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

Joshua L. Price\textsuperscript{a,b}, David L. Powers\textsuperscript{c}, Evan T. Powers\textsuperscript{a,b,1}, and Jeffery W. Kelly\textsuperscript{a,b,d,1}

\textsuperscript{a}Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037; \textsuperscript{b}The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037; \textsuperscript{c}Department of Mathematics, Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699; and \textsuperscript{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 naive 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 \( \beta \)-turn harboring a G1 \( \beta \)-bulge (using a Phe–Yyy–Asn–Xxx–Thr sequon) and a type II \( \beta \)-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 glycosylation of a novel type II \( \beta \)-bulge, Phe–Asn–Xxx–Thr, is a type I \( \beta \)-turn in three distinct host WW domains reveals that the glycosylation-mediated stabilization is greatest for the enhanced aromatic sequon in its respective reverse turn context. Adding the Phe–Asn–Xxx–Thr motif (in a type I \( \beta \)-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.

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\textsubscript{3}Man\textsubscript{5}GlcNAc\textsubscript{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 translated into the endoplasmic reticulum (2–4). The N-linked oligosaccharide, or O-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), so installing the Phe–Asn–Xxx–Thr enhanced aromatic sequon could be an attractive strategy for increasing the stability of the many proteins that harbor type I \( \beta \)-bulge turns. Identifying other suitable reverse turn types could position Phe, GlcNAc and Thr close enough to facilitate a tripartite interaction. We have expanded the number of proteins that can 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), which in which antiparallel \( \beta \)-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 \( \beta \)-turn (Fig. 1B); 0.1% of the reverse turns in the PDB have this conformation (33). The side-chain beta carbons (CPs) 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, GlcNAc and Thr, similar to the interactions observed in the glycosylated type I \( \beta \)-bulge turn of HsCD2ad (Fig. 1A). 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, GlcNAc\textsubscript{1}) 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–Zzz–Asn–Xxx–Thr sequon in the type I \( \beta \)-bulge turns of RnCD2ad (Fig. 1B). 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).

Published structural data (31) from the human ortholog of RnCD2ad (HsCD2ad, Fig. 1A) suggest that placement of an N-glycan at i + 2 in the type I \( \beta \)-bulge turn context allows the \( \alpha \)-face of GlcNAc\textsubscript{1} of the N-glycan to engage in stabilizing hydrophobic interactions with the aromatic ring of Phe at the i position (a stabilizing C-H\(\pi\) 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).

Nearby 9% of the reverse turns in the Protein Data Bank (PDB) are type I \( \beta \)-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 \( \beta \)-bulge turns. Identifying other suitable reverse turn types that could position Phe, GlcNAc and Thr close enough to facilitate a tripartite interaction would further expand the number of proteins that can benefit from the increased stability and possibly the increased glycosylation efficiency afforded by the enhanced aromatic sequon.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To 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.
The stabilizing tripartite interaction between Phe, GlcNAc, and Thr has been characterized. Crystal structures exist for WW domains harboring the type II β-turn within the six-residue loop (Fig. 1B), a five-residue type I β-bulge turn (Fig. 1C), and a four-residue type I β-turn (Fig. 1D) as loop 1 (34, 36). We hypothesized that a type I β-turn, which makes up 1% of the reverse turns in the PDB (33), would also be an additional conformational host for a complementary enhanced aromatic sequon: the CP’s of the side chains at the $i$, $i + 1$, and $i + 3$-positions are close enough (<5.6 Å; see Fig 1D) 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 Phe–Asn(GlcNAc1)19–Gly20–Thr21, Phe–Asn(GlcNAc1)19–Gly20–Thr21, and Phe–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, 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, 4g-F, T, 5-F, 5g-F, T, 6-F, and 6g-F, T and...
Table 1. Sequences, melting temperatures, and folding free energies of Pin WW-derived glycoproteins and their nonglycosylated counterparts*

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence’</th>
<th>$T_m$ (°C)</th>
<th>$\Delta T_m$ (°C)</th>
<th>$\Delta G_f$ (kcal/mol)</th>
<th>$\Delta \Delta G_f$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>MS-NGR</td>
<td>64.4 ± 0.4</td>
<td>2.2 ± 0.6</td>
<td>0.06 ± 0.04</td>
<td>-0.23 ± 0.06</td>
</tr>
<tr>
<td>4g</td>
<td>MS-NGR</td>
<td>66.6 ± 0.4</td>
<td></td>
<td>-0.17 ± 0.04</td>
<td></td>
</tr>
<tr>
<td>4-F</td>
<td>MF-NGR</td>
<td>66.7 ± 0.5</td>
<td>1.5 ± 0.7</td>
<td>-0.18 ± 0.08</td>
<td>-0.18 ± 0.08</td>
</tr>
<tr>
<td>4g-F</td>
<td>MF-NGR</td>
<td>68.2 ± 0.5</td>
<td></td>
<td>-0.36 ± 0.05</td>
<td></td>
</tr>
<tr>
<td>4-T</td>
<td>MS-NGT</td>
<td>62.2 ± 0.4</td>
<td>-0.8 ± 0.6</td>
<td>0.30 ± 0.04</td>
<td>0.07 ± 0.07</td>
</tr>
<tr>
<td>4g-T</td>
<td>MS-NGT</td>
<td>61.4 ± 0.5</td>
<td></td>
<td>0.37 ± 0.05</td>
<td></td>
</tr>
<tr>
<td>4-F,T</td>
<td>MF-NGT</td>
<td>63.5 ± 0.3</td>
<td></td>
<td>0.18 ± 0.03</td>
<td>-0.39 ± 0.09</td>
</tr>
<tr>
<td>4g-F,T</td>
<td>MF-NGT</td>
<td>66.7 ± 0.6</td>
<td>3.2 ± 0.7</td>
<td>-0.21 ± 0.08</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>MS-NGR</td>
<td>68.7 ± 0.2</td>
<td>0.6 ± 0.3</td>
<td>-0.38 ± 0.02</td>
<td>-0.07 ± 0.04</td>
</tr>
<tr>
<td>5</td>
<td>MS-NGR</td>
<td>69.3 ± 0.2</td>
<td></td>
<td>-0.46 ± 0.03</td>
<td></td>
</tr>
<tr>
<td>5-F</td>
<td>MF-NGR</td>
<td>65.2 ± 0.3</td>
<td>5.0 ± 0.4</td>
<td>-0.55 ± 0.04</td>
<td></td>
</tr>
<tr>
<td>5g-F</td>
<td>MF-NGR</td>
<td>70.3 ± 0.2</td>
<td></td>
<td>-0.58 ± 0.02</td>
<td></td>
</tr>
<tr>
<td>5-T</td>
<td>MS-ANGT</td>
<td>68.9 ± 0.2</td>
<td>2.4 ± 0.3</td>
<td>-0.42 ± 0.02</td>
<td>-0.23 ± 0.03</td>
</tr>
<tr>
<td>5g-T</td>
<td>MS-ANGT</td>
<td>71.3 ± 0.3</td>
<td></td>
<td>-0.65 ± 0.03</td>
<td></td>
</tr>
<tr>
<td>5-F,T</td>
<td>MF-NGT</td>
<td>66.0 ± 0.2</td>
<td>9.2 ± 0.2</td>
<td>-0.11 ± 0.02</td>
<td>-0.94 ± 0.03</td>
</tr>
<tr>
<td>5g-F,T</td>
<td>MF-ANGT</td>
<td>75.2 ± 0.2</td>
<td></td>
<td>-1.05 ± 0.02</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>MS-NGR</td>
<td>56.2 ± 0.3</td>
<td>-2.6 ± 0.4</td>
<td>0.95 ± 0.04</td>
<td>0.21 ± 0.06</td>
</tr>
<tr>
<td>6g</td>
<td>MS-NGR</td>
<td>53.6 ± 0.3</td>
<td></td>
<td>1.16 ± 0.04</td>
<td></td>
</tr>
<tr>
<td>6-F</td>
<td>MF-NGR</td>
<td>51.0 ± 0.3</td>
<td>0.7 ± 0.4</td>
<td>1.45 ± 0.06</td>
<td>-0.17 ± 0.08</td>
</tr>
<tr>
<td>6g-F</td>
<td>MF-NGR</td>
<td>51.7 ± 0.3</td>
<td></td>
<td>1.28 ± 0.04</td>
<td></td>
</tr>
<tr>
<td>6-T</td>
<td>MS-ANGT</td>
<td>52.5 ± 0.3</td>
<td>-0.2 ± 0.5</td>
<td>1.22 ± 0.05</td>
<td>0.04 ± 0.07</td>
</tr>
<tr>
<td>6g-T</td>
<td>MS-ANGT</td>
<td>52.3 ± 0.3</td>
<td></td>
<td>1.26 ± 0.05</td>
<td></td>
</tr>
<tr>
<td>6-F,T</td>
<td>MF-NGR</td>
<td>47.4 ± 0.4</td>
<td>7.6 ± 0.5</td>
<td>1.72 ± 0.09</td>
<td>-0.70 ± 0.10</td>
</tr>
<tr>
<td>6g-F,T</td>
<td>MF-NGR</td>
<td>55.0 ± 0.3</td>
<td></td>
<td>1.02 ± 0.04</td>
<td></td>
</tr>
</tbody>
</table>

*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.

N = 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 $\Delta G_f$ (at 65 °C) for each protein and corresponding glycoprotein, along with the effect of glycosylation on the $T_m$ and $\Delta G_f$ (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 $\Delta G_f$ to temperatures outside the transition region using thermodynamic parameter estimates from fits to variable temperature CD data is unreliable (because errors in $\Delta C_p$, 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 $\Delta C_p$, the differences between their $T_m$ values should reflect the differences between their $\Delta G_f$ 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 ($\Delta \Delta G_f = -0.39 ± 0.09$ 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 $\beta$-turn stabilizes WW. Glycosylating the Phe–Yyy–Asn–Xxx–Thr sequon in the context of the five-residue type I $\beta$-turn also stabilizes WW ($\Delta T_m = 9.2 ± 0.2$ °C; $\Delta \Delta G_f = -0.94 ± 0.03$ kcal mol$^{-1}$ at 65 °C), as does glycosylating the Phe–Yyy–Zzz–Asn–Xxx–Thr sequon in the type II $\beta$-turn in a six-residue loop ($\Delta T_m = 7.6 ± 0.5$ °C, $\Delta \Delta G_f = -0.70 ± 0.10$ 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 $\beta$-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. 24), reveals that glycosylation of Asn19 (in the presence of Arg21) stabilizes glycoprotein 4g relative to protein 4 ($\Delta \Delta G_f = -0.23 ± 0.06$ kcal mol$^{-1}$ at 65 °C). Similarly, glycosylation of Asn19 (in the presence of Arg21) stabilizes 4g-F relative to 4g ($\Delta \Delta G_f = -0.18 ± 0.08$ kcal mol$^{-1}$ at 65 °C). The difference between $\Delta \Delta G_f$ and $\Delta \Delta G_{f,front}$ ($0.05 ± 0.10$ 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 $\beta$-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).](image-url)
Chang 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) reveals that in the presence of Thr21 (instead of Arg21), Phe16 and Asn(GlcNAc1) interact favorably (\( \Delta \Delta G_{i\text{-back}} = -0.46 \pm 0.11 \text{ kcal mol}^{-1} \) 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), and Thr21. The large difference between \( \Delta \Delta G_{i\text{-front}} \) and \( \Delta \Delta G_{i\text{-back}} \) for the four-residue type I \( \beta \)-turn (\( \Delta \Delta \Delta \Delta G_i = -0.51 \pm 0.15 \text{ kcal mol}^{-1} \) at 65°C) indicates that Phe16, Asn(GlcNAc1), 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 (Figs. 2B and C) reveal a favorable interaction between Phe16, Asn(GlcNAc1), and Thr21, in the five-residue type I \( \beta \)-bulge turn (\( \Delta \Delta \Delta \Delta G_i = -0.23 \pm 0.07 \text{ kcal mol}^{-1} \) at 65°C) and in the type II \( \beta \)-turn in a six-residue loop (\( \Delta \Delta \Delta \Delta G_i = -0.36 \pm 0.15 \text{ kcal mol}^{-1} \) at 65°C). This three-way interaction between Phe16, Asn(GlcNAc1), and Thr21 is similarly favorable in each reverse turn context (perhaps more favorable in the type I \( \beta \)-turn than in the type I \( \beta \)-bulge turn, but recall that this is only part of the overall stabilizing effect of N-glycosylation).

The attraction of \( \Delta \Delta \Delta \Delta G_{i\text{-front}} \) and \( \Delta \Delta \Delta \Delta G_{i\text{-back}} \) values to the interaction between Phe16 and Asn(GlcNAc1), and of \( \Delta \Delta \Delta \Delta G_{i\text{-i}} \) to the tripartite interaction among Phe16, Asn(GlcNAc1), and Thr21, assumes that the Ser16 side chain does not interact with the main chain (see Fig. 1B–D). The lone exception is an interaction between the side-chain hydroxil of Ser16 and the side-chain carboxylate of Asp31 in the type I \( \beta \)-bulge turn (Fig. 1C). However, the equivalent interaction (between the Ser16 hydroxil and the Asn19 side-chain carboxyl in) the variants of 5 that have Ser at position 16 (5, 5g, 5-T, and 5g-T) should be the same whether Asn 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 \( \beta \)-turns and type I \( \beta \)-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:

\[
\Delta G_i = \Delta G_{i}' + C_{EFN} \cdot W_{EF} + C_{EN} \cdot W_N + C_{CT} \cdot W_T + C_{FNT} \cdot W_{FNT} + C_{NT} \cdot W_{NT} + C_{T} \cdot W_T .
\]

**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*

<table>
<thead>
<tr>
<th>Type I ( \beta )-turn</th>
<th>Type I ( \beta )-bulge turn</th>
<th>Type II ( \beta )-turn in six-residue loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta G_i )</td>
<td>0.06 ± 0.06 (0.287)</td>
<td>-0.38 ± 0.04 (0.000)</td>
</tr>
<tr>
<td>( C_{EF} )</td>
<td>-0.24 ± 0.08 (0.005)</td>
<td>0.36 ± 0.06 (0.000)</td>
</tr>
<tr>
<td>( C_{EN} )</td>
<td>-0.23 ± 0.06 (0.009)</td>
<td>-0.07 ± 0.06 (0.248)</td>
</tr>
<tr>
<td>( C_{CT} )</td>
<td>0.24 ± 0.08 (0.557)</td>
<td>-0.04 ± 0.06 (0.557)</td>
</tr>
<tr>
<td>( C_{FNT} )</td>
<td>0.05 ± 0.11 (0.661)</td>
<td>-0.48 ± 0.09 (0.000)</td>
</tr>
<tr>
<td>( C_{NT} )</td>
<td>0.15 ± 0.11 (0.168)</td>
<td>-0.05 ± 0.09 (0.562)</td>
</tr>
<tr>
<td>( C_{T} )</td>
<td>0.31 ± 0.12 (0.015)</td>
<td>-0.16 ± 0.09 (0.088)</td>
</tr>
<tr>
<td>( C_{FNT} )</td>
<td>-0.54 ± 0.15 (0.001)</td>
<td>-0.23 ± 0.12 (0.078)</td>
</tr>
</tbody>
</table>

*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.

**Citations:**

and Asn(GlcNAc)19 (Thr21 (three-way interaction between Phe16, Asn(GlcNAc)19, and Thr21 stabilizes each reverse turn type by similar amounts.

In the type I reverse turn context is stabilizing, though the origins of this Glycosylating an enhanced aromatic sequon in its correlated Discussion

19, and Thr21 stabilizes each reverse turn type by similar amounts.

The two-way interaction between Asn(GlcNAc)19 and Thr21 stabilizes the five- and six-residue turns (C β-β-β-turns within a six-residue loop. Each appears to facilitate the three-way interaction between Phe16, Asn(GlcNAc)19, and Thr21 stabilizes each reverse turn type by similar amounts.

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 β-β-β-turns within a six-residue loop. Each appears to facilitate the three-way interaction between Phe16, Asn(GlcNAc)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(GlcNAc)19, and Thr21 (C β-β-β-turn) and from the Asn19 to Asn(GlcNAc)19 mutation (C N, respectively) comprise less than 9% of all reverse turns in the PDB (28, 33). By successfully applying the Phe–Asn–Xxx–Thr for type I β-turns, Phe–Yyy–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 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–Zxx–Asn–Xxx–Thr sequences, 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-(4-(benzotriazol-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(ΔGlcNAc)-Ort (Nα-Fmoc-Asn[3,4,6-tri-O-acetyl-2-(acetylamino)-deoxy-2-β-D-glucopyranosyl]-L-asparagine). Piperidine and N,N-disopropylethylamine were procured from Aldrich, and N,N-dimethylformamide was obtained from Fisher.

Protected groups are being 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 mL), thiaoisole (500 mL), ethanediol (250 mL), and triisopropylsilane (100 mL) 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-hydroxy 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 spectrophotometrically from tyr- 
osine and tryptophan absorbance at 280 nm in 6 M guanidine hydrochloride 
20 mM sodium phosphate (ϵ<sub>280</sub> = 5690 M<sup>-1</sup>cm<sup>-1</sup>, ϵ<sub>295</sub> = 1280 M<sup>-1</sup>cm<sup>-1</sup>) as described previously (41, 60). CD spectra were obtained by monitoring molar 
elipticity from 227 nm from 0.2 °C to 98.2 °C or 108.2 °C at 2 °C intervals, with 
elipticity at 227 nm from 0.2 °C to 98.2 °C or 108.2 °C at 2 °C intervals, with five-second averaging 
itivity to the variable temperature CD data to obtain 
th between data points and 30 s averaging times. We 
22. O
31. Wyss DF, et al. (1995) Conformation and function of the N-linked glycan in the adhe-
33. Oliva B, Baten PA, Querol E, Aviles FX, Sternberg MJE (1997) An automated classifica-
tional switching. 
32. Li HJ, d
27. Hutchinson EG, Thornton JM (1994) A revised set of potentials for beta-turn formation in 
28. O
a systematic classification with applications to modeling by homology, — 
23. O
20. O
55. Hackenberger CPR, Friel CT, Imperiali B (2005) Semisynthesis of a glyco-
32. Wyss DF, et al. (1995) Conformation and function of the N-linked glycan in the adhe-
27. Hutchinson EG, Thornton JM (1994) A revised set of potentials for beta-turn formation in 
37. Hackenberger CPR, Friel CT, Imperiali B (2005) Semisynthesis of a glyco-
50. Grant GM051105 (J.W.K. and E.T.P), the Skaggs Institute for Chemical Biology, 
55. Hackenberger CPR, Friel CT, Imperiali B (2005) Semisynthesis of a glyco-
41. O
47. Li HJ, d
31. Wyss DF, et al. (1995) Conformation and function of the N-linked glycan in the adhe-

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 cysteamine and utilized in folding studies. The complete amino acid sequences identified by matrix-assisted laser desorption/ionization time-of-flight spectrometry (MALDI-TOF, Table S1), and purity was established by analytical HPLC.

**Deprotecting Glycosylated Pin WW Domain Proteins.** Acetate protecting groups were removed from the 3-, 4-, and 6-hydroxy groups on the Asn-linked GlcNAc residues in glycopeptides 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 hydrolysis 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 spectrophotometrically based on tyrosine and tryptophan absorbance at 280 nm in 6 M GdnHCl + 20 mM sodium phosphate (ε_Tyr = 5690 M⁻¹ cm⁻¹, ε_Trp = 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:

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

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:
\[
K_f = \exp \left[ \frac{-\Delta G_f(T)}{RT} \right].
\]  

[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:

\[
\Delta G_f(T) = \frac{\Delta H(T_m)}{T_m} (T_m - T) + \Delta C_p \left[ T - T_m - T \ln \left( \frac{T}{T_m} \right) \right].
\]  

[S3]

where \( \Delta H(T_m) \) is the enthalpy of folding at the melting temperature and \( \Delta C_p \) is the heat capacity of folding (\( \Delta H(T_m), \Delta C_p, \) 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:

\[
\Delta G_f(T) = \Delta G_0 + \Delta G_1 \times (T - T_m) + \Delta G_2 \times (T - T_m)^2.
\]  

[S4]

where \( \Delta G_0, \Delta G_1, \) and \( \Delta G_2 \) 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, \( \Delta G_0 \) is the free energy of folding at \( T_m \) (and is therefore very close to 0), \( \Delta G_1 \) is equal to \( d\Delta G_f/dT \), or \( -\Delta S_f(T_m) \) (the entropy at \( T_m \)) and \( \Delta G_2 \) is 

\[\begin{align*}
\end{align*}\]
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-F, and 4g-F 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.
Fig. S2. CD spectra, variable-temperature CD data, and fit parameters (see Eqs. S1–S4) for (A) protein S and glycoprotein Sg; (B) protein S-F and glycoprotein Sg-F; (C) protein S-T and glycoprotein Sg-T; and (D) protein S-F,T, and glycoprotein Sg-F,T in 20 mM sodium phosphate buffer, pH 7. CD experiments were conducted at a protein concentration of 10 μM.
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>
<thead>
<tr>
<th>Protein</th>
<th>Sequence*</th>
<th>Expected $[M + H^+]$ (amu)†</th>
<th>Observed $[M + H^+]$ (amu)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>KLPQG WEKRM S-NG RVYYF NHITN ASQFE RPSG</td>
<td>3,766.9</td>
<td>3,766.8</td>
</tr>
<tr>
<td>4g</td>
<td>KLPQG WEKRM S-NG RVYYF NHITN ASQFE RPSG</td>
<td>3,969.9</td>
<td>3,972</td>
</tr>
<tr>
<td>4-F</td>
<td>KLPQG WEKRM F-NG RVYYF NHITN ASQFE RPSG</td>
<td>3,826.9</td>
<td>3,826.4</td>
</tr>
<tr>
<td>4g-F</td>
<td>KLPQG WEKRM S-NG RVYYF NHITN ASQFE RPSG</td>
<td>4,030.0</td>
<td>4,030.5</td>
</tr>
<tr>
<td>4-T</td>
<td>KLPQG WEKRM S-NG TVYYF NHITN ASQFE RPSG</td>
<td>3,711.8</td>
<td>3,711.8</td>
</tr>
<tr>
<td>4g-T</td>
<td>KLPQG WEKRM S-NG TVYYF NHITN ASQFE RPSG</td>
<td>3,914.9</td>
<td>3,916.3</td>
</tr>
<tr>
<td>4-F,T</td>
<td>KLPQG WEKRM F-NG TVYYF NHITN ASQFE RPSG</td>
<td>3,771.9</td>
<td>3,770.8</td>
</tr>
<tr>
<td>4g-F,T</td>
<td>KLPQG WEKRM F-NG TVYYF NHITN ASQFE RPSG</td>
<td>3,974.9</td>
<td>3,975.1</td>
</tr>
<tr>
<td>5</td>
<td>KLPQG WEKRM S-ANG RVYYF NHITN ASQFE RPSG</td>
<td>3,837.9</td>
<td>3,837.4</td>
</tr>
<tr>
<td>5g</td>
<td>KLPQG WEKRM S-ANG RVYYF NHITN ASQFE RPSG</td>
<td>4,041.0</td>
<td>4,041.7</td>
</tr>
<tr>
<td>5-F</td>
<td>KLPQG WEKRM F-ANG RVYYF NHITN ASQFE RPSG</td>
<td>3,897.9</td>
<td>3,898.1</td>
</tr>
<tr>
<td>5g-F</td>
<td>KLPQG WEKRM F-ANG RVYYF NHITN ASQFE RPSG</td>
<td>4,101.0</td>
<td>4,101.6</td>
</tr>
<tr>
<td>5-T</td>
<td>KLPQG WEKRM S-ANG TVYYF NHITN ASQFE RPSG</td>
<td>3,782.9</td>
<td>3,783.2</td>
</tr>
<tr>
<td>5g-T</td>
<td>KLPQG WEKRM S-ANG TVYYF NHITN ASQFE RPSG</td>
<td>3,985.9</td>
<td>3,986.2</td>
</tr>
<tr>
<td>5-F,T</td>
<td>KLPQG WEKRM F-ANG TVYYF NHITN ASQFE RPSG</td>
<td>3,842.9</td>
<td>3,842.7</td>
</tr>
<tr>
<td>5g-F,T</td>
<td>KLPQG WEKRM F-ANG TVYYF NHITN ASQFE RPSG</td>
<td>4,046.0</td>
<td>4,045.4</td>
</tr>
<tr>
<td>6</td>
<td>KLPQG WEKRM SRSNG RVYYF NHITN ASQFE RPSG</td>
<td>4,010.0</td>
<td>4</td>
</tr>
<tr>
<td>6g</td>
<td>KLPQG WEKRM SRSNG RVYYF NHITN ASQFE RPSG</td>
<td>4,213.1</td>
<td></td>
</tr>
<tr>
<td>6-F</td>
<td>KLPQG WEKRM FRSGN RVYYF NHITN ASQFE RPSG</td>
<td>4,070.0</td>
<td></td>
</tr>
<tr>
<td>6g-F</td>
<td>KLPQG WEKRM FRSGN RVYYF NHITN ASQFE RPSG</td>
<td>4,273.1</td>
<td></td>
</tr>
<tr>
<td>6-T</td>
<td>KLPQG WEKRM FRSGN TVYYF NHITN ASQFE RPSG</td>
<td>3,954.9</td>
<td></td>
</tr>
<tr>
<td>6g-T</td>
<td>KLPQG WEKRM FRSGN TVYYF NHITN ASQFE RPSG</td>
<td>4,158.0</td>
<td></td>
</tr>
<tr>
<td>6-F,T</td>
<td>KLPQG WEKRM FRSGN TVYYF NHITN ASQFE RPSG</td>
<td>4,015.0</td>
<td></td>
</tr>
<tr>
<td>6g-F,T</td>
<td>KLPQG WEKRM FRSGN TVYYF NHITN ASQFE RPSG</td>
<td>4,218.1</td>
<td></td>
</tr>
</tbody>
</table>

*N = Asn(GlcNAc).
†Monoisotopic masses.
‡Determined previously (1).