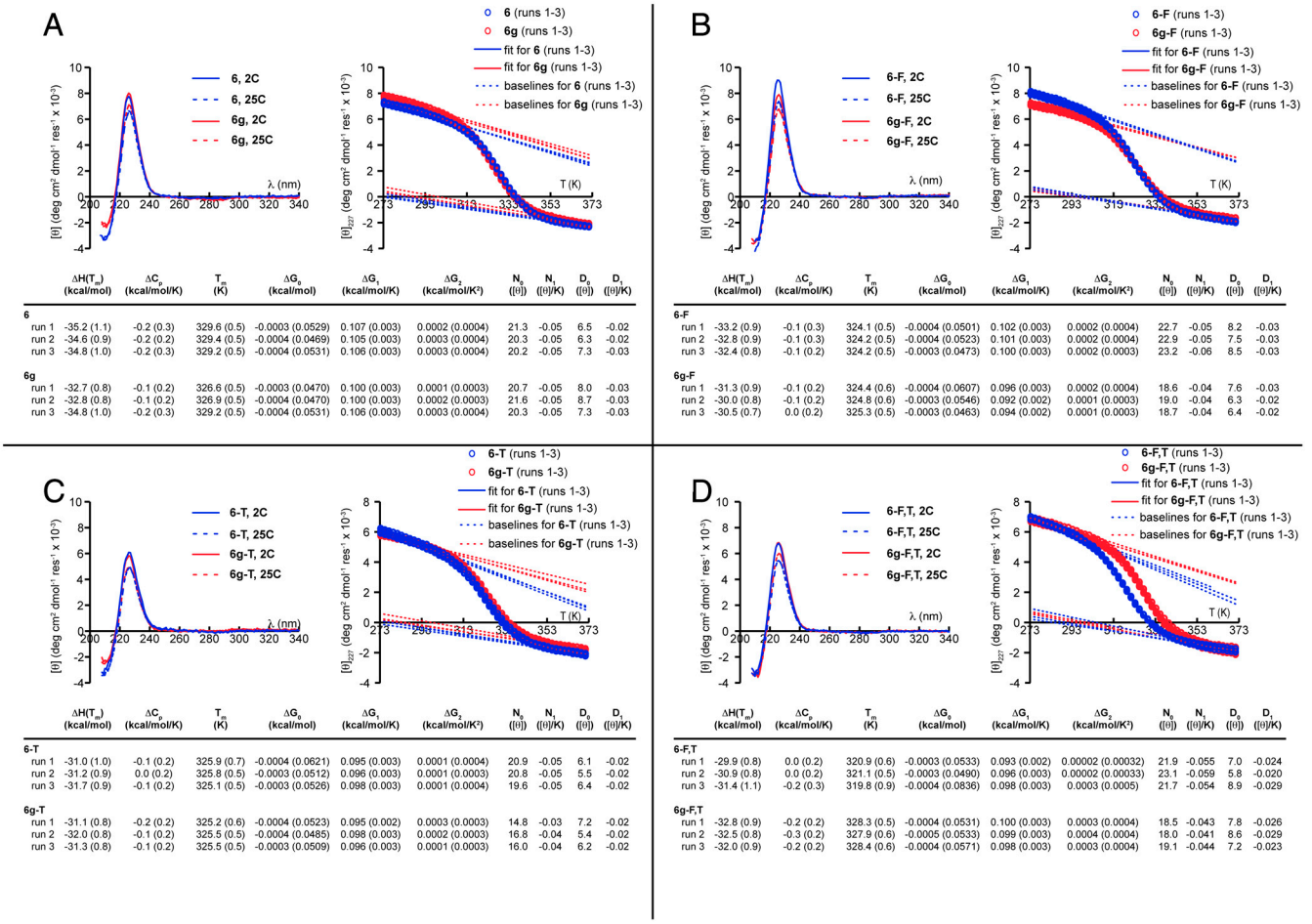
	$\Delta H(T_m)$ (kcal/mol)	ΔC_p (kcal/mol/K)	T_m (K)	ΔG_s (kcal/mol)	ΔG_1 (kcal/mol/K)	ΔG_2 (kcal/mol/K²)	N_0 ([θ])	N_1 ([θ]/K)	D_0 ([θ])	D_1 ([θ]/K)
5										
run 1	-38.8 (0.7)	-0.5 (0.2)	342.5 (0.3)	-0.0006 (0.0324)	0.113 (0.002)	0.0007 (0.0003)	22.6	-0.048	5.3	-0.020
run 2	-35.6 (1.2)	-0.6 (0.2)	341.3 (0.5)	-0.0021 (0.0551)	0.104 (0.004)	0.0009 (0.0004)	18.8	-0.032	8.2	-0.026
run 3	-35.2 (0.7)	-0.5 (0.2)	341.6 (0.4)	-0.0010 (0.0459)	0.103 (0.002)	0.0007 (0.0003)	22.1	-0.046	8.7	-0.028
5g										
run 1	-39.1 (1.0)	-0.3 (0.3)	343.0 (0.4)	-0.0004 (0.0451)	0.114 (0.003)	0.0005 (0.0004)	25.3	-0.059	6.0	-0.023
run 2	-36.7 (0.8)	-0.5 (0.2)	342.5 (0.4)	-0.0008 (0.0440)	0.107 (0.002)	0.0007 (0.0003)	24.1	-0.053	7.3	-0.025
run 3	-35.8 (0.8)	-0.4 (0.2)	341.8 (0.5)	-0.0008 (0.0487)	0.105 (0.003)	0.0006 (0.0003)	22.5	-0.048	8.7	-0.028

	$\Delta H(T_m)$ (kcal/mol)	ΔC_p (kcal/mol/K)	T_m (K)	ΔG_s (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)
5-F										
run 1	-34.5 (0.8)	-0.5 (0.2)	338.7 (0.5)	-0.0011 (0.0474)	0.102 (0.002)	0.0007 (0.0003)	26.1	-0.056	5.6	-0.020
run 2	-35.7 (0.6)	-0.4 (0.2)	338.9 (0.3)	-0.0007 (0.0327)	0.105 (0.002)	0.0006 (0.0002)	25.0	-0.055	5.0	-0.018
run 3	-31.8 (1.2)	-0.4 (0.3)	337.5 (0.8)	-0.0015 (0.0792)	0.094 (0.003)	0.0007 (0.0004)	19.2	-0.035	6.1	-0.020
5g-F										
run 1	-40.5 (0.7)	-0.5 (0.2)	343.4 (0.3)	-0.0005 (0.0329)	0.118 (0.002)	0.0007 (0.0003)	23.1	-0.048	8.8	-0.028
run 2	-38.4 (0.6)	-0.4 (0.2)	343.7 (0.3)	-0.0005 (0.0322)	0.112 (0.002)	0.0006 (0.0002)	23.3	-0.049	5.6	-0.021
run 3	-38.0 (0.7)	-0.5 (0.2)	343.1 (0.3)	-0.0009 (0.0387)	0.111 (0.002)	0.0008 (0.0003)	20.9	-0.040	7.6	-0.025

	$\Delta H(T_m)$ (kcal/mol)	ΔC_p (kcal/mol/K)	T_m (K)	ΔG_s (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)
5-T										
run 1	-36.2 (0.8)	-0.6 (0.2)	341.5 (0.3)	-0.0017 (0.0385)	0.106 (0.002)	0.0009 (0.0003)	21.3	-0.03	7.7	-0.03
run 2	-37.9 (0.6)	-0.5 (0.2)	342.0 (0.3)	-0.0009 (0.0328)	0.111 (0.002)	0.0008 (0.0003)	24.7	-0.05	7.3	-0.03
run 3	-39.4 (0.7)	-0.4 (0.2)	342.6 (0.3)	-0.0005 (0.0308)	0.115 (0.002)	0.0007 (0.0003)	27.2	-0.05	5.2	-0.02
5g-T										
run 1	-38.8 (0.8)	-0.5 (0.2)	345.1 (0.4)	-0.0009 (0.0451)	0.112 (0.003)	0.0008 (0.0003)	18.4	-0.04	8.8	-0.03
run 2	-38.2 (0.9)	-0.6 (0.2)	344.7 (0.3)	-0.0016 (0.0418)	0.111 (0.003)	0.0009 (0.0003)	16.6	-0.03	10.7	-0.03
run 3	-34.9 (1.6)	-0.6 (0.2)	343.6 (0.6)	-0.0025 (0.0651)	0.101 (0.005)	0.0009 (0.0004)	14.8	-0.02	12.2	-0.04

	$\Delta H(T_m)$ (kcal/mol)	ΔC_p (kcal/mol/K)	T_m (K)	ΔG_s (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)
5-F,T										
run 1	-38.2 (0.7)	-0.5 (0.2)	339.2 (0.3)	-0.0008 (0.0317)	0.112 (0.002)	0.0008 (0.0003)	30.4	-0.07	7.7	-0.03
run 2	-37.9 (0.6)	-0.5 (0.2)	339.3 (0.3)	-0.0008 (0.0295)	0.112 (0.002)	0.0007 (0.0003)	29.4	-0.07	7.4	-0.03
run 3	-37.3 (0.7)	-0.5 (0.2)	339.0 (0.3)	-0.0008 (0.0313)	0.110 (0.002)	0.0007 (0.0003)	29.2	-0.06	6.6	-0.03
5g-F,T										
run 1	-38.9 (1.0)	-0.7 (0.1)	348.3 (0.3)	-0.0031 (0.0352)	0.112 (0.003)	0.0010 (0.0002)	18.9	-0.03	12.6	-0.04
run 2	-40.1 (0.7)	-0.6 (0.1)	348.8 (0.2)	-0.0008 (0.0483)	0.112 (0.003)	0.0008 (0.0003)	20.4	-0.03	12.6	-0.04
run 3	-38.4 (1.1)	-0.7 (0.1)	348.0 (0.3)	-0.0032 (0.0360)	0.110 (0.004)	0.0010 (0.0002)	20.5	-0.03	12.6	-0.04

Fig. S2. CD spectra, variable-temperature CD data, and fit parameters (see Eqs. **S1**–**S4**) for (*A*) protein **5** and glycoprotein **5g**; (*B*) protein **5-F** and glycoprotein **5g-F**; (*C*) protein **5-T** and glycoprotein **5g-T**; and (*D*) protein **5-F,T**, and glycoprotein **5g-F,T** in 20 mM sodium phosphate buffer, pH 7. CD experiments were conducted at a protein concentration of 10 μM.



	ΔH(Tm) (kcal/mol)	ΔCp (kcal/mol/K)	Tm (K)	ΔG0 (kcal/mol)	ΔG1 (kcal/mol/K)	ΔG2 (kcal/mol/K²)	N0 ([θ])	N1 ([θ]/K)	D0 ([θ])	D1 ([θ]/K)
6										
run 1	-35.2 (1.1)	-0.2 (0.3)	329.6 (0.5)	-0.0003 (0.0529)	0.107 (0.003)	0.0002 (0.0004)	21.3	-0.05	6.5	-0.02
run 2	-34.6 (0.9)	-0.2 (0.2)	329.4 (0.5)	-0.0004 (0.0469)	0.105 (0.003)	0.0003 (0.0004)	20.3	-0.05	6.3	-0.02
run 3	-34.8 (1.0)	-0.2 (0.3)	329.2 (0.5)	-0.0004 (0.0531)	0.106 (0.003)	0.0003 (0.0004)	20.2	-0.05	7.3	-0.03
6g										
run 1	-32.7 (0.8)	-0.1 (0.2)	326.6 (0.5)	-0.0003 (0.0470)	0.100 (0.003)	0.0001 (0.0003)	20.7	-0.05	8.0	-0.03
run 2	-32.8 (0.8)	-0.1 (0.2)	326.9 (0.5)	-0.0004 (0.0470)	0.100 (0.003)	0.0002 (0.0003)	21.6	-0.05	8.7	-0.03
run 3	-34.8 (1.0)	-0.2 (0.3)	329.2 (0.5)	-0.0004 (0.0531)	0.106 (0.003)	0.0003 (0.0004)	20.3	-0.05	7.3	-0.03

	ΔH(Tm) (kcal/mol)	ΔCp (kcal/mol/K)	Tm (K)	ΔG0 (kcal/mol)	ΔG1 (kcal/mol/K)	ΔG2 (kcal/mol/K²)	N0 ([θ])	N1 ([θ]/K)	D0 ([θ])	D1 ([θ]/K)
6-F										
run 1	-33.2 (0.9)	-0.1 (0.3)	324.1 (0.5)	-0.0004 (0.0501)	0.102 (0.003)	0.0002 (0.0004)	22.7	-0.05	8.2	-0.03
run 2	-32.8 (0.9)	-0.1 (0.3)	324.2 (0.5)	-0.0004 (0.0523)	0.101 (0.003)	0.0002 (0.0004)	22.9	-0.05	7.5	-0.03
run 3	-32.4 (0.8)	-0.1 (0.2)	324.2 (0.5)	-0.0003 (0.0473)	0.100 (0.003)	0.0002 (0.0003)	23.2	-0.06	8.5	-0.03
6g-F										
run 1	-31.3 (0.9)	-0.1 (0.2)	324.4 (0.6)	-0.0004 (0.0607)	0.096 (0.003)	0.0002 (0.0004)	18.6	-0.04	7.6	-0.03
run 2	-30.0 (0.8)	-0.1 (0.2)	324.8 (0.6)	-0.0003 (0.0546)	0.092 (0.002)	0.0001 (0.0003)	19.0	-0.04	6.3	-0.02
run 3	-30.5 (0.7)	0.0 (0.2)	325.3 (0.5)	-0.0003 (0.0463)	0.094 (0.002)	0.0001 (0.0003)	18.7	-0.04	6.4	-0.02

	ΔH(Tm) (kcal/mol)	ΔCp (kcal/mol/K)	Tm (K)	ΔG0 (kcal/mol)	ΔG1 (kcal/mol/K)	ΔG2 (kcal/mol/K²)	N0 ([θ])	N1 ([θ]/K)	D0 ([θ])	D1 ([θ]/K)
6-T										
run 1	-31.0 (1.0)	-0.1 (0.2)	325.9 (0.7)	-0.0004 (0.0621)	0.095 (0.003)	0.0001 (0.0004)	20.9	-0.05	6.1	-0.02
run 2	-31.2 (0.9)	0.0 (0.2)	325.8 (0.5)	-0.0003 (0.0512)	0.096 (0.003)	0.0001 (0.0003)	20.8	-0.05	5.5	-0.02
run 3	-31.7 (0.9)	-0.1 (0.2)	325.1 (0.5)	-0.0003 (0.0526)	0.098 (0.003)	0.0001 (0.0004)	19.6	-0.05	6.4	-0.02
6g-T										
run 1	-31.1 (0.8)	-0.2 (0.2)	325.2 (0.6)	-0.0004 (0.0523)	0.095 (0.002)	0.0003 (0.0003)	14.8	-0.03	7.2	-0.02
run 2	-32.0 (0.8)	-0.1 (0.2)	325.5 (0.5)	-0.0004 (0.0485)	0.098 (0.003)	0.0002 (0.0003)	16.8	-0.04	5.4	-0.02
run 3	-31.3 (0.8)	-0.1 (0.2)	325.5 (0.5)	-0.0003 (0.0509)	0.096 (0.003)	0.0001 (0.0003)	16.0	-0.04	6.2	-0.02

	ΔH(Tm) (kcal/mol)	ΔCp (kcal/mol/K)	Tm (K)	ΔG0 (kcal/mol)	ΔG1 (kcal/mol/K)	ΔG2 (kcal/mol/K²)	N0 ([θ])	N1 ([θ]/K)	D0 ([θ])	D1 ([θ]/K)
6-F,T										
run 1	-29.9 (0.8)	0.0 (0.2)	320.9 (0.6)	-0.0003 (0.0533)	0.093 (0.002)	0.00002 (0.00032)	21.9	-0.055	7.0	-0.024
run 2	-30.9 (0.8)	0.0 (0.2)	321.1 (0.5)	-0.0003 (0.0490)	0.096 (0.003)	0.00002 (0.00033)	23.1	-0.059	5.8	-0.020
run 3	-31.4 (1.1)	-0.2 (0.3)	319.8 (0.9)	-0.0004 (0.0836)	0.098 (0.003)	0.0003 (0.0005)	21.7	-0.054	8.9	-0.029
6g-F,T										
run 1	-32.8 (0.9)	-0.2 (0.2)	328.3 (0.5)	-0.0004 (0.0531)	0.100 (0.003)	0.0003 (0.0004)	18.5	-0.043	7.8	-0.026
run 2	-32.5 (0.8)	-0.3 (0.2)	327.9 (0.6)	-0.0005 (0.0533)	0.099 (0.003)	0.0004 (0.0004)	18.0	-0.041	8.6	-0.029
run 3	-32.0 (0.9)	-0.2 (0.2)	328.4 (0.6)	-0.0004 (0.0571)	0.098 (0.003)	0.0003 (0.0004)	19.1	-0.044	7.2	-0.023

Fig. S3. CD spectra, variable-temperature CD data, and fit parameters (see Eqs. **S1**–**S4**) for (*A*) protein **6** and glycoprotein **6g**; (*B*) protein **6-F** and glycoprotein **6g-F**; (*C*) protein **6-T** and glycoprotein **6g-T**; and (*D*) protein **6-F,T**, and glycoprotein **6g-F,T** in 20 mM sodium phosphate buffer, pH 7. Variable-temperature CD experiments were conducted at a protein concentration of 10 μM. CD spectra for **6**, **6g**, **6-F**, and **6g-F** were obtained at 100 μM protein concentration. CD spectra for **6-T**, **6g-T**, **6-F,T**, and **6g-F,T** were obtained at 10 μM protein concentration. These spectra were reported previously (ref. 1), but are included here for comparison with the CD spectra of **4** and **5** and their derivatives.

Table S1. Pin WW domain protein sequences and MALDI-TOF data

Protein	Sequence*	Expected [M + H$^+$] (amu)†	Observed [M + H$^+$] (amu)
	6 10 15 20 25 30 35 39		
4	KLPPG WEKRM **S**–NG **R**VYYF NHITN ASQFE RPSG	3,766.9	3,766.8
4g	KLPPG WEKRM **S**–**N**G **R**VYYF NHITN ASQFE RPSG	3,969.9	3,972
4-F	KLPPG WEKRM **F**–NG **R**VYYF NHITN ASQFE RPSG	3,826.9	3,826.4
4g-F	KLPPG WEKRM **F**–**N**G **R**VYYF NHITN ASQFE RPSG	4,030.0	4,030.5
4-T	KLPPG WEKRM **S**–NG **T**VYYF NHITN ASQFE RPSG	3,711.8	3,711.8
4g-T	KLPPG WEKRM **S**–**N**G **T**VYYF NHITN ASQFE RPSG	3,914.9	3,916.3
4-F,T	KLPPG WEKRM **F**–NG **T**VYYF NHITN ASQFE RPSG	3,771.9	3,770.8
4g-F,T	KLPPG WEKRM **F**–**N**G **T**VYYF NHITN ASQFE RPSG	3,974.9	3,975.1
5	KLPPG WEKRM **S**-ANG **R**VYYF NHITN ASQFE RPSG	3,837.9	3,837.4
5g	KLPPG WEKRM **S**-A**N**G **R**VYYF NHITN ASQFE RPSG	4,041.0	4,041.7
5-F	KLPPG WEKRM **F**-ANG **R**VYYF NHITN ASQFE RPSG	3,897.9	3,898.1
5g-F	KLPPG WEKRM **F**-A**N**G **R**VYYF NHITN ASQFE RPSG	4,101.0	4,101.6
5-T	KLPPG WEKRM **S**-ANG **T**VYYF NHITN ASQFE RPSG	3,782.9	3,783.2
5g-T	KLPPG WEKRM **S**-A**N**G **T**VYYF NHITN ASQFE RPSG	3,985.9	3,986.2
5-F,T	KLPPG WEKRM **F**-ANG **T**VYYF NHITN ASQFE RPSG	3,842.9	3,842.7
5g-F,T	KLPPG WEKRM **F**-A**N**G **T**VYYF NHITN ASQFE RPSG	4,046.0	4,045.4
6	KLPPG WEKRM **S**R**S**NG **R**VYYF NHITN ASQFE RPSG	4,010.0	‡
6g	KLPPG WEKRM **S**RS**N**G **R**VYYF NHITN ASQFE RPSG	4,213.1	‡
6-F	KLPPG WEKRM **F**RS**N**G **R**VYYF NHITN ASQFE RPSG	4,070.0	‡
6g-F	KLPPG WEKRM **F**RS**N**G **R**VYYF NHITN ASQFE RPSG	4,273.1	‡
6-T	KLPPG WEKRM **S**R**S**NG **T**VYYF NHITN ASQFE RPSG	3,954.9	‡
6g-T	KLPPG WEKRM **S**RS**N**G **T**VYYF NHITN ASQFE RPSG	4,158.0	‡
6-F,T	KLPPG WEKRM **F**RS**N**G **T**VYYF NHITN ASQFE RPSG	4,015.0	‡
6g-F,T	KLPPG WEKRM **F**RS**N**G **T**VYYF NHITN ASQFE RPSG	4,218.1	‡

*N = Asn(GlcNAc).

†Monoisotopic masses.

‡Determined previously (1).