THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

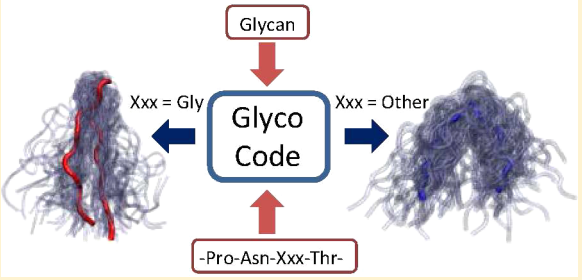
Deciphering the Glycosylation Code

Christopher R. Ellis and William G. Noid*

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, United States

Supporting Information

ABSTRACT: Asparagine-linked carbohydrates profoundly impact glycoprotein folding, stability, and structure. However, the "glycosylation code" that relates these effects to protein sequence remains unsolved. We report atomically detailed replica exchange molecular dynamics simulations in explicit solvent that systematically investigate the impact of glycosylation upon peptides with the central sequon Pro-Asn-Gly/Ala-Thr-Trp/Ala. These simulations suggest that the effects of glycosylation may be quite sensitive to steric crowding by the side chain immediately following the glycosylation site but less sensitive to stacking interactions with the aromatic Trp residue. In addition, we compare our simulated ensembles with the known structures for full length glycoproteins. These structures corroborate the simulations and also suggest a remarkable consistency between the intraprotein and protein-glycan interactions of natural glycoproteins. Moreover, our analysis highlights the significance of left-handed conformations for compact β-hairpins at glycosylation sites. In summary, these studies elucidate basic biophysical principles for the glycosylation code.



INTRODUCTION

Many eukaryotic proteins are modified by the covalent attachment of a bulky carbohydrate to an Asn side chain within the canonical sequon Asn-Xxx-Thr/Ser, where Xxx is any residue other than proline.[1,2] The Asn(N)-linked glycan determines the biological function of glycoproteins[3,4] and also dramatically impacts their structure, dynamics, folding, and stability.[5,6] It has been suggested that a "glycosylation code" relates the biophysical effects of glycosylation to protein sequence and structure.[7] The key to this code would not only elucidate important biophysical principles for glycobiology but also hold significant ramifications for engineering novel protein folds[8−10] and improving glycoprotein therapeutics.[11] Unfortunately, the glycosylation code remains unsolved.

Previous studies have proposed several general principles for the biophysical effects of glycosylation. For instance, elegant studies have highlighted the significance of steric crowding by N-linked glycans.[12,13] However, more recent studies suggest that the glycosylation code reflects specific, sequence-dependent interactions,[14−18] such as van der Waals interactions,[19] hydrogen bonds,[20] and "stacking interactions" between glycans and aromatic side chains.[21] Significantly, although N-linked glycosylation involves a complex branched carbohydrate, previous studies have demonstrated that the "intrinsic" effects of glycosylation upon protein biophysics result primarily from the core chitobiose disaccharide that is directly linked to the Asn side chain.[22,23]

In a series of important experiments, Imperiali and co-workers demonstrated that an N-linked chitobiose disaccharide triggered a "conformational switch" from extended Asx-turns to compact β-turns in a model peptide with the central sequon Pro-Asn-Gly-Thr-Trp.[24−27] Here we investigate the sensitivity of this conformational switch to the peptide sequence via atomistic molecular dynamics simulations in explicit solvent. Furthermore, in order to corroborate our simulations and assess their significance for actual glycoproteins, we analyzed the Structural Assessment of Glycosylation Sites (SAGS) database of nonredundant glycosylation sites in full length glycoproteins.[28−30]

In particular, our simulations considered a series of peptides with the sequence Ace1-Ile2-Thr3-Pro4-Asn5-Gly/Ala6-Thr7-Trp/Ala8-Ala9-NH_2, as summarized in Table 1. While the Gly6 → Ala6 mutation assesses the importance of flexible residues adjacent to the glycosylation site, the Trp8 → Ala8 mutation assesses the importance of aromatic−glycan stacking interactions. The simulated Gly6-Trp8 sequence differs from the model system of Imperiali and co-workers[24−27] only in the omission of an ornithine residue prior to Ile2. The Gly6-Trp8 sequence also bears important similarities and differences with the "enhanced aromatic sequons" that have been previously studied by Kelly and co-workers.[8−10] Both the simulated Gly6-

Table 1. Simulated Peptide Sequons

simulated residue	Res4	Res5	Res6	Res7	Res8
aligned residue	−1	0	+1	+2	+3
Gly6-Trp8	Pro4	Asn5	Gly6	Thr7	Trp8
Gly6-Ala8	Pro4	Asn5	Gly6	Thr7	**Ala8**
Ala6-Trp8	Pro4	Asn5	**Ala6**	Thr7	Trp8

Received: August 23, 2014
Revised: September 3, 2014
Published: September 4, 2014

   dx.doi.org/10.1021/jp508535f | J. Phys. Chem. B 2014, 118, 11462−11469

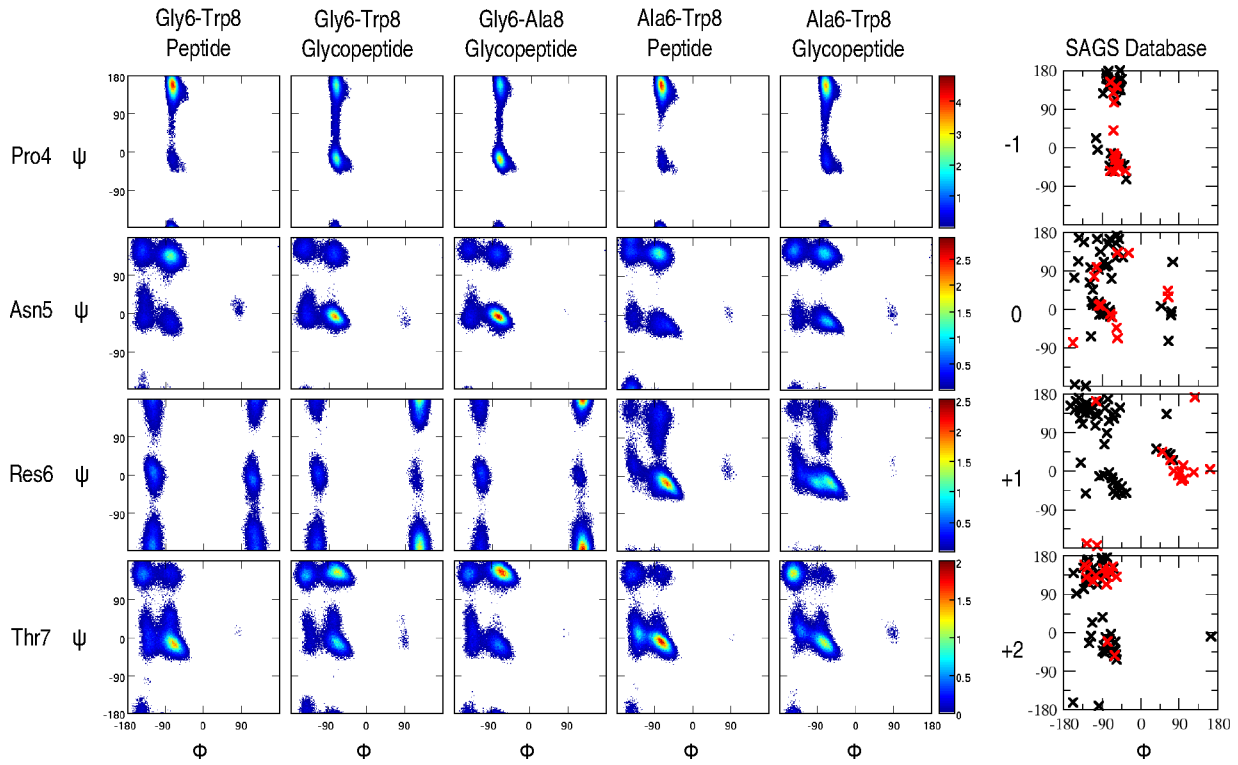


Figure 1. Impact of glycosylation upon Ramachandran maps of neighboring residues. Columns 1−5 correspond to the REMD simulations, while the last column corresponds to the 61 glycosylated Pro-Asn-Xxx-Thr/Ser sequons in the SAGS database. In this column, red X's indicate the 14 glycosylated sequons with Gly at the Xxx position, while black X's indicate the remaining 47 sequons.

Trp8 sequon and also the enhanced aromatic sequons feature a prominent aromatic residue that interacts with the N-linked glycan, as well as a Gly residue immediately after the glycosylation site. However, the key aromatic residue precedes the glycosylation site in the enhanced aromatic sequons (i.e., it aligns with Ile2 or Thr3), while it follows the glycosylation site in the simulated Gly6-Trp8 sequon. Moreover, in contrast to the enhanced aromatic sequons, a Pro residue immediately precedes the glycosylation site in the simulated sequons.

METHODS

Molecular Dynamics Simulations. We employed Gromacs 4.5.3[31,32] to perform a 100 ns replica exchange molecular dynamics[33] simulation of each peptide and also the corresponding glycopeptide with an N-linked chitobiose disaccharide. The initial conditions for these peptides were constructed with either the PRODRG server[34] or Pymol.[35] These simulations employed the SPC/E model[36] to explicitly describe the solvent and the OPLS-AA model[37,38] to describe the peptide and the glycan in atomic detail, while adapting 18 parameters for the Asn-linkage. Each simulation was performed in the constant NPT ensemble with 60 replicas at temperatures between 298 and 495 K, according to standard protocols.[32,39−44] The Supporting Information completely describes these simulations.

Bioinformatic Analysis. We analyzed the structural assessment of glycosylation sites (SAGS) database of nonredundant glycosylation sites in full length glycoproteins.[28−30] We excluded all sequons that did not provide complete structural information for the protein backbone within 3 residues of the glycosylation site. In addition, we excluded sequons in which the proximal glycan (i.e., the glycan covalently linked to the Asn side chain) was fucosylated.[45,46] In the following, we report results from our analysis of the remaining 1524 structures for N-linked glycosylation sites. Interestingly, though, Figure S5 of the Supporting Information suggests that fucosylation of the proximal glycan does not significantly influence the trends that are reported below, except in the 3 cases that the proximal glycan is fucosylated at both O3 and O6. In these 3 cases, the glycosylated sequon adopts an extended conformation. We also note that, while our simulations considered sequons with Thr two residues after Asn, we did not observe any remarkable differences in the SAGS database between sequons with Thr or with Ser in this position (see Figure S6 of the Supporting Information.) Thus, the following bioinformatic analysis does not distinguish between these two cases. The Supporting Information provides greater detail regarding our bioinformatic analysis.

RESULTS AND DISCUSSION

Local Torsional Preferences. Figure 1 presents Ramachandran maps that characterize the impact of glycosylation upon local torsional preferences. The first 5 columns analyze simulations of systems with the central Pro4-Asn5-Gly/Ala6-Thr7 sequon. The last column analyzes the 61 nonredundant instances of glycosylated Pro-Asn-Xxx-Thr/Ser sequons in the SAGS database, where Xxx is any residue. We number the residues in these aligned sequons as −1 through +2 with 0 being the glycosylated Asn.

The first three columns of Figure 1 present simulated Ramachandran maps for the nonglycosylated Gly6-Trp8 peptide, the Gly6-Trp8 glycopeptide, and the Gly6-Ala8 glycopeptide, respectively. The Gly6-Trp8 and Gly6-Ala8 peptides exhibit similar torsional preferences (see the

dx.doi.org/10.1021/jp508535f | J. Phys. Chem. B 2014, 118, 11462−11469

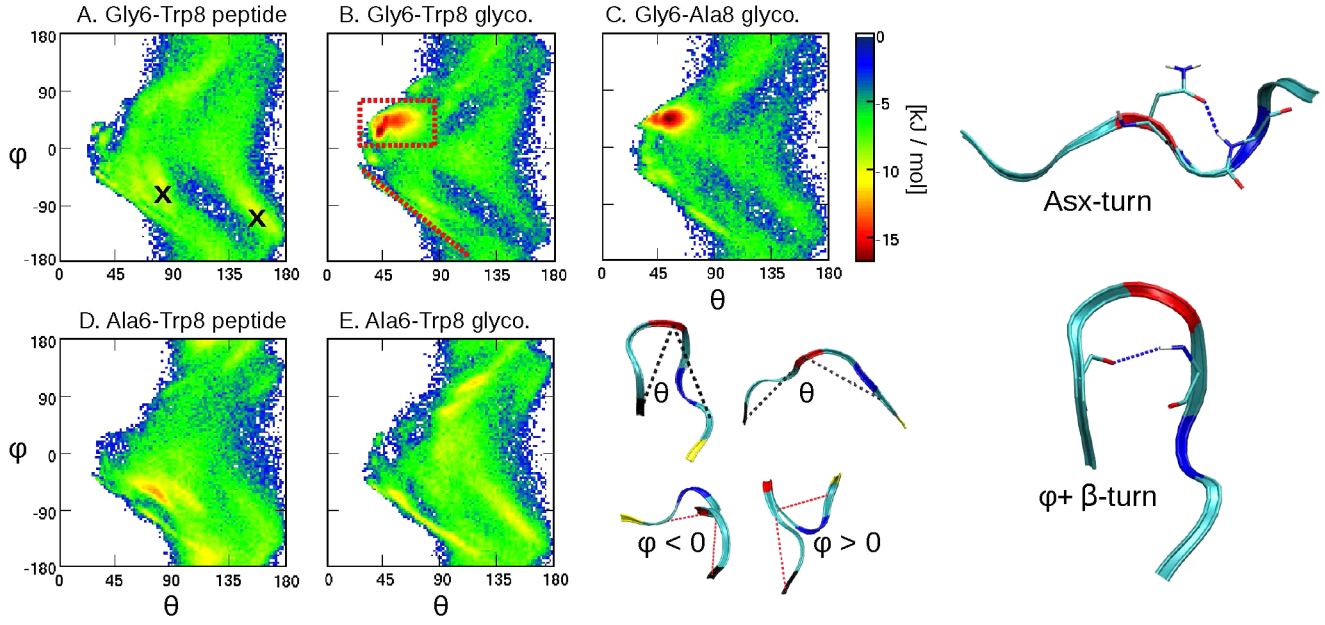


Figure 2. Simulated free-energy surfaces and prominent structures. Red and blue indicate Asn5 and Thr7, respectively.

Supporting Information). Glycosylation dramatically impacts the torsional preferences of the Gly6-Trp8 peptide and exerts strikingly similar, or perhaps even slightly greater, influence upon the Gly6-Ala8 peptide. In both cases, glycosylation shifts Pro4 and Asn5 from extended toward helical conformations but shifts Thr7 from helical toward extended conformations. Most significantly, while Gly6 samples left- ($\phi > 0°$) and right- ($\phi < 0°$) handed conformations with similar probability prior to glycosylation, glycosylation significantly stabilized left-handed conformations of Gly6. Moreover, the Supporting Information demonstrates that glycosylation introduces strong correlations between the conformations of Pro4 and the following residues.

The red X's in the last column indicate the Ramachandran angles sampled by the 14 glycosylated instances of the Pro-Asn-Gly-Thr/Ser sequon in the SAGS data set. These structures are remarkably consistent with the glycan-induced shifts that we observe in REMD simulations of short octapeptides. In these 14 structures, Pro and Asn sample helical conformations in 71% and 64% of the cases, respectively, while Thr is extended in 86% of the cases. Moreover, although Gly samples the basin at $\phi \approx 150°$, $\psi \approx \pm 180°$ in only 1 of the 14 structures, it samples $\phi > 0°$ in 79% of the structures.

The fourth and fifth columns of Figure 1 present simulated Ramachandran maps for the nonglycosylated Ala6-Trp8 peptide and for the Ala6-Trp8 glycopeptide. With the exception of Ala6, which primarily samples right-handed helical conformations, the Ala6-Trp8 peptide demonstrates similar torsional preferences to the Gly6-Trp8 peptide. However, in contrast to the Gly6 peptides, glycosylation does not dramatically alter the torsional preferences of the Ala6-Trp8 peptide. In particular, glycosylation only minimally impacts the torsional preferences of Ala6.

The black X's in the last column indicate the Ramachandran angles sampled in the 47 glycosylated instances in the SAGS database of the Pro-Asn-Yyy-Thr/Ser sequon, where Yyy is any residue other than Gly. Pro, Asn, Yyy, and Thr/Ser adopt extended conformations in 57%, 55%, 60%, and 60% of these structures, respectively. In comparison to Pro-Asn-Gly-Thr/Ser sequons, the N-linked glycan appears to exert relatively little influence upon the torsional preferences of Pro-Asn-Yyy-Thr/Ser sequons, which is quite consistent with the Ala6-Trp8 peptide simulations.

In summary, Figure 1 demonstrates that glycosylation significantly altered the local torsional preferences of the Gly6 sequon and also introduced significant coupling between these residues. Interestingly, the Trp8→Ala8 mutation slightly enhanced the effects of glycosylation, while the Gly6→Ala6 mutation mitigated these effects. Furthermore, these simulated conformational tendencies are quite consistent with known glycoprotein structures.

Simulated Ensembles. Figure 2 characterizes the impact of glycosylation upon the simulated ensembles. The hinging angle, θ, which is defined by the α carbons of Ile2, Asn5, and Trp8/Ala8, characterizes the peptide compaction. The twisting angle, φ, which is defined by the (negative of the) pseudodihedral angle formed by the α carbons of Ile2, Pro4, Gly6/Ala6, and Trp8/Ala8, characterizes the peptide twist.

Figure 2A presents the simulated free-energy surface (FES) for the Gly6-Trp8 peptide as a function of θ and φ. The Gly6-Ala8 peptide samples a very similar FES (see the Supporting Information). Prior to glycosylation, the Gly6 peptides sample a diverse and highly disordered ensemble of fairly extended conformations with a slightly negative twist on average. The black X's indicate the two most stable regions of the FES. These shallow basins correspond to Asx-turns, which are defined by a hydrogen bond between the side chain carbonyl of Asn5 and the backbone amide of Thr7. While the Gly6-Trp8 peptide samples Asx-turns in 12% of conformations, it samples β-turns in less than 3% of conformations. For simplicity, we define β-turns for these short peptides by a single hydrogen bond[47] between the backbone carbonyl of Thr3 and the backbone amide of Gly6.

Figure 2B demonstrates that glycosylation dramatically impacts the conformational ensemble sampled by the Gly6-Trp8 glycopeptide. The glycan reverses the average peptide twist and destabilizes extended conformations such as Asx-turns, which are sampled in less than 4% of conformations. Instead, the glycan stabilizes compact conformations and, in

dx.doi.org/10.1021/jp508535f | J. Phys. Chem. B 2014, 118, 11462−11469

particular, β-turns, which are sampled in 27% of the Gly6-Trp8 glycopeptide conformations. Most significantly, the glycan stabilizes a native state (NS) of β-turns at the "eye" of the FES (indicated by the red box), which was rarely sampled prior to glycosylation. This NS is stabilized by peptide-glycan hydrogen bonds and also by a stacking interaction between the glycan and the Trp8 side chain that significantly reduces the hydrophobic solvent accessible surface area (SASA) for both groups (see the Supporting Information). This glycan-induced shift from Asx- to β-turns is highly consistent with the experiments of Imperiali and co-workers for a closely related peptide.[24−27]

Figure 2B reveals two distinct populations of i-to-i + 3 β-turns with a hydrogen bond between the Thr3 carbonyl and the Gly6 amide. In both populations, Thr3 samples extended conformations, while Pro4 and Asn5 sample helical conformations, as expected for a type I β-turn.[48] The conformation of the i + 3 residue (i.e., Gly6) distinguishes the two populations. In the dominant turn population, Gly6 samples left-handed conformations ($\phi > 0°$) and, in particular, the basin at $\phi \approx +150°$, $\psi \approx \pm 180°$. Interestingly, the glycan not only stabilized these conformations of Pro4, Asn5, and Gly6 but also introduced significant correlations among them. These β-turns, which we shall refer to as φ_+ β-turns, adopt a positive twist ($\varphi > 0°$) and appear in the NS at the eye of the FES. In the minor turn population, Gly6 samples right-handed conformations ($\phi < 0°$). These turns, which we shall refer to as ϕ_- β-turns, adopt a negative twist ($\varphi < 0°$) and appear in the shallow basin that is indicated by the dashed line.

Figure 2C presents the simulated FES for the Gly6-Ala8 glycopeptide. As suggested by Figure 1, glycosylation appears to exert very similar, and perhaps slightly greater, influence upon the Gly6-Ala8 FES. In comparison to the Gly6-Trp8 glycopeptide ensemble, the Gly6-Ala8 glycopeptide ensemble is more compact with a greater propensity for β-turns (31%) and conformations with $\varphi > 0°$. Moreover, the eye of φ_+ β-turns is slightly larger in the Gly6-Ala8 FES.

The enhanced effect of glycosylation upon the Gly6-Ala8 peptide is quite surprising, since the Trp8→Ala8 mutation abolishes the favorable aromatic−glycan stacking interactions which have been previously demonstrated to stabilize turns at enhanced aromatic sequons.[8−10] However, as noted above, the simulated peptide sequences are quite distinct from enhanced aromatic sequons, which have aromatic residues before the glycosylation site and lack Pro before the glycosylation site. Moreover, the Supporting Information reveals that the Ala8 side chain also interacts quite favorably with the disaccharide in the simulated NS. Because the Ala8 side chain does not interact with the glycan in non-native conformations, the NS provides a similar reduction in the glycan hydrophobic SASA for both Gly6 glycopeptides. The Supporting Information also demonstrates that, in comparison to the Gly6-Trp8 NS, the Gly6-Ala8 NS more effectively hydrogen bonds with the N-linked glycan.

Figure 2 (panels D and E) presents the simulated FES's for the Ala6-Trp8 peptide and for the Ala6-Trp8 glycopeptide, respectively. Prior to glycosylation, the Gly6-Trp8 and Ala6-Trp8 peptides sample fairly similar conformational ensembles, although the Gly6→Ala6 mutation stabilizes Asx-turns (30%). However, Figure 2E demonstrates that glycosylation exerts considerably less effect upon the Ala6-Trp8 peptide than upon the Gly6 peptides. Glycosylation only minimally impacts either the compaction or the twist of the ensemble. Glycosylation destabilizes Asx-turns (12%) and slightly stabilizes φ_- β-turns (8%), although Asx-turns remain one of the most stable conformations in the ensemble.

More importantly, Figure 2 (panels D and E) demonstrate that glycosylation does not significantly stabilize φ_+ β-turns (<1%) or any other particular conformation for the simulated Ala6 glycopeptides. The Supporting Information demonstrates that steric clashes with the Ala6 side chain preclude the Ala6 peptides from sampling φ_+ β-turns. When the Gly6 glyco/peptides adopt φ_+ β-turns, one proton of the Gly6 α carbon eclipses the backbone carbonyl oxygen of Asn5. In the Ala6 glyco/peptides this proton is replaced by the Ala6 side chain β carbon, which sterically clashes with the Asn5 backbone carbonyl when the peptide samples the φ_+ β-turn conformation. Since any other residue includes a β carbon, the simulations suggest that Gly may play an important role in stabilizing φ_+ β-turns at glycosylated Pro-Asn-Xxx-Thr sequons.

In summary, Figure 2 demonstrates that glycosylation triggers a dramatic conformational switch from extended Asx-turns to compact β-turns for both Gly6 peptides but exerts much less influence upon the Ala6 peptide. Thus, aromatic-glycan stacking interactions appear neither necessary nor sufficient for significantly stabilizing β-turns in these peptides. Instead, in the context of the Pro-Asn-Xxx-Thr sequon, the glycosylation code appears much more sensitive to the side chain β carbon of the residue that is adjacent to the glycosylated Asn. In particular, the simulations suggest that glycosylation may readily stabilize φ_+ β-turns at Pro-Asn-Xxx-Thr sequons when the Xxx residue is Gly but that the side chain β carbon of any other residue in this position may sterically destabilize φ_+ β turns.

Glycoprotein Structure Analysis. Of course, the REMD simulations are limited by the accuracy of the empirical classical force field. Moreover, it is not obvious that the ensembles sampled by short glycopeptides should provide meaningful insight into the structures adopted by glycosylated sequons in full length glycoproteins. Accordingly, Figure 3 compares the simulated ensembles with the structures of glycosylated sequons in the SAGS database of full length glycoproteins.[28,30] The black points in panels A and B of Figure 3 identify simulated conformations for the Gly6-Trp8 and Ala6-Trp8 glycopeptides, respectively, while the X's indicate corresponding glycosylated sequons in the SAGS database. We identified hydrogen bonds in the SAGS database by protonating the structures and allowing for a slightly relaxed acceptor−donor proton angle ($\theta_{ADH} \leq 40°$).

Figure 3A considers glycosylated sequons with Gly in position +1. The red X's indicate the 145 glycosylated Zzz-Asn-Gly-Thr/Ser sequons in the SAGS data set, where Zzz is any residue. These 145 sequons do not adopt any unique structure, although approximately 19% sample β-turns. The green X's indicate 14 glycosylated Pro-Asn-Gly-Thr/Ser sequons, which match the simulated Gly6 glycopeptides. The striking difference between the conformations indicated by the red and green X's suggests significant coupling between the conformations of Pro and Gly in these sequons, which was noted above in the context of Figure 1. As expected from the simulations, the 14 glycosylated Pro-Asn-Gly-Thr/Ser sequons tend to adopt compact structures with a positive twist. Moreover, and in remarkable agreement with the simulations, eight of these sequons, which are highlighted by blue circles, fold near the NS of the simulated FES and adopt φ_+ β-turns with the +1 Gly residue sampling left-handed conformations.

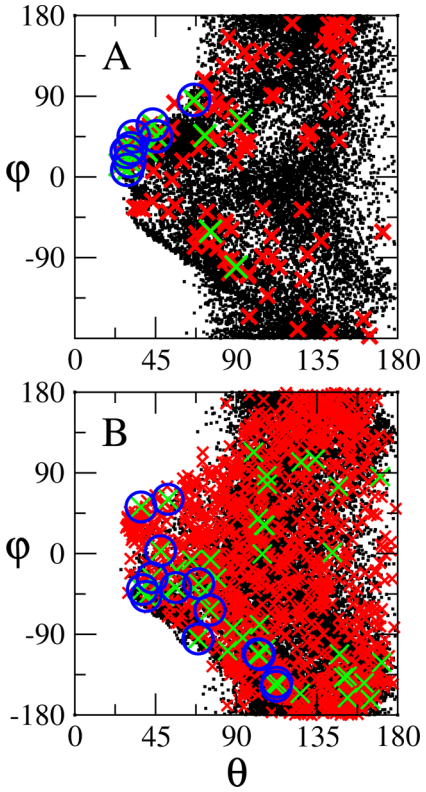


Figure 3. Conformations of glycosylated sequons. In panels A and B , the black points correspond to simulated conformations for the Gly6-Trp8 and Ala6-Trp8 glycopeptides, respectively, while the X's indicate conformations from the SAGS data set. (A) X's correspond to glycosylated Zzz-Asn-Gly-Thr/Ser sequons where Zzz is any residue (red) or Pro (green). (B) X's correspond to glycosylated Zzz-Asn-Yyy-Thr/Ser sequons, where Yyy is any residue other than Gly and Zzz is any residue (red) or Pro (green). In both panels, blue circles indicate SAGS structures that form β-turns and have Pro in position Zzz.

Figure 3B considers glycosylated sequons that do not have Gly in position +1. The red X's indicate the 1379 glycosylated Zzz-Asn-Yyy-Thr/Ser sequons in the SAGS data set, where Zzz is any residue and Yyy is any residue other than Gly. These sequons do not adopt any particular conformation and only 12% adopt β-turns. The green X's indicate 47 glycosylated Pro-Asn-Yyy-Thr/Ser sequons, which align with the simulated Ala6 glycopeptide. As observed in Figure 3A, the Pro residue appears to significantly impact the conformation of the glycosylated sequon. Only 15 of the glycosylated Pro-Asn-Yyy-Thr/Ser sequons adopt β-turns. As expected from the simulations, 12 of these 15 sequons form φ_- turns with the +1 Yyy residue sampling right-handed conformations. Thus, the SAGS database also suggests that the combination of Pro before and a side chain immediately after the glycosylation site may significantly destabilize φ_+ β-turns. Curiously, the two glycosylated Pro-Asn-Yyy-Thr/Ser sequons that form ϕ_+ β-turns near the center of the Gly6 NS region both have Cys in the +1 position. Moreover, these are the only two glycosylated Pro-Asn-Cys-Thr/Ser sequons in the SAGS data set and, in both cases, the Cys residue forms a disulfide bridge. Figure S7 of the Supporting Information provides further analysis of the glycosylated Pro-Asn-Yyy-Thr/Ser sequons in the SAGS database.

In summary, Figure 3 demonstrates remarkable consistency between the simulations of short glycopeptides and the structures adopted in full length glycoproteins. This consistency corroborates the accuracy of the simulation model and also its relevance for full length glycoproteins. Moreover, this agreement indicates a remarkable degree of consistency[49] between the protein−protein and protein−glycan interactions of natural glycoproteins. This, in turn, suggests that natural glycoproteins may be minimally frustrated to simultaneously optimize not only intraprotein interactions but also the relevant glycan−protein interactions.[50]

Glycosylated β-Turns. Finally, Figure 4 analyzes the 188 glycosylated β-turns for which the SAGS data set provides structural information for the 10 residues immediately before and after the glycosylation site. We label the residues such that 0 indicates the glycosylated Asn, −10··· −1 indicate the 10 residues immediately preceding the glycosylated Asn, and +1··· +10 indicate the 10 residues immediately following the glycosylated Asn. As above, we identified these β-turns based upon a backbone hydrogen bond between the residues in position −2 and +1 (i.e., the residues that align with Thr3 and Gly/Ala6). Panels A and B characterize the 56 sequons that adopt φ_+ β-turns (i.e., $\varphi > 0°$), while panels C and D characterize the remaining 132 sequons that adopt φ_- β-turns (i.e., $\varphi < 0°$). In panels A and C, the x axis indicates the β-turns, while the y axis indicates the residues within each β-turn. The colors indicate the secondary structures that are predicted by DSSP[51,52] for each residue. The transparent images in panels B and D trace the backbone structures for the φ_+ and φ_- β-turns, respectively, while the solid images indicate the corresponding average structures.

Figure 4 demonstrates that the 56 glycosylated φ_+ β-turns tend to form compact β-hairpins between two sheets. A diverse set of sequons fold to φ_+ β-turns and, in particular, only 22 of the 56 sequons contain Gly at position +1. However, in all 22 of these sequences, the +1 Gly residue adopts left-handed helical conformations, which is quite consistent with the glycopeptide simulations. In fact, the +1 residue samples a left-handed conformation in 71% of the 56 φ_+ β-turns. Moreover, either the +1 or −1 residue adopts a left-handed conformation in 96% of these φ_+ β-turns. Conversely, Figure 4 (panels C and D) demonstrates that φ_- β-turns occur in wide-range structures, including many disordered coils but also a few fairly long α helices. In contrast to φ_+ β-turns, the +1 residue samples a left-handed helical conformation in only 2% of the φ_- β-turns. Thus, left-handed helical conformations, especially at position +1, appear to significantly promote the formation of compact β-hairpins at glycosylation sites.

CONCLUSIONS

In conclusion, this study investigates the impact of N-linked glycosylation upon sequons of the form Pro-Asn-Xxx-Thr-Yyy. The REMD simulations of short glycopeptides suggest that, at least for this sequon, the impact of glycosylation is quite sensitive to the Xxx residue in position +1. The simulations demonstrate that, when Gly is in this position, glycosylation drives a conformational switch from extended conformations to compact β-turns, as previously observed by Imperiali and coworkers.[24−27] Moreover, the simulations suggest that the side chain β carbon of any other residue in this position may sterically destabilize compact β-turns and may quite generally mitigate the effects of glycosylation upon sequons of this form. Furthermore, aromatic−glycan stacking interactions are neither necessary nor sufficient for stabilizing β-turns in the simulated glycopeptides. Thus, at least for the present class of sequons,

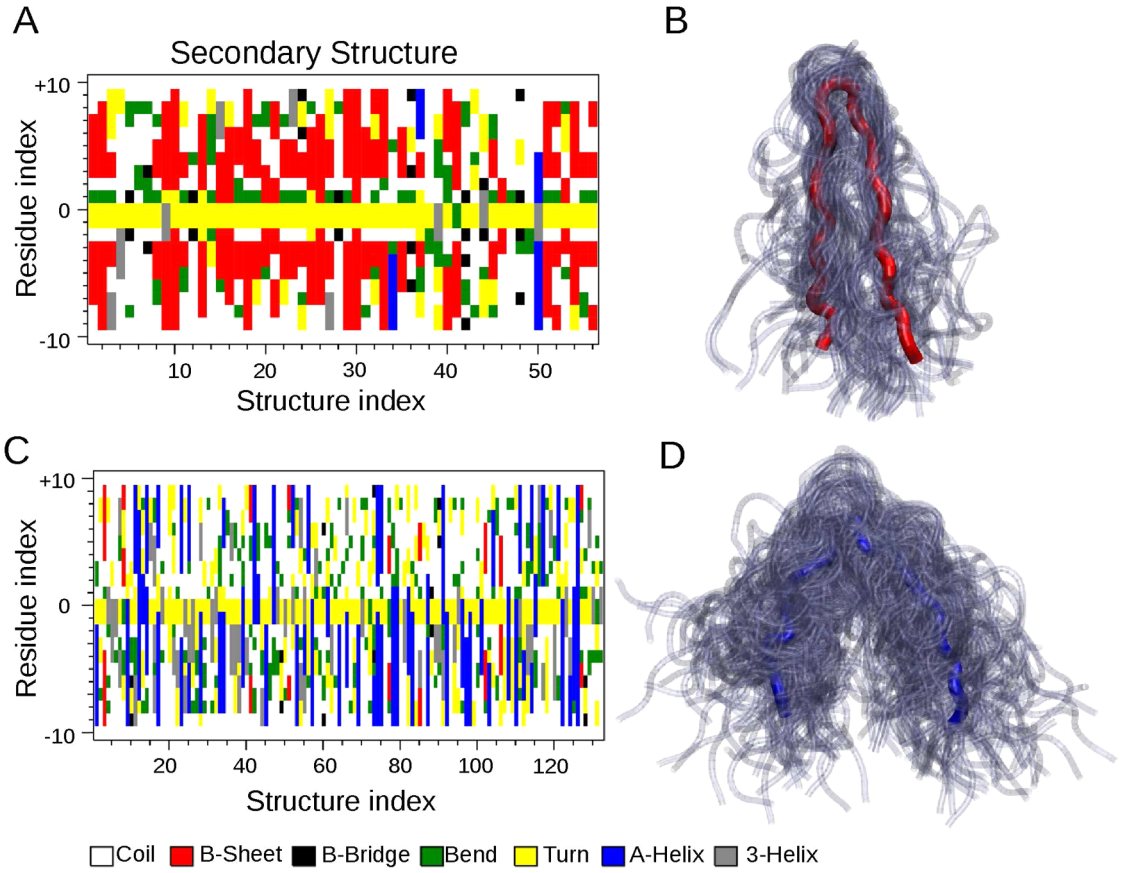


Figure 4. Structural context of 188 glycosylated β-turns from the SAGS glycoprotein database. (A and C) characterize the secondary structures surrounding φ_+ and φ_- β-turns, respectively. (B and D) Transparent images trace the backbone of these structures, while the solid red and blue images trace the average structures.

the glycosylation code appears more sensitive to the side chain β carbon at position +1 than to an aromatic side chain in position +3. The REMD simulations of short glycopeptides appear remarkably predictive of the structures observed in the SAGS database of glycoproteins. This not only corroborates the simulations but also suggests a remarkable degree of consistency between the intraprotein and protein−glycan interactions of natural glycoproteins. Significantly, the simulations and bioinformatic analysis both indicate significant coupling between the Pro and Gly residues that flank the glycosylated Asn. Thus, these conclusions may be limited to sequons of the form Pro-Asn-Xxx-Thr/Ser-Yyy, which clearly motivates future studies that investigate the importance of the residue preceding the glycosylation site. Finally, the computational studies suggest that left-handed conformations play a critical role for inducing compact β-hairpins at glycosylated sequons, which may provide a new design principle for engineering novel structural motifs into glycoproteins.

■ ASSOCIATED CONTENT

Supporting Information

Additional information and analysis for the reported simulations regarding the simulated force field, equilibration procedures, and observed interactions. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: wnoid@chem.psu.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has been financially supported by an NSF CAREER award (Grant MCB 1053970) from the National Science Foundation. The authors gratefully acknowledge Profs. Joshua L. Price and Scott A. Showalter for very helpful comments and the Petrescu lab for access to the SAGS database. Numerical calculations used support and resources from Research Computing and Cyberinfrastructure, a unit of Information Technology Services at Penn State. Figures 2 and 4 were made with VMD. VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana−Champaign.

■ REFERENCES

(1) Taylor, M. E.; Drickamer, K. *Introduction to GlycoBiology*; Oxford Press: New York, NY, 1987.

(2) Apweiler, R.; Hermjakob, H.; Sharon, N. On the Frequency of Protein Glycosylation, as Deduced from Analysis of the SWISS-PROT Database. *Biochim. Biophys. Act.* **1999**, *1473*, 4−8.

(3) Helenius, A.; Aebi, M. Intracellular Functions of N-Linked Glycans. *Science* **2001**, *291*, 2364−2369.

(4) Molinari, M. N-Glycan Structure Dictates Extension of Protein Folding or Onset of Disposal. *Nat. Chem. Biol.* **2007**, *3*, 313−320.
(5) Rademacher, T.; Parekh, R.; Dwek, R. Glycobiology. *Annu. Rev. Biochem.* **1988**, *57*, 785−838.
(6) Imperiali, B.; O'Connor, S. Effect of N-Linked Glycosylation on Glycopeptide and Glycoprotein Structure. *Curr. Opin. Chem. Biol.* **1999**, *3*, 643−649.
(7) Shental-Bechor, D.; Levy, Y. Folding of Glycoproteins: Toward Understanding the Biophysics of the Glycosylation Code. *Curr. Opin. Struc. Biol.* **2009**, *19*, 524−533.
(8) Culyba, E. K.; Price, J. L.; Hanson, S. R.; Dhar, A.; Wong, C.-H.; Gruebele, M.; Powers, E. T.; Kelly, J. W. Protein Native-State Stabilization by Placing Aromatic Side Chains in N-Glycosylated Reverse Turns. *Science* **2011**, *331*, 571−575.
(9) Price, J. L.; Powers, D. L.; Powers, E. T.; Kelly, J. W. Glycosylation of the Enhanced Aromatic Sequon is Similarly Stabilizing in Three Distinct Reverse Turn Contexts. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 14127−14132.
(10) Price, J. L.; Culyba, E. K.; Chen, W.; Murray, A. N.; Hanson, S. R.; Wong, C.-H.; Powers, E. T.; Kelly, J. W. N-Glycosylation of Enhanced Aromatic Sequons to Increase Glycoprotein Stability. *Biopolymers* **2012**, *98*, 195−211.
(11) Sola, R. J.; Griebenow, K. Effects of Glycosylation on the Stability of Protein Pharmaceuticals. *J. Pharm. Sci.* **2009**, *98*, 1223−1245.
(12) Hoffmann, D.; Florke, H. A Structural Role for Glycosylation: Lessons from the hp Model. *Folding Des.* **1998**, *3*, 337−343.
(13) Shental-Bechor, D.; Levy, Y. Effect of Glycosylation on Protein Folding: A Close Look at Thermodynamic Stabilization. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8256−8261.
(14) Price, J. L.; Shental-Bechor, D.; Dhar, A.; Turner, M. J.; Powers, E. T.; Gruebele, M.; Levy, Y.; Kelly, J. W. Context-Dependent Effects of Asparagine Glycosylation on Pin WW Folding Kinetics and Thermodynamics. *J. Am. Chem. Soc.* **2010**, *132*, 15359−15367.
(15) Chen, M. M.; Bartlett, A. I.; Nerenberg, P. S.; Friel, C. T.; Hackenberger, C. P. R.; Stultz, C. M.; Radford, S. E.; Imperiali, B. Perturbing the Folding Energy Landscape of the Bacterial Immunity Protein Im7 by Site-Specific N-Linked Glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 22528−22533.
(16) Ellis, C. R.; Maiti, B.; Noid, W. G. Specific and Nonspecific Effects of Glycosylation. *J. Am. Chem. Soc.* **2012**, *134*, 8184−8193.
(17) Ellis, C. R.; Maiti, B.; Noid, W. G. Addition to "Specific and Nonspecific Effects of Glycosylation. *J. Am. Chem. Soc.* **2014**, *136*, 8485−8485.
(18) Pandey, B. K.; Enck, S.; Price, J. L. Stabilizing Impact of N-Glycosylation on the WW Domain Depends Strongly on the Asn-GlcNAc Linkage. *ACS Chem. Biol.* **2013**, *8*, 2140−2144.
(19) Jitsuhara, Y.; Toyoda, T.; Itai, T.; Yamaguchi, H. Chaperone-like Functions of High-Mannose Type and Complex-Type N-Glycans and Their Molecular Basis. *J. Biochem.* **2002**, *132*, 803−811.
(20) Wyss, D.; Choi, J.; Li, J.; Knoppers, M.; Willis, K.; Arulanandam, A.; Smolyar, A.; Reinherz, E.; Wagner, G. Conformation and Function of the N-Linked Glycan in the Adhesion Domain of Human CD2. *Science* **1995**, *269*, 1273−1278.
(21) Luis Asensio, J.; Arda, A.; Javier Canada, F.; Jimenez-Barbero, J. Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46*, 946−954.
(22) Hanson, S. R.; Culyba, E. K.; Hsu, T.-L.; Wong, C.-H.; Kelly, J. W.; Powers, E. T. The Core Trisaccharide of an N-Linked Glycoprotein Intrinsically Accelerates Folding and Enhances Stability. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3131−3136.
(23) O'Connor, S.; Pohlmann, J.; Imperiali, B.; Saskiawan, I.; Yamamoto, K. Probing the Effect of the Outer Saccharide Residues of N-Linked Glycans on Peptide Conformation. *J. Am. Chem. Soc.* **2001**, *123*, 6187−6188.
(24) Imperiali, B.; Rickert, K. Conformational Implications of Asparagine-Linked Glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 97−101.
(25) O'Connor, S.; Imperiali, B. Conformational Switching by Asparagine-Linked Glycosylation. *J. Am. Chem. Soc.* **1997**, *119*, 2295−2296.
(26) O'Connor, S.; Imperiali, B. A Molecular Basis for Glycosylation-Induced Conformational Switching. *Chem. Biol.* **1998**, *5*, 427−437.
(27) Bosques, C.; Tschampel, S.; Woods, R.; Imperiali, B. Effects of Glycosylation on Peptide Conformation: A Synergistic Experimental and Computational Study. *J. Am. Chem. Soc.* **2004**, *126*, 8421−8425.
(28) Petrescu, A.; Milac, A.; Petrescu, S.; Dwek, R.; Wormald, M. Statistical Analysis of the Protein Environment of N-Glycosylation Sites: Implications for Occupancy, Structure, and Folding. *Glycobiology* **2004**, *14*, 103−114.
(29) Petrescu, A.-J.; Wormald, M. R.; Dwek, R. A. Structural Aspects of Glycomes with a Focus on N-Glycosylation and Glycoprotein Folding. *Curr. Opin. Struct. Biol.* **2006**, *16*, 600−607.
(30) *Glycosylation*; Petrescu, S., Ed.; InTech: Rijeka, Croatia, 2012.
(31) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(32) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(33) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(34) Schuttelkopf, A.; van Aalten, D. PRODRG: A Tool for High-throughput Crystallography of Protein-Ligand Complexes. *Acta Crystallogr.* **2004**, *60*, 1355−1363.
(35) Schrödinger, L. L. C. *The PyMOL Molecular Graphics System*, version 1.5; Schrödinger, LLC: New York, 2010.
(36) Berendsen, H.; Grigera, J.; Straatsma, T. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(37) Jorgensen, W.; Maxwell, D.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(38) Kony, D.; Damm, W.; Stoll, S.; van Gunsteren, W. An Improved OPLS-AA Force Field for Carbohydrates. *J. Comput. Chem.* **2002**, *23*, 1416−1429.
(39) Allen, M. P.; Tildesley, D. P. *Computer Simulation of Liquids*; Oxford Press: New York, NY, 1987.
(40) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(41) Parrinello, M.; Rahman, A. Strain Fluctuations and Elastic-constants. *J. Chem. Phys.* **1982**, *76*, 2662−2666.
(42) York, D.; Darden, T.; Pedersen, L. The Effect of Long-Range Electrostatic Interactions in Simulations of Macromolecular Crystals: A Comparison of the Ewald and Truncated List Methods. *J. Chem. Phys.* **1993**, *99*, 8345−8348.
(43) Patriksson, A.; van der Spoel, D. A Temperature Predictor for Parallel Tempering Simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073−2077.
(44) Sindhikara, D. J.; Emerson, D. J.; Roitberg, A. E. Exchange Often and Properly in Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2010**, *6*, 2804−2808.
(45) Becker, D. J.; Lowe, J. B. Fucose: Biosynthesis and Biological Function in Mammals. *Glycobiology* **2003**, *13*, 41R−53R.
(46) Ma, B.; Simala-Grant, J. L.; Taylor, D. E. Fucosylation in Prokaryotes and Eukaryotes. *Glycobiology* **2006**, *16*, 158R−184R.
(47) Luzar, A.; Chandler, D. Structure and Hydrogen-Bond Dynamics of Water-Dimethyl Sulfoxide Mixtures by Computer-Simulations. *J. Chem. Phys.* **1993**, *98*, 8160−8173.
(48) Sibanda, B.; Blundell, T.; Thornton, J. 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.* **1989**, *206*, 759−777.
(49) Go, N. Theoretical Studies of Protein Folding. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183−210.

(50) Bryngelson, J.; Wolynes, P. Spin-Glasses and the Statistical-Mechanics of Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524−7528.
(51) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577−2637.
(52) Joosten, R. P.; Beek, T. A. H. T.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W. W.; Schneider, R.; Sander, C.; Vriend, G. A Series of PDB Related Databases for Everyday Needs. *Nucleic Acids Res.* **2011**, *39*, D411−D419.