HHS Public Access

Author manuscript

J Am Chem Soc. Author manuscript; available in PMC 2017 June 22.

Published in final edited form as:

J Am Chem Soc. 2016 June 22; 138(24): 7636–7648. doi:10.1021/jacs.6b02879.

The Dependence of Carbohydrate–Aromatic Interaction Strengths on the Structure of the Carbohydrate

Che-Hsiung Hsu[†,‡], **Sangho Park**[§], **David E. Mortenson**[†], **B. Lachele Foley**[#], **Xiaocong Wang**[#], **Robert J. Woods**[#], **David A. Case**[‖], **Evan T. Powers**[‡,*], **Chi-Huey Wong**[‡,=,⊥,*], **H. Jane Dyson**[§,*], and **Jeffery W. Kelly**[†,‡,⊥,*]

[†]Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA

[‡]Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037, USA

[§]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

[#]Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, Georgia 30602, USA

[‖]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854

[=]Genomics Research Center, Academia Sinica, Taipei 115, Taiwan

[⊥]The Skaggs Institute for Chemical Biology

Abstract

Interactions between proteins and carbohydrates are ubiquitous in biology. Therefore, understanding the factors that determine their affinity and selectivity are correspondingly important. Herein, we have determined the relative strengths of intramolecular interactions between a series of monosaccharides and an aromatic ring close to the glycosylation site in an N-glycoprotein host. We employed the enhanced aromatic sequon, a structural motif found in the reverse turns of some N-glycoproteins, to facilitate face-to-face monosaccharide–aromatic interactions. A protein host was used because the dependence of the folding energetics on the identity of the monosaccharide can be accurately measured to assess the strength of the carbohydrate–aromatic interaction. Our data demonstrate that the carbohydrate–aromatic interaction strengths are moderately affected by changes in the stereochemistry and identity of the

[*]**Corresponding Author** J.W.K.: jkelly@scripps.edu; H.J.D.: dyson@scripps.edu; C.-H.W.: chwong@gate.sinica.edu.tw; E.T.P.: epowers@scripps.edu.

Present Address C.-H.H.: Adheren Biologics, Inc. 5858 Horton Street, Suite 255, Emeryville, CA 94608

Supporting information Detailed thermodynamics data from fitting CD-monitored thermal denaturation data of Pin WW glycovariants (Table S1); summary of carbohydrate ring chemical shift perturbations comparing carbohydrates attached to Asn_{19} in enhanced aromatic sequons in Pin WW to the same carbohydrate attached to free asparagine (Table S2); statistics from NMR structure determinations (Table S3); ESI-MS data for all Pin WW glycovariants (Table S4); additional information regarding the molecular dynamics simulations; experimental procedures and characterization for Fmoc-protected asparagine derivatives used in the solid phase peptide synthesis of Pin WW glycovariants. All structures reported herein are available in the RCSB protein data bank with the following PDB IDs: galactose (Gal) Pin WW glycovariant **5**, PDB ID: 2NC4; allose (All) Pin WW glycovariant **7**, PDB ID: 2NC3; L-idose (L-Ido) Pin WW glycovariant **8**, PDB ID: 2NC6; xylose Pin WW glycovariant **9**, PDB ID: 2NC5.

The authors declare no competing financial interest.

substituents on the pyranose rings of the sugars. Galactose seems to make the weakest and allose the strongest sugar–aromatic interactions, with glucose, N-acetylglucosamine (GlcNAc) and mannose in between. The NMR solution structures of several of the monosaccharide-containing N-glycoproteins were solved to further understand the origins of the similarities and differences between the monosaccharide–aromatic interaction energies. Peracetylation of the monosaccharides substantially increases the strength of the sugar–aromatic interaction in the context of our N-glycoprotein host. Finally, we discuss our results in light of recent literature regarding the contribution of electrostatics to CH–π interactions and speculate on what our observations imply about the absolute conservation of GlcNAc as the monosaccharide through which N-linked glycans are attached to glycoproteins in eukaryotes.

INTRODUCTION

N-linked glycosylation is a major co- or post-translational protein modification. Protein N-glycosylation occurs in all three domains of life—bacteria, archaea, and eukarya—and is mediated by the enzyme oligosaccharyltransferase (OST).[–] In eukaryotes, OST catalyzes the *en bloc* transfer of the preassembled, lipid linked oligosaccharide $Glc_3Man_9GlcNAc_2$– onto the side chain amide nitrogen of an asparagine (Asn) residue in the acceptor sequence (or "sequon") Asn-Xxx-Ser/Thr (where Xxx can be any amino acid except Pro) harbored within polypeptide chains that are inserted into the endoplasmic reticulum (ER).[,] N-glycosylation influences both the folding and function of glycoproteins. Protein N-glycosylation enables key interactions with the lectin-type chaperones of the ER glycoproteostasis network, which influence whether the N-glycoprotein will be folded and trafficked or degraded.[–] N-Glycosylation can also intrinsically alter the folding kinetics and thermodynamic stability of N-glycoproteins, especially via native state intramolecular protein–carbohydrate interactions. For example, in human CD2, protein–glycan interactions stabilize the native state by −3.1 kcal mol^{-1} relative to the unfolded state.[,] In CD2, the protein–carbohydrate interactions occur within an "enhanced aromatic sequon".[–] Enhanced aromatic sequons are structural modules having the sequence Aro-$(Xxx)_n$-Asn(R_1)-Yyy-Thr; where Aro is an aromatic amino acid; n = 1 or 2 (n can be 0 as well, but the stabilization is greatly diminished); $(Xxx)_n$ is any single amino acid or pair of amino acids; Yyy can be any amino acid other than Pro, but is often Gly; and R_1 is the carbohydrate attached via N-glycosylation to the Asn side chain. When enhanced aromatic sequons adopt certain conformations (e.g., a type 1 β-bulge turn if n = 1 or a type 2 β-turn within a six-residue loop if n = 2) an interaction between the Aro side chain and R_1 is enforced (Figure 1). This contact is typically a face-to-face interaction between the α-face of the sugar residue directly attached to Asn (GlcNAc-1) and the aromatic side chain of Aro that is driven by the hydrophobic effect and by CH–π interactions.

Enhanced aromatic sequons exist in a number of N-glycoproteins. In addition to their ability to stabilize protein native states, they also improve N-glycosylation efficiency by OST and increase the homogeneity of N-glycan structures, which is accomplished by restricting N-glycan processing in the Golgi. Herein, we use the enhanced aromatic sequon in the context of a type 1 β-bulge turn to reliably position the sugar-of-interest relative to the aromatic side chain in order to study the energetics of sugar–aromatic interactions. In a previous study, we

transplanted five- and six-residue enhanced aromatic sequons into loop 1 of the 34-residue WW domain from human Pin1, referred to hereafter as “Pin WW”, to create a simple GlcNAcylated protein. Pin WW, consisting of three β strands connected by two loops (Figure 2), folds reliably and has been used extensively to study the energetics of protein folding.[9] Recently, we reported a study wherein we varied Aro in order to parse the contributions from the hydrophobic effect and CH–π interactions to the net energy of the GlcNAc-1–aromatic interaction. We found that the hydrophobic effect and CH–π interactions both contributed substantially to protein–carbohydrate interactions. We also found that the net energy of the CH–π component of the interaction between GlcNAc and aromatic side chains (that is, the energy change upon going from the unfolded state in which the GlcNAc CH groups and the aromatic protein side chain are solvated to the folded state in which they are interacting with each other) was dominated by dispersion forces, with at most a minor contribution from electrostatics.

Here we examine the energetics of protein–carbohydrate interactions from the point of view of the carbohydrate; that is, by varying the structure of the sugar attached to the Asn side chain amide nitrogen in the context of a five-residue enhanced aromatic sequon adopting a type 1 β-bulge reverse turn within Pin WW, while keeping Aro constant as Phe. In addition, we examined the influence of peracetylation of the sugars, as this has been shown to strengthen protein–carbohydrate interactions.[9] We also explored the energetic consequences of replacing the sugars with purely hydrophobic cycloalkanes. We find that the stabilization free energy derived from sugar–aromatic interactions is moderately affected by changes in the stereochemistry of the hydroxyl groups on the sugar (for example, glucose (Glc) vs. galactose (Gal) or mannose (Man) or allose (All)) or by hydroxyl group deletion (for example, glucose vs. cyclohexane). In addition, peracetylation of the sugars greatly and generally increases the strength of the protein–carbohydrate interactions, in agreement with previous findings.[9] We analyze the measured interaction energies in the context of NMR-derived solution structures of several of the Pin WW glycovariants, and discuss the potential for carbohydrate–aromatic interactions to contribute to the affinity and selectivity of carbohydrate recognition. Finally, we revisit the driving forces for CH–π interactions and speculate on the evolutionary selection pressures that may have contributed to GlcNAc being absolutely conserved as the residue by which N-glycans are attached to proteins in eukaryotes.[9]

RESULTS

Pin WW domains with five-residue enhanced aromatic sequons are stabilized by carbohydrate–aromatic interactions

The folding kinetics and thermodynamics of Pin WW domains have been extensively characterized.[10–15] Its three-stranded antiparallel β-sheet structure (Figure 2)[10,15] folds by a two state mechanism wherein loop 1 formation is generally rate-limiting.[10] The five-residue enhanced aromatic sequon was incorporated into Pin WW by converting the native sequence of loop 1 (Ser_{16}-Arg_{17}-Ser_{18}-Ser_{19}-Gly_{20}-Arg_{21}) to Phe_{16}-Ala_{18}-$Asn_{19}(R_1)$-Gly_{20}-Thr_{21}; note that there is no residue 17 in the latter sequence, and that R_1 is the N-glycan attached to the side chain amide nitrogen atom of Asn_{19}. We previously solved the solution structure of

Pin WW harboring the five-residue enhanced aromatic sequon by NMR spectroscopy, both in its unglycosylated form (**1**; PDB ID: 2M9E, Figure 2A) and with R_1 = GlcNAc (**2**; PDB ID: 2M9F, Figure 2B). In both cases, loop 1 adopted a type 1 β-bulge reverse turn conformation. In variant **2**, this conformation enforces the face-to-face interaction between GlcNAc and Phe that is the signature of the enhanced aromatic sequon (Figures 1 and 2B).

Chemical synthesis of Pin WW glycovariants

The sugar substituents that were attached to the side chain amide nitrogen of Asn_{19} in Pin WW are shown in Chart 1. All of these Pin WW glycovariants were prepared by solid phase peptide synthesis (SPPS) employing an Fmoc chemistry strategy. The synthesis of Fmoc-$(OAc)_4$allose-Asn, depicted in Scheme 1, is representative of the preparation of the protected monosaccharide-Asn conjugates used for SPPS. Allosyl amine **16** with a β-configuration at the C1 or anomeric carbon was afforded by reduction of the corresponding β-glycosyl azide **15**, followed by the subsequent coupling with Fmoc-Asp-OtBu using benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) for dehydration. This previously unexplored approach for attaching 1-amino sugars to Fmoc-Asp-OtBu was efficient (60 to 90% yields). After acid-mediated removal of the *t*-butyl group, the Fmoc- and acetyl-protected glycosyl-asparagine amino acids (for example, **18**) were incorporated into Pin WW sequences using a previously published Fmoc-based solid phase synthesis of Pin WW glycovariants.[9] This strategy, exemplified in Scheme 1, was employed to make the 15 monosaccharide-containing Pin WW variants shown in Chart 1. In addition, Pin WW variants with cycloalkyl substituents attached to the side chain amide nitrogen of Asn_{19} (**10**, **11**, **12**, and **13**) were prepared by a similar procedure. The necessary Fmoc-protected asparagine derivatives were prepared by coupling the desired cycloalkyl amines to the unprotected side chain carboxylic acid of Fmoc-Asp(OH)-OtBu, followed by deprotection of the *t*-butyl group and incorporation of the protected amino acid residues into Pin WW by Fmoc SPPS.

Sugar stereochemistry and structure influences the folding energetics of the Pin WW glycovariants

Far-UV circular dichroism (CD) spectroscopy-monitored thermal denaturation curves were recorded and processed as described previously to extract the folding free energies of the Pin WW glycovariants summarized in Figure 3 (detailed thermodynamic data are reported in Table S1).[10] The midpoint temperature for thermal denaturation (T_m) of unglycosylated Pin WW (variant **1**) was found to be 66.1 ± 0.2 °C. Modification of Asn_{19} by attaching a GlcNAc to its side chain amide nitrogen increased the T_m to 75.2 ± 0.2 °C. This change in T_m (ΔT_m) of 9.1 ± 0.3 °C indicates that GlcNAcylation of Asn_{19} stabilizes variant **2** by a free energy of folding difference (ΔG_{glyc}) of −1.03 ± 0.07 kcal mol^{-1} at 60 °C relative to **1** (Table S1; all data are reported as the mean ± standard error unless otherwise noted). These data are in excellent agreement with previously reported values for these variants.[9]

Examining the folding energetics of Pin WW glycovariants **2** through **9**, which all have fully deprotected monosaccharides attached to the side chain amide nitrogen of Asn_{19}, reveals that N-glycosylation of Asn_{19} with virtually any monosaccharide stabilizes the native state of Pin WW (Figure 3, Table S1). Note that in Figure 3 and our discussion of the data

presented therein we use changes in T_m as a surrogate parameter for the changes in folding energetics because it can be measured quite accurately and because changes in T_m faithfully reflect folding free energies. In particular, in Pin WW variants a difference of 1 °C in T_m corresponds to a difference of about −0.11 kcal mol^{-1} in ΔG_f at 60 °C (see Table S1). In addition, the values of ΔG_f at 60 °C for all Pin WW glycovariants are reported in Table S1.

The data in Figure 3 and Table S1 demonstrate that the stereochemistry and structure of the sugar ring can affect the degree to which it stabilizes Pin WW, though the effect is modest. For example, replacing the C2 hydroxyl group with an acetamide (see Chart 1 for monosaccharide ring numbering scheme) slightly increases the native state stabilization due to glycosylation; compare the T_m values of variants **2** and **4** (R_1 = GlcNAc and glucose) and those of **3** and **5** (R_1 = N-acetylgalactosamine (GalNAc) and galactose). This effect is statistically significant only for the latter pair (for **2** *vs.* **4** $\Delta T_m = 0.7 \pm 0.3$ °C, $p = 0.06$; for **3** *vs.* **5** $\Delta T_m = 1.8 \pm 0.3$ °C, $p = 0.005$; p-values are for two-sided Student's T-tests), though **3** is much less stable than **2**. Changing the stereochemistry of the functional groups that project from the pyranose ring from their disposition in glucose (in which they are all equatorial) has variable effects on the native state stabilization due to glycosylation. Epimerization at C4 consistently diminishes the stability of Pin WW glycovariants, as seen from the difference in the folding energetics of glycovariants **2** *vs.* **3** (R_1 = GlcNAc *vs.* GalNAc, $\Delta T_m = -1.3 \pm 0.3$ °C) and glycovariants **4** *vs.* **5** ($\Delta T_m = -2.2 \pm 0.3$ °C). Epimerization at either C2 or C3 increases the stability of Pin WW glycovariants, as seen by comparing the differences in folding energetics of glycovariants **4** and **6**, which differ by epimerization at C2 (R_1 = glucose *vs.* mannose, $\Delta T_m = 1.6 \pm 0.3$), and between glycovariants **4** and **7**, which differ by epimerization at C3 (R_1 = glucose *vs.* allose, $\Delta T_m = 2.7 \pm 0.3$).

Based on the previously reported structure of glycovariant **2**, we anticipated that epimerizing C5 would lead to a steric clash between the hydroxymethyl substituent on C5 and the phenyl ring of Phe_{16}. As expected, epimerization at C5 greatly diminishes the stability of Pin WW, as seen by comparing glycovariants **4** *vs.* **8** (Figure 3; R_1 = glucose *vs.* L-idose, $\Delta T_m = -5.4 \pm 0.3$ °C). However, it does not completely eliminate the stabilization due to glycosylation with L-idose, as the T_m of variant **8** is 3.0 ± 0.3 °C higher than that for non-glycosylated Pin WW **1**. Surprisingly, deletion of the exocyclic hydroxymethyl group, affording Pin WW variant **9** in which R_1 = xylose, has only a small effect on the glycosylation-mediated stabilization of the native state relative to glucose ($R_1 = \Delta T_m = 0.7 \pm 0.2$ °C; *cf.* variants **4** *vs.* **9**, Figure 3). Thus, the negative effect caused by changing the NHAc group to OH, as shown between variants **2** and **4**, is compensated by the removal of the hydroxymethyl group from **4**. It is, however, noteworthy that a gain of stabilization was observed by replacing the equatorial NHAc group of variant **2** with the axial OH group in variant **6** although the basis for this is unclear. We next explored the effects on Pin WW stability of eliminating the free hydroxyl groups in the monosaccharides either by peracetylation or by deletion, *i.e.*, replacing the monosaccharide substructure with a cycloalkane substructure.

Peracetylation enhances protein–carbohydrate interactions in Pin WW glycovariants

The peracetylated version of each Pin WW glycovariant is actually that glycovariant's synthetic precursor (see Scheme 1, *e.g.*, **7-Ac**), and is therefore very conveniently obtained. The only peracetylated variant that could not be obtained in pure form from its synthetic precursor was **8**, in which R_1 = L-idose. In this case the fully peracetylated version could not be separated from partially deacetylated contaminants. Far-UV CD monitored thermal denaturation experiments were carried out as described above to characterize the folding thermodynamics of each peracetylated Pin WW glycovariant (variants **2-Ac** through **9-Ac**; Figure 3, Table S1).

In agreement with previously reported findings that peracetylation of carbohydrates strengthens carbohydrate–aromatic interactions,[9] we found that peracetylation of the hydroxyl groups on the sugar substantially increased the thermodynamic stability of all of the monosaccharide-containing Pin WW glycovariants. The difference in T_m values between the normal and peracetylated Pin WW glycovariants ranged from 1.6 ± 0.2 °C for the mannose-based variants (**6** vs. **6-Ac**) to 6.8 ± 0.4 °C for the GlcNAc-based variants (**2** vs. **2-Ac**), though the trend is different from that of the nonacetylated glycovariants. A likely explanation for this stabilization, as proposed by Laughrey and co-workers, is that acetylation of a sugar's hydroxyl groups increases sugar hydrophobicity, making it more favorable for the sugar to transfer from being solvated to being packed against an aromatic ring. However, there are alternative explanations that cannot be ruled out based on our data. For example, peracetylation could increase the polarization of the C–H bonds, leading to stronger CH–π interactions. In addition, the acetyl groups could themselves interact with the aromatic ring in the enhanced aromatic sequon or with other side chains or the main chain via hydrophobic interactions (with the acetyl methyl group) or hydrogen bonding (with the ester carbonyl oxygen). The latter possibility is consistent with the variability of the effect of peracetlyation as a function of the stereochemistry and orientation of individual acetylated hydroxyls.

Complete deletion of the pyranose hydroxyl groups: the hydrophobic effect in carbohydrate–aromatic interactions

To further understand the forces involved in the carbohydrate–aromatic interaction in enhanced aromatic sequons, we synthesized Pin WW variant **10** in which a cyclohexyl group rather than a monosaccharide is attached to the side chain amide nitrogen of Asn_{19}. Pin WW variants **4** and **10** (R_1 = glucose and cyclohexyl, respectively) have similar T_m values (for **4**, T_m = 74.5 ± 0.2; for **10**, T_m = 75.2 ± 0.2; ΔT_m = 0.7 ± 0.3; Figure 3, Table S1).

It is perhaps surprising that two R_1 groups as different as cyclohexane and glucose should interact with the Phe_{16} side chain with similar energetics. However, upon careful examination it is in fact reasonable to expect that the two main energetic components of this interaction—the CH–π interaction and the hydrophobic effect—will be similar between these two moieties. First, while it is true that the C–H bonds in carbohydrates are more polarized than those in cyclohexane, this difference in polarization is not sufficient to result in a substantive difference in the strength of the CH–π interaction with the Phe side chain. This statement is based on our previous observation that electrostatics make at most a small

contribution to the strength of the CH–π interaction in the enhanced aromatic sequon, as well as computational/experimental studies demonstrating that the electrostatic energies of benzene–carbohydrate (fucose, in particular)[9] and benzene–cyclohexane gas-phase complexes are both small, while their dispersion energies are similar and much larger. Second, despite glucose being highly hydrophilic overall, it interacts with the Phe_{16} side chain via only its α-face, which is composed entirely of methine groups, and the methylene portion of the exocyclic C6 hydroxymethyl group. The T_m values of Pin WW variants **4** and **10** being similar implies that these parts of glucose taken in isolation may be as hydrophobic as a pure hydrocarbon like cyclohexane.

The rigidity of six-membered rings is an important factor in carbohydrate binding

Most of the monosaccharides that are common in nature are six-membered pyranose rings. Six-membered rings stand out among the smaller rings formed by saturated organic compounds in that they are quite rigid conformationally. The conformational rigidity of six-membered rings could play a role in protein–carbohydrate interactions (as well as other ligand binding processes), since packing of the ring against a protein surface could restrict the motion of the ring. Unfortunately, it is difficult to test this notion with carbohydrates because many factors would be confounded with the conformational entropy difference between the pyranose and furanose forms of a carbohydrate, in particular the types and positions of the substituents on the ring. Therefore, to determine whether the rigidity of six-membered pyranose rings might enhance their ability to pack with aromatic rings in protein–carbohydrate interactions, we compared the folding energetics of Pin WW variant **10**, in which R_1 = cyclohexyl, to variants **11**, **12**, and **13**, in which R_1 = cyclobutyl, cyclopentyl, and cycloheptyl, respectively, since four- five- and seven-membered rings are more conformationally heterogeneous than six-membered rings.[9]

Based on hydrophobicity alone, one would expect the native state stabilization by cycloalkyl substituents appended to Asn_{19} to simply increase with the size of the ring. However, the data in Figure 3 and Table S1 show that the native state stabilization peaks when R_1 = cyclohexyl and is diminished for the cycloheptyl, cyclobutyl and cyclopentyl rings. This decrease is unlikely to be due to steric constraints, since the cycloheptyl group is considerably smaller than GlcNAc. Rather, it is more consistent with the notion that the interaction of the Asn_{19} substituents with the phenyl ring of Phe_{16} is accompanied by a decrease in their conformational mobility, which places a greater entropic burden on four-, five-, and seven-membered rings than on six-membered rings. This result suggests that the rigidity of pyranose rings likely increases the strength of protein–carbohydrate interactions.

Structure-stability assessment by NMR

We used NMR spectroscopy to determine the solution structures of the Pin WW glycovariants, and thereby to gain a better understanding of the energetics and of the interaction geometry between Asn_{19}-bound sugars and the Phe_{16} side chain in our Pin WW glycovariants. Four monosaccharide-containing Pin WW variants were investigated: the galactose variant **5** (in which C4 is inverted relative to glucose), the allose variant **7** (in which C3 is inverted), the L-idose variant **8** (in which C5 is inverted), and the xylose variant **9** (in which C6 is deleted). A set of 2D-TOCSY, NOESY, and COSY spectra for each of the

glycovariants were collected to enable structure determination, as described previously. To assign the Pin WW resonances, the backbone NH and CHα protons were first assigned utilizing the TOCSY and NOESY spectra. The protein side chains were then assigned using the TOCSY spectrum. The sugar resonance assignments were made after Pin WW had been assigned, again with the aid of the COSY, TOCSY, and NOESY spectra.

Sugar resonances are shifted upfield due to their proximity to the Phe_{16} phenyl ring

When a CH–π interaction occurs, the chemical shift(s) of the proton(s) attached to the sugar should be shifted depending on their position relative to the aromatic ring current.[9] In particular, the ^1H chemical shift of a sugar C–H bond involved in a CH–π interaction will be shifted upfield since the C–H protons will be in the shielding cone of the aromatic ring. Thus, significant sugar chemical shift perturbations serve as an indicator of a CH–π interaction. In order to produce relevant reference spectra for the monosaccharides used in Pin WW variants **5**, **7**, **8**, and **9** (that is, spectra lacking aromatic chemical shift perturbations), we prepared asparagine conjugates with galactose (**19**), allose (**20**), L-idose (**21**), and xylose (**22**)—depicted in Chart 2—from their protected precursors (compounds **S5**, **18**, **S22**, and **S27**, respectively) by treating them with piperidine to remove the Fmoc group followed by 10% hydrazine to remove the acetyl groups (as described for the conversion of **7-Ac** to **7**). The resonances of the monosaccharide protons were assigned with the aid of two-dimensional DQF-COSY ^1H spectra (Figures S1–S4).

The chemical shift perturbations of the monosaccharide proton resonances in Pin WW glycovariants **5**, **7**, **8**, and **9** relative to the corresponding protons in the Asn-monosaccharide conjugates **19**, **20**, **21**, and **22** are shown in Figure 4 (see also Table S2). A negative value indicates an upfield chemical shift (shielding), whereas a positive value indicates a downfield shift (deshielding). For galactose variant (**5**) and allose variant (**7**), the C–H protons H5, H6a, and H6b exhibited the largest upfield shifts, with H5a and H6a being the most significant, whereas the chemical shift perturbations for the remainder of the sugar ring protons were moderate or small, strictly analogous to the previously reported chemical shift perturbations of GlcNAc in glycovariant **2** (also shown in Figure 4). This indicates that the orientations of the phenyl rings and the α-faces of the monosaccharide rings are likely very similar in Pin WW glycovariants **2**, **5**, and **7**. It is noteworthy, however, that the H6b proton of Gal in variant **5** displays a much smaller upfield shift ($\Delta\delta = -0.48$ ppm) relative to the H6 protons in other variants. We will address this observation below in light of the three-dimensional structure of Pin WW variant **5**.

For the xylose Pin WW variant **9**, which lacks the C6 carbon, the two H5 protons exhibit large upfield chemical shifts. In addition, the other four C–H groups of xylose all displayed moderate upfield shifts, which were greatest for the H3 and H4 protons. This result suggests that the aryl–sugar packing arrangement in the xylose Pin WW glycovariant (**9**) is likely somewhat different than it is for the GlcNAc (**2**), galactose (**5**), and allose (**7**) Pin WW glycovariants.

In the L-idose Pin WW glycovariant **8**, the C5 and C6 proteins are not shifted upfield to the extent that they are in Pin WW variants **2**, **5**, and **7**. Instead, all of the L-idose C–H protons

are moderately upfield shifted, consistent with multiple phenyl ring–sugar packing arrangements that are in fast exchange on the NMR timescale or an altogether different packing geometry relative to **2**, **5** and **7**.

General features of the structures of Pin WW glycovariants 5, 7, 8, and 9

To solve the NMR solution structures of Pin WW glycovariants **5**, **7**, **8**, and **9**, we used the assigned NOEs to identify distance restraints (Table S3) and generated 100 structures for each N-glycovariant by restrained molecular dynamics simulated annealing using the AMBER 11 software package as described previously. As was the case for the previously published structure of GlcNAcylated Pin WW glycovariant **2**, the Pin WW backbone structure did not change relative to unglycosylated Pin WW for any of the monosaccharide-containing Pin WW glycovariants studied herein, *i.e.*, **5**, **7**, **8**, and **9**. An overlay of the twenty lowest energy structures derived from the 100 calculated structures for each glycovariant is shown in Figure 2, enabling comparison to **2** and unglycosylated Pin WW **1**. The tertiary structure of Pin WW is not significantly affected by alteration of the carbohydrate ring, as shown by the small backbone RMSD (~ 1 Å) between the average structures of **2**, **5**, **78**, and **9**. These structures displayed minor variations only at the extreme N- and C-termini. Structural alterations within the type 1 β-bulge enhanced aromatic sequon of the glycovariants **5**, **7**, **8**, and **9** relative to the GlcNAc glycovariant **2** were also minor (exhibiting a RMSD of about 0.3 Å for residues 9–16).

Conformations of the Asn_{19}–glycan linkages in Pin WW glycovariants 5, 7, 8, and 9

A previously reported analysis of the conformation of Asn–GlcNAc linkages based on glycoprotein structures in the PDB shows that this bond has one dominant conformer with ϕ_N, ψ_N dihedral angles of 260.2° ± 21.1°, 176.8° ± 12.0° (all dihedrals are reported as mean ± standard deviation). The average ϕ_N, ψ_N dihedral angles of the N-linked sugars in **5**, **7**, and **9** were all close to these ranges (Figure 5): ϕ_N, ψ_N = 246.8° ± 2.5°, 171.6° ± 2.4° for **5**; ϕ_N, ψ_N = 245.1° ± 3.6°, 172.6° ± 1.9° for **7**; and ϕ_N, ψ_N = 259.9° ± 8.8°, 185.5° ± 1.5° for **9**. These dihedral angles are consistent with those in the previously reported NMR structure of **2** (ϕ_N, ψ_N = 255.3° ± 16.3°, 174.4° ± 21.2°; Figure 5A). Notably, the ϕ_N dihedral angle in the L-idose Pin WW glycovariant **8** is quite different (100.3° ± 8.5°) relative to the average ϕ_N angle in the other structures (252°), indicating that the glycosidic linkage populates a different rotamer in **8** than it does in **2**, **5**, **7**, and **9**. Thus, the sugar in **8** must be rotated relative to its orientation in **2**, **5**, **7**, and **9** such that L-idose engages in a different interaction with the phenyl ring of Phe_{16} (Figure 5D). Thus, the decrease in the ability of L-idose to stabilize the native state of Pin WW is accompanied by a major change in the protein–carbohydrate interaction geometry. We revisit the unusual nature of the interaction between L-idose and Phe_{16} below.

NMR structures are consistent with the chemical shift perturbation data and provide a potential explanation for the relatively weak interaction between Phe_{16} and galactose

Overall, the sugar–Phe_{16} packing arrangement within Pin WW glycovariants **5**, **7**, **8**, and **9** is consistent with the chemical shift perturbation data summarized in Figure 4. Similar interactions are observed between the aromatic ring of Phe_{16} and the galactose (**5**, Figure

5B), allose (**7**, Figure 5C) and GlcNAc (**2**, Figure 5A) substructures. The Phe_{16} aryl ring interacts predominantly with the axial H5 hydrogen and the two H6 hydrogens (Figures 5B and C). This result is consistent with the chemical shift perturbation data, wherein the H5 and H6 protons exhibit the largest upfield chemical shifts (Figure 4). However, a unique structural feature of the galactose Pin WW glycovariant **5**, not seen in the GlcNAc glycovariant **2** or the allose glycovariant **7**, is that the exocyclic hydroxymethyl group that projects from C5 of the pyranose ring populates two rotamers, the *gauche*$^-$ and *gauche*$^+$ rotamers (also known as the gg and gt rotamers, torsion angles $\omega \sim 300°$ and 60°, respectively). Ten of the twenty lowest energy structures adopt each rotamer (Figure 6). In contrast, only one of the twenty lowest energy structures of **7** and none of the twenty lowest energy structures of **2** adopted the *gauche*$^+$ rotamer. This finding is notable because in the *gauche*$^+$ rotamer only one of the two diastereotopic hydrogens attached to C6 (specifically, the *pro*-S hydrogen) is pointed toward the Phe_{16} phenyl group, whereas both of the C6 hydrogens are pointed toward the Phe_{16} phenyl group in the *gauche*$^-$ rotamer (Figure 6). This result reflects the intrinsic conformational preferences of the exocyclic hydroxymethyl group in the free monosaccharides. In free β-D-galactose, the *gauche*$^+$:*gauche*$^-$ ratio is about 3:1, whereas in β-D-glucose the ratio is 1:1. It also explains why the upfield shift of one of the H6 protons in the galactose Pin WW glycovariant **5** is relatively small (−0.48 ppm; Figure 4). Finally, it is worth noting that this galactose conformational preference may inhibit its ability to make optimal stabilizing contacts with the Phe_{16} phenyl ring. Both C6 hydrogens make contact with the Phe_{16} phenyl ring in the *gauche*$^-$ rotamer. The *gauche*$^+$ rotamer places one C6 hydrogen and an OH group proximal to the Phe_{16} phenyl ring—thereby providing a rationale for why glycosylating the enhanced aromatic sequon with galactose stabilizes the native state less than glycosylating it with GlcNAc, glucose, mannose, or allose (Figure 3).

A rationale for the relatively strong interaction between allose and the Phe16 phenyl ring based on the NMR structure of **7** is more elusive. It is noteworthy that the allose ring, which stabilizes Pin WW to a greater extent ($\Delta T_m = 11.1 \pm 0.3$ °C) than any of the non-peracetylated monosaccharides, is pitched somewhat higher in the structure of **7** (Figure 5C), i.e., the angle between the carbohydrate and phenyl rings (pitch angle = 28.4° ± 2.0°) is greater than the angle exhibited by GlcNAc in the structure of **2** (pitch angle = 13.9° ± 2.1°). This places the H5 of allose closer to being above the center of the phenyl ring ($d_{offset} = 0.46 \pm 0.06$ Å) than the H5 of GlcNAc is in **2** ($d_{offset} = 0.78 \pm 0.10$ Å). However, the geometry of the galactose–phenyl ring interaction in **5** is similar to that of the allose–phenyl ring interaction in **7** (Figure 5B, angle = 22.1° ± 3.6°, $d_{offset} = 0.54 \pm 0.16$ Å; all quantities relating to the geometry of the carbohydrate–aromatic interactions are reported as the mean ± standard deviation from the corresponding 20 lowest-energy NMR structures). Thus, while the geometry of the allose–phenyl ring interaction may contribute to the stabilization energy, other factors probably contribute as well. For example, the stereochemistry of the hydroxyl groups around the pyranose ring can affect how the ring is solvated, and the sugar's electrostatic surface potential. In addition, carbohydrate–protein interactions other than the CH–π interaction could be involved. In particular, Thr_{21} was found previously to make a small contribution to the stabilization afforded by five-residue enhanced aromatic sequons. In fact, the γ-hydroxyl group of Thr_{21} accepts a hydrogen bond from, or donates a hydrogen

bond to, the C2 substituent of the carbohydrate (acetamido in **2** and hydroxyl in **5**, **7**, and **9**) in many of the calculated structures of Pin WW glycovariants **2**, **5**, **7**, and **9**. However, given that the mannosylated Pin WW glycovariant **6**—in which the C2 hydroxyl cannot make this interaction because C2 is epimerized—is more stable than most of the monosaccharidylated Pin WW glycovariants (Figure 3), it seems unlikely that this (highly solvent exposed) hydrogen bond makes a substantial contribution to native state stability.[39–] If carbohydrate–Thr_{21} interactions modulate stabilization by enhanced aromatic sequons, it probably would be related to the burial of the non-polar parts of Thr_{21} by the carbohydrate.

L-idose in Pin WW glycovariant 8 interacts differently with Phe_{16} than other monosaccharides do

The L-idose in Pin WW glycovariant **8** features a C5 inversion relative to GlcNAc. In the chair conformer that is usually adopted by pyranose sugar rings—the $^4C_1$ conformer that can be seen in Pin WW variants **2**, **5**, and **7** (Figure 5A, B, and C)—this stereochemistry at C5 would place the exocyclic C6 hydroxymethyl group in an axial position, leading to a steric clash with the Phe_{16} phenyl ring (as well as a desolvation requirement). This arrangement is conformationally unstable, and as a result the pyranose ring of L-idose can flip to the other $^1C_4$ chair conformer (Figure 5D; L-idose is also depicted in the $^1C_4$ conformation in Chart 1). Specific evidence that L-idose in Pin WW glycovariant **8** adopts a $^1C_4$ chair conformation is apparent in the NOESY spectrum, which shows strong intramolecular NOEs between the H2, H3, and H4 ring protons (which would only be observed when they are in equatorial positions). These considerations suggest that L-idose will have to interact very differently with the Phe_{16} phenyl ring than other monosaccharides. This expectation is consistent with the unusual ϕ_N and ψ_N dihedral angles for the L-idose glycosidic bond noted above. Moreover, this atypical interaction is demonstrated by the full three-dimensional structure of Pin WW glycovariant **8** (Figure 5D).

Only two medium-strength NOEs were observed between L-idose ring protons and the rest of Pin WW, indicating that L-idose does not pack as well with Pin WW as other sugars do. The sugar hydrogens H1, H2, and H3 are the closest hydrogens to the phenyl ring of Phe_{16} in the twenty lowest-energy structures of Pin WW variant **8** (Figure 5D), which agrees with the chemical shift perturbation data (Figure 4), showing that the H1, H2, and H3 protons exhibit the largest (yet still modest) upfield shifts. Another unique feature of L-idose Pin WW glycovariant **8** was the location of the axial C3 hydroxyl group, which is positioned just above the Phe_{16} ring. Although the polar hydroxyl group can make an OH–π interaction with the Phe_{16} side chain, its desolvation is likely to be energetically costly and would diminish the net interaction free energy. In fact, the small increase in T_m upon L-idosylation of Pin WW ($\Delta T_m = 3.0 \pm 0.3$ °C) corresponds to a folding free energy difference of only about −0.3 kcal mol^{-1} (see Table S1). This energy is barely larger than the stabilization energy due simply to the effect of N-glycosylation on the conformational preferences of Asn_{19} (−0.2 kcal mol^{-1}) determined previously. Thus, the interaction of L-idose with the Phe_{16} side chain is likely a very weak one, if it occurs at all.

The NMR structure of Pin WW glycovariant 9 suggests how xylose maintains a strong interaction with Phe_{16} despite lacking a C6 exocyclic hydroxymethyl group

The exocyclic C6 hydroxymethyl group that is present in hexoses like GlcNAc, galactose, and allose is absent in xylose, which is a pentose; in other words, xylose has two hydrogen atoms on C5 instead of one hydrogen atom and a hydroxymethyl group. As a result, xylose cannot interact with Phe_{16} via the two H6 protons, as occurs in Pin WW glycovariants **2**, **5**, and **7**. To compensate for the loss of this interaction, the pitch angle between the planes of the aryl ring of Phe_{16} and the xylose ring in Pin WW glycovariant **9** is significantly greater (57.7 ± 9.3°; Figure 5E) than the corresponding angles in the GlcNAc, galactose, and allose Pin WW glycovariants **2**, **5**, and **7** (Figure 5A, B, and C) The H5a hydrogen of glycovariant **9** is situated almost directly above the center of the aryl ring of Phe_{16}, with d_{offset} = 0.24 ± 0.22 Å. The H5b hydrogen atom is not as close to the center of the Phe16 phenyl ring as H5a (d_{offset} = 1.68 ± 0.16 Å), but because the xylose ring is pitched at such a steep angle relative to the Phe16 phenyl ring, H5b is at a similar height above the ring ($d_{CH-\pi}$ = 2.80 ± 0.06 Å) as H5a ($d_{CH-\pi}$ = 2.85 ± 0.21 Å), and likely can contribute to the carbohydrate–aromatic interaction. This observation could explain how xylose, despite lacking an exocyclic hydroxymethyl group, is as stabilizing in the enhanced aromatic sequon as GlcNAc and slightly more stabilizing than glucose (Figure 3 and Table S1). Apparently the dual interaction between H5a and H5b with the Phe_{16} phenyl ring balances the loss of any interactions between the exocyclic hydroxymethyl group and the Phe_{16} phenyl ring.

DISCUSSION

CH–π interactions enable affinity and, to a lesser degree, selectivity in carbohydrate binding

Carbohydrates appear, at first glance, as if they should be relatively difficult species for proteins to bind. They are often uncharged, removing the possibility of attractive electrostatic interactions, and they are exceptionally well solvated by water. Nevertheless, proteins and carbohydrates can interact productively, and this interaction is often driven by CH–π interactions between CH groups projecting from the carbohydrate and aromatic amino acid side chains on the binding protein.[11] Such interactions occur both intramolecularly, between the glycan and protein components of glycoproteins,[7–10] as well as intermolecularly, between lectins and their carbohydrate ligands or the carbohydrate-binding modules of carbohydrate-active enzymes and their substrates.[1–6] Furthermore, our result that monosaccharides and cyclohexane interact with Phe_{16} with similar energetics suggests that the α-faces of carbohydrates may have local hydrophobicities that are comparable to hydrocarbons. Thus, hydrophobic burial contributes substantially to the energetics of carbohydrate–aromatic interactions.[11]

These factors together provide a foundation on which to build strong protein–carbohydrate interactions. In fact, we found that, with the exception of L-idose, all of the monosaccharides used in this study substantially stabilized Pin WW when attached to the side chain amide nitrogen of Asn_{19} in the five-residue enhanced aromatic sequon. The average ΔT_m for the non-peracetylated Pin WW glycovariants (excluding the L-idose glycovariant **8**) relative to unglycosylated Pin WW was 8.8 °C, and the ΔT_m values for all of the Pin WW

glycovariants were within 2.6 °C of this value. Thus, the effects of changing the stereochemistry of the functional groups on the sugar are relatively modest compared to the overall effect of glycosylation. This finding demonstrates the robustness of carbohydrate–aromatic interactions—and in particular the CH–π interactions that are their foundation—to sugar structural perturbations in enhanced aromatic sequons. Thus, CH–π interactions can improve the affinity with which proteins bind to virtually any carbohydrate. It is important to note, however, that the effect (or lack thereof) of sugar structural perturbations on Pin WW stabilization could be a function of the geometry of the carbohydrate–aromatic interaction that is enforced by the enhanced aromatic sequon. In the enhanced aromatic sequon, only one corner of the monosaccharide (the part in the vicinity of C5) interacts with the aromatic ring. Cases in which the monosaccharide is centered above the aromatic ring could be more sensitive to changes in the stereochemistry of the substituents on the monosaccharide.

While the differences between the stabilizing effects of particular monosaccharides are small compared to the overall interaction strength, carbohydrate–aromatic interactions can nevertheless contribute to the selectivity of carbohydrate binding. For example, in a binding site that had a carbohydrate–aromatic interaction similar to that in an enhanced aromatic sequon, one would expect binding of galactose to be disfavored relative to other monosaccharides. The degree of discrimination between monosaccharides that is possible solely via carbohydrate–aromatic interactions is modest, but can nevertheless supplement other sources of selectivity (for example, hydrogen bonding complementarity).

Electrostatics vs. dispersion forces in the energetics of CH–π interactions

We reported previously that the contribution of electrostatic forces to the net energy of CH–π interactions was small in our system. This finding was based on our observation of a lack of correlation between the extent to which N-GlcNAcylation of Asn_{19} stabilized Pin WW and the presence of electron-donating or electron-withdrawing substituents on the Phe_{16} ring. Consistent with this earlier result, here we have found that galactose, which has a more electropositive α-face than glucose, stabilizes Pin WW to a lesser extent than glucose does. Moreover, this trend holds for the N-acetyl hexosamine versions of galactose and glucose (GalNAc vs. GlcNAc) as well as for the peracetylated versions of these monosaccharides (Ac_4Gal vs. Ac_4Glc and Ac_3GalNAc vs. Ac_3GlcNAc; see Figure 3 and Table S1). Thus, the higher electropositivity of the α-face of galactose does not lead to its interacting more strongly than glucose with the Phe_{16} phenyl side chain in the enhanced aromatic sequon. This finding again suggests that the contribution of electrostatics to CH–π interactions between carbohydrates and the Phe_{16} side chain is small in our system.

Our conclusion about the contribution of electrostatics to CH–π interactions are in contrast with results recently reported by Hudson et al. and Jiménez-Moreno et al.[9] The former group of authors found by mining the RCSB Protein Data Bank that binding sites for galactose, which as noted above has the most positively charged α-face of the common sugars, were richer in aromatic amino acids (and especially tryptophan) than binding sites for other monosaccharides. Moreover, they showed that galactose associated more strongly than glucose with indole in aqueous solution, and that the strength of the galactose–indole interaction depended on how electron rich the indole ring system was. The latter group used

a dynamic combinatorial approach to measure the facial interaction strength for all combinations of a large series of monosaccharides and aromatics. They found that electron withdrawing substituents on the monosaccharides and electron donating substituents on the aromatics tended to increase the strength of the CH–π interaction.[9] In both cases, the results are consistent with a substantial electrostatic contribution to CH–π interactions.

A likely explanation for the differences between our system and those used by Hudson et al. and Jiménez-Moreno et al. is that, in our studies with the enhanced aromatic sequon, we have used relatively weakly charged substructures as reference points. In our previous work, we used GlcNAc as the sugar and varied the aromatic, but the α-face of GlcNAc is much less electropositive than the α-face of galactose. Similarly, in this work we have used the phenyl side chain of Phe_{16} as the CH–π acceptor, which is much less electron rich than the side chains of Tyr or Trp. We hypothesize that this choice muted the electrostatic contributions to the CH–π interactions that we studied and thereby brought the contributions from dispersion forces to the foreground. Thus, we view our results as complementary to (and not in conflict with) those obtained in other systems. Importantly, our results combined with those of Hudson et al. and Jiménez-Moreno et al.[9] suggest that CH–π interactions between different partners can be strong for different reasons: CH donors with highly polarized C–H bonds interact strongly with π systems that are electron rich via dispersion forces that are substantially supplemented by electrostatic interactions (for example, galactose with indole rings), whereas CH donors with weakly polarized C–H bonds interact strongly with π systems that are less electron rich via dispersion forces alone (for example, GlcNAc with phenyl rings). This observation is in some ways reminiscent of the concept of hardness and softness of Lewis acids and bases: hard acids prefer to react with hard bases and soft acids prefer to react with soft bases. It should also be noted that the geometry of the carbohydrate–aromatic interaction in our system is constrained by the covalent linkage of the carbohydrate to Asn_{19}. This factor may limit the contribution of electrostatic forces to the CH–π interaction in our system, as electrostatic interactions can be very sensitive to the relative orientations of the groups involved.[9]

Why are N-glycans always attached to proteins via GlcNAc residues in eukaryotes?

Most of the proteins that traverse the secretory pathway in eukaryotes are glycosylated. Thus, the intramolecular interaction between a glycoprotein and its attached glycan is an especially common type of protein–carbohydrate interaction. N-glycans are particularly interesting because, as noted in the Introduction, the sequence and composition of the glycan that is initially transferred by OST to asparagine side chains are almost universally conserved among eukaryotes.[9] N-glycans are always attached to asparagine side chains via a GlcNAc residue. This privileged position means that GlcNAc-1 is involved in many protein–carbohydrate interactions, which in turn suggests the question: why was GlcNAc evolutionarily selected for in this position? According to our results, GlcNAc forms neither the weakest nor the strongest interactions with Phe_{16} among the common monosaccharides. While undoubtedly many factors drive the evolutionary choice of one chemical substructure over another at a particular site in a biomacromolecule, metabolic constraints or compatibility with the relevant biosynthetic machinery are always possibilities. However, it can still be useful to speculate about such issues in terms of the suitability of the chemical

substructure to its role at the site in question. Such speculation has led to insights into, for example, the choice of phosphates *vs.* alternatives such as arsenate or silicate and the choice of pentoses vs. hexoses in the backbones of nucleic acids. If the choice of GlcNAc for the stem of N-glycans is to be understood in this way, it seems that there may be a "Goldilocks principle" in operation. GlcNAc can interact reasonably strongly with proteins, enabling it to stabilize protein native states when in certain structural contexts (for example, in enhanced aromatic sequons[7–]). On the other hand, GlcNAc–aromatic interactions are not so strong that they cause N-glycans to always adhere tightly to protein surfaces, which could inhibit the very important biological process of N-glycan maturation in the Golgi.[7] Thus, it seems possible that GlcNAc may have been selected for as the monosaccharide through which N-glycans are attached to proteins because it interacts with aromatics with a free energy that is "just right".

EXPERIMENTAL SECTION

General procedure for the preparation of peracetylated β-linked glycosyl amino acids

2,3,4,6-per-*O*-acetyl-β-D-allopyranosyl amine (16)—In a round-bottomed flask containing 1,2,3,4,6-per-*O*-acetyl-β-D-allopyranose **14** (2.2 g, 5.64 mmol) was added HBr in AcOH (30 mL) at 0 °C. The mixture was stirred overnight before removing the solvent *in vacuo*. The resulting mixture was dissolved in THF (50 mL) followed by the addition of 1M TBAF in THF (8 mL, 8 mmol) and $TMSN_3$ (1.05 mL, 8 mmol). After 5 h, the mixture was diluted with CH_2Cl_2, and washed successively with water, sat. aq. $NaHCO_3$, and brine. The organic portion was collected, dried over $MgSO_4$, and concentrated. The crude product was purified by EtOAc/Hex 1:2 to obtain **15** (1.4 g) in 72% as a colorless oil. To a round-bottomed flask containing **15** (1.4 g, 4.02 mmol) was added THF (30 mL). The solution was treated with PtO_2 (50 mg). The flask was evacuated and filled with H_2, repeating several times. The reaction mixture was then stirred under a hydrogen atmosphere for 3 h. The mixture was filtered through a pad of Celite and the solvent was removed *in vacuo*. The reaction residue was purified by silica gel chromatography (EtOAc/Hex 4:1) to afford compound **16** (1.05 g, 75%) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.58 (t, *J* = 2.9 Hz, 1H), 4.86 (dd, *J* = 10.3, 2.9 Hz, 1H), 4.74 (dd, *J* = 9.2, 2.9 Hz, 1H), 4.49 (d, *J* = 9.2 Hz, 1H), 4.13 (d, *J* = 3.1 Hz, 2H), 4.04 – 3.95 (m, 1H), 2.11 (s, 3H), 2.05 (s, 3H), 2.01 (s, 3H), 1.95 (s, 3H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 170.30, 170.25, 82.22, 70.98, 70.25, 68.99, 67.05, 63.10, 21.13, 20.94. The NMR spectra for **16** are shown in full in the Supporting Information. HRMS (ESI-TOF) calcd for $C_{14}H_{21}NO_9$ $(M + H)^+$, 348.1273; found, 348.1348.

Fmoc-Asn(Ac$_4$-β-D-allose)-OtBu (17)—To a stirred solution of **16** (530 mg, 1.53 mmol) in CH_2Cl_2 (16 mL), *N*-Fmoc-Asp-OtBu (628.4 mg, 1.53 mmol), benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) (953 mg, 1.83 mmol), and diisopropylamine (DIPEA) (0.55 mL) were added successively. The reaction was stirred for 18 h and then washed successively with water and brine. The collected organic portion was dried over $MgSO_4$ and concentrated. The residue was purified by flash chromatography (EtOAc/Hex 1:1) to afford **17** (918 mg) as a white solid in 81 % yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.79 (d, *J* = 7.6 Hz, 2H), 7.63 (d, *J* = 7.5 Hz, 2H), 7.42 (t, *J* = 7.5 Hz, 2H), 7.33 (t,

J = 7.5 Hz, 2H), 6.35 (d, *J* = 9.5 Hz, 1H), 5.94 (d, *J* = 8.6 Hz, 1H), 5.69 (t, *J* = 2.9 Hz, 1H), 5.62 – 5.56 (m, 1H), 4.93 (td, *J* = 9.9, 9.4, 2.8 Hz, 2H), 4.53 (dt, *J* = 9.3, 4.8 Hz, 1H), 4.46 (dd, *J* = 10.6, 7.2 Hz, 1H), 4.38 – 4.31 (m, 4H), 4.30 – 4.22 (m, 2H), 4.16 (ddd, *J* = 10.3, 4.3, 1.9 Hz, 1H), 4.11 (d, *J* = 12.5 Hz, 1H), 2.87 (dd, *J* = 16.3, 4.6 Hz, 1H), 2.75 (dd, *J* = 16.2, 4.7 Hz, 1H), 2.21 (s, 3H), 2.07 (s, 3H), 2.04 (s, 3H), 2.03 (s, 3H), 1.49 (s, 9H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 170.74, 170.63, 169.89, 169.79, 169.31, 156.11, 143.88, 143.75, 141.29, 127.73, 127.07, 125.17, 125.13, 120.00, 82.49, 75.62, 71.28, 68.21, 68.13, 67.20, 66.06, 61.92, 51.09, 47.13, 38.22, 27.89, 20.78, 20.71, 20.68, 20.53. HRMS (ESI-TOF) calcd for $C_{33}H_{37}N_3O_{13}$ $(M + H)^+$, 741.2865; found, 741.2861.

Fmoc-Asn(Ac$_4$-β-D-allose)-OH (18)—To a round-bottomed flask containing **17** (300 mg, 0.41 mmol) was added 10 mL of TFA/H_2O (95% (v/v)) and triisopropylsilane (100 μL). The reaction was stirred for 3 h. The reaction mixture was concentrated followed by lyophilization to obtain compound **18** as a white powder. ^{1}H NMR (600 MHz, CD_3OD) δ 7.79 (d, *J* = 7.6 Hz, 2H), 7.66 (dd, *J* = 7.5, 3.0 Hz, 2H), 7.39 (t, *J* = 7.4 Hz, 2H), 7.31 (td, *J* = 7.4, 4.8 Hz, 2H), 5.70 (t, *J* = 2.9 Hz, 1H), 5.52 (d, *J* = 9.5 Hz, 1H), 4.92 (dd, *J* = 10.2, 2.9 Hz, 1H), 4.57 – 4.52 (m, 1H), 4.33 (qd, *J* = 10.5, 7.1 Hz, 2H), 4.26 – 4.19 (m, 2H), 4.18 (ddd, *J* = 10.4, 4.5, 2.0 Hz, 1H), 4.13 (dd, *J* = 12.2, 2.1 Hz, 1H), 2.81 (dd, *J* = 16.3, 4.1 Hz, 1H), 2.75 (dd, *J* = 15.8, 7.1 Hz, 1H), 2.16 (s, 3H), 1.99 (s, 3H), 1.99 (s, 3H), 1.94 (s, 3H). ^{13}C NMR (151 MHz, CD_3OD) δ 170.90, 170.40, 169.77, 169.59, 143.90, 143.78, 141.15, 141.13, 127.38, 126.80, 126.76, 124.93, 124.89, 119.51, 74.94, 71.26, 68.43, 67.82, 66.71, 66.16, 62.03, 19.16, 19.14, 19.11, 19.06. The NMR spectra for **18** are shown in full in the Supporting Information. HRMS (ESI-TOF) calcd for $C_{33}H_{36}N_2O_{14}$ $(M + H)^+$, 685.2273; found, 685.2273.

Per-*O*-acetyl-allopyranosyl Pin WW (7-Ac)—Fmoc-glycine-*O*-tritylchloride resin (0.05 mmol) was swollen in CH_2Cl_2 for 30 min before conducting the synthesis. Peptide couplings were then performed to yield the sequence NH_2-KLPPGWEKRMFAN(per-*O*-acetyl-allopyranose)GTVYYFNHITNASQFERPS-CO_2H. HBTU (0.25 mmol), HOBT (0.25 mmol), and DIEA (0.5 mmol) were used for each amino acid coupling. The per-*O*-acetylated allose-attached Asn (**18**; 0.1 mmol) was double coupled. Each coupling was taken for 30 min with gentle shaking. The deprotection of Fmoc group was achieved by using 20% piperidine in DMF. After each coupling and deprotection, the resin was washed by DMF (3×), MeOH (3×), and CH_2Cl_2 (3×) successively. After the last step of Fmoc deprotection of lysine, the peptide was deprotected and cleaved from the resin by mixing with a deprotection cocktail (trifluoroacetic acid 8.5 mL + 0.5 mL water + 500 mg thiophenol + 0.25 mL thioanisole + 0.1 mol ethane dithiol + 0.1 mL triisopropylsilane) for 3 h. The crude peptide was then dissolved in MeCN/water and purified by preparative HPLC (20% MeCN to 40% MeCN) to afford glycopeptide **7-Ac**. The identity of **7-Ac** was confirmed by electrospray ionization mass spectrometry (ESI-MS; Table S4).

Allopyranosyl Pin WW (7)—The crude peptide **7-Ac** was treated with 5 mL of 10% hydrazine and 100 mM dithiothreitol (DTT) in MeCN/water for 4 h. The reaction mixture was then quenched by trifluoroacetic acid and lyophilized. The residue was purified by

preparative HPLC (20% MeCN to 35% MeCN) to afford Pin WW glycovariant **7**. The identity of **7** was confirmed by ESI-MS (Table S4).

CD thermal unfolding experiments—Unfolding measurements were obtained by monitoring the ellipticity signal at 227 nm from 0.2 to 108.2 °C in 2 °C increments on an Aviv 62A DS circular dichroism spectrometer using quartz cuvettes with a 1 cm path length. Peptide concentrations were 5–10 μM in 20 mM phosphate buffer (pH 7.0). Thermal denaturation data were fit to the van't Hoff equation assuming a two-state transition and linear pre- and post-transition baselines, as described previously.[29]

Determination of NMR Solution Structures of Pin1 WW domain Glycovariants—^{1}H 2D NMR spectra were obtained on Bruker DRX 600 MHz spectrometers at 285 K. Proton-Proton distance restraints were derived from 2D ^{1}H NOESY spectra acquired with a 100 ms mixing time. These distance restraints were employed in molecular dynamics simulations as described below. Details of the NMR restraints and resulting structure statistics are included in Table S3. The 20 lowest energy structures were analyzed using PROCHECK-NMR. All structures are available in the RCSB protein data bank with the following PDB IDs: galactose (Gal) Pin WW glycovariant **5**, PDB ID: 2NC4; allose (All) Pin WW glycovariant **7**, PDB ID: 2NC3; L-idose (L-Ido) Pin WW glycovariant **8**, PDB ID: 2NC6; xylose Pin WW glycovariant **9**, PDB ID: 2NC5.

Generation of Molecular-Dynamics-Refined Structures for use in NMR Structure Determination—Details of the simulations are provided in the Supporting Experimental Section of the Supplemental Information. Briefly, input coordinates were based on the protein structure of Pin WW glycovariant **2** in PDB ID 2M9E and generated by computational glycosylations of the protein by the Glycoprotein Builder utility at GLYCAM_Web (http://glycam.org/gp) or by the Nucleic Acids Builder utility in AmberTools. Parameter-topology files were built from the resulting structures using a combination of the ILDN and NMR modifications to the ff99SB force field and an updated form of the GLYCAM-06 force field. The structures were subjected to a simulated annealing protocol using the SANDER utility in AMBER 11 in implicit solvent, and using atomic restraints as dictated by the NMR results described above.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We would like to thank Dr. Sebastian Enck for his insightful advice on NMR data processing and structure determination and Dr. Gerard Kroon for his assistance with setting up NMR experiments and data processing. C.-H.H. was supported by a Talent Development Program postdoctoral fellowship between Academia Sinica and The Scripps Research Institute. This work was supported by the Skaggs Institute for Chemical Biology, the Lita Annenberg Hazen Foundation, and by National Institutes of Health Grants GM100058 (R.J.W. and B.L.F.), GM045811 (D.A.C), AI072155 (C.-H.W.), GM071862 (H.J.D.), and GM051105 (J.W.K.).

REFERENCES

(1). Weerapana E, Imperiali B. Glycobiology. 2006; 16:91R.

(2). Abu-Qarn M, Eichler J, Sharon N. Curr. Opin. Struct. Biol. 2008; 18:544. [PubMed: 18694827]
(3). Schwarz F, Aebi M. Curr. Opin. Struct. Biol. 2011; 21:576. [PubMed: 21978957]
(4). Mohorko E, Glockshuber R, Aebi M. J. Inherit. Metab. Dis. 2011; 34:869. [PubMed: 21614585]
(5). Kelleher DJ, Gilmore R. Glycobiology. 2006; 16:47R.
(6). Helenius A, Aebi M. Annu. Rev. Biochem. 2004; 73:1019. [PubMed: 15189166]
(7). Hebert DN, Lamriben L, Powers ET, Kelly JW. Nat. Chem. Biol. 2014; 10:902. [PubMed: 25325701]
(8). Ferris SP, Kodali VK, Kaufman RJ. Dis. Models Mech. 2014; 7:331.
(9). Wyss DF, Choi JS, Li J, Knoppers MH, Willis KJ, Arulanandam ARN, Smolyar A, Reinherz EL, Wagner G. Science. 1995; 269:1273. [PubMed: 7544493]
(10). Hanson SR, Culyba EK, Hsu TL, Wong CH, Kelly JW, Powers ET. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:3131. [PubMed: 19204290]
(11). Culyba EK, Price JL, Hanson SR, Dhar A, Wong CH, Gruebele M, Powers ET, Kelly JW. Science. 2011; 331:571. [PubMed: 21292975]
(12). Price JL, Powers DL, Powers ET, Kelly JW. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:14127. [PubMed: 21825145]
(13). Price JL, Culyba EK, Chen W, Murray AN, Hanson SR, Wong CH, Powers ET, Kelly JW. Biopolymers. 2012; 98:195. [PubMed: 22782562]
(14). Chen WT, Enck S, Price JL, Powers DL, Powers ET, Wong CH, Dyson HJ, Kelly JW. J. Am. Chem. Soc. 2013; 135:9877. [PubMed: 23742246]
(15). Murray AN, Chen WT, Antonopoulos A, Hanson SR, Wiseman RL, Dell A, Haslam SM, Powers DL, Powers ET, Kelly JW. Chem. Biol. 2015; 22:1052. [PubMed: 26190824]
(16). Dave K, Jager M, Nguyen H, Kelly JW, Gruebele M. J. Mol. Biol. 2016; 428:1617. [PubMed: 26880334]
(17). Jager M, Nguyen H, Crane JC, Kelly JW, Gruebele M. J. Mol. Biol. 2001; 311:373. [PubMed: 11478867]
(18). Laughrey ZR, Kiehna SE, Riemen AJ, Waters ML. J. Am. Chem. Soc. 2008; 130:14625. [PubMed: 18844354]
(19). Kiehna SE, Laughrey ZR, Waters ML. Chem. Commun. 2007:4026.
(20). Banerjee S, Vishwanath P, Cui J, Kelleher DJ, Gilmore R, Robbins PW, Samuelson J. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:11676. [PubMed: 17606910]
(21). Jager M, Dendle M, Kelly JW. Protein Sci. 2009; 18:1806. [PubMed: 19565466]
(22). Jager M, Deechongkit S, Koepf EK, Nguyen H, Gao JM, Powers ET, Gruebele M, Kelly JW. Biopolymers. 2008; 90:751. [PubMed: 18844292]
(23). Jager M, Dendle M, Fuller AA, Kelly JW. Protein Sci. 2007; 16:2306. [PubMed: 17766376]
(24). Jager M, Nguyen H, Dendle M, Gruebele M, Kelly JW. Protein Sci. 2007; 16:1495. [PubMed: 17586778]
(25). Nguyen H, Jager M, Kelly JW, Gruebele M. J. Phys. Chem. B. 2005; 109:15182. [PubMed: 16852923]
(26). Jager M, Zhang Y, Bieschke J, Nguyen H, Dendle M, Bowman ME, Noel JP, Gruebele M, Kelly JW. Proc. Natl. Acad. Sci. U. S. A. 2006; 103:10648. [PubMed: 16807295]
(27). Deechongkit S, Nguyen H, Jager M, Powers ET, Gruebele M, Kelly JW. Curr. Opin. Struct. Biol. 2006; 16:94. [PubMed: 16442278]
(28). Nguyen H, Jager M, Moretto A, Gruebele M, Kelly JW. Proc. Natl. Acad. Sci. U. S. A. 2003; 100:3948. [PubMed: 12651955]
(29). Deechongkit S, Dawson PE, Kelly JW. J. Am. Chem. Soc. 2004; 126:16762. [PubMed: 15612714]
(30). Deechongkit S, Nguyen H, Powers ET, Dawson PE, Gruebele M, Kelly JW. Nature. 2004; 430:101. [PubMed: 15229605]
(31). Fuller AA, Du D, Liu F, Davoren JE, Bhabha G, Kroon G, Case DA, Dyson HJ, Powers ET, Wipf P, Gruebele M, Kelly JW. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:11067. [PubMed: 19541614]

(32). Liu F, Du DG, Fuller AA, Davoren JE, Wipf P, Kelly JW, Gruebele M. Proc. Natl. Acad. Sci. U. S. A. 2008; 105:2369. [PubMed: 18268349]

(33). Weikl TR. Biophys. J. 2008; 94:929. [PubMed: 17905840]

(34). Mittal J, Best RB. Biophys J. 2010; 99:L26. [PubMed: 20682244]

(35). Karanicolas J, Brooks CL 3rd. Proc. Natl. Acad. Sci. U. S. A. 2004; 101:3432. [PubMed: 14981252]

(36). Sharpe T, Jonsson AL, Rutherford TJ, Daggett V, Fersht AR. Protein Sci. 2007; 16:2233. [PubMed: 17766370]

(37). Ji CG, Xiao X, Zhang JZ. J. Chem. Theory Comput. 2012; 8:2157. [PubMed: 26593846]

(38). Pandey BK, Smith MS, Torgerson C, Lawrence PB, Matthews SS, Watkins E, Groves ML, Prigozhin MB, Price JL. Bioconjug. Chem. 2013; 24:796. [PubMed: 23578107]

(39). Lawrence PB, Gavrilov Y, Matthews SS, Langlois MI, Shental-Bechor D, Greenblatt HM, Pandey BK, Smith MS, Paxman R, Torgerson CD, Merrell JP, Ritz CC, Prigozhin MB, Levy Y, Price JL. J. Am. Chem. Soc. 2014; 136:17547. [PubMed: 25409346]

(40). Chao SH, Matthews SS, Paxman R, Aksimentiev A, Gruebele M, Price JL. J. Phys. Chem. B. 2014; 118:8388. [PubMed: 24821319]

(41). Davis CM, Dyer RB. J. Am. Chem. Soc. 2016; 138:1456. [PubMed: 26750867]

(42). Ranganathan R, Lu KP, Hunter T, Noel JP. Cell. 1997; 89:875. [PubMed: 9200606]

(43). Kowalski JA, Liu K, Kelly JW. Biopolymers. 2002; 63:111. [PubMed: 11786999]

(44). Mortenson DE, Kreitler DF, Yun HG, Gellman SH, Forest KT. Acta Crystallogr. D Biol. Crystallogr. 2013; 69:2506. [PubMed: 24311591]

(45). Carpino LA, Han GY. J. Org. Chem. 1972; 37:3404.

(46). Price JL, Shental-Bechor D, Dhar A, Turner MJ, Powers ET, Gruebele M, Levy Y, Kelly JW. J. Am. Chem. Soc. 2010; 132:15359. [PubMed: 20936810]

(47). Wheeler SE, Houk KN. J. Chem. Theory Comput. 2009; 5:2301. [PubMed: 20161573]

(48). Tsuzuki S, Uchimaru T, Mikami M. J. Phys. Chem. B. 2009; 113:5617. [PubMed: 19331351]

(49). Tsuzuki S, Uchimaru T, Mikami M. J. Phys. Chem. A. 2011; 115:11256. [PubMed: 21812469]

(50). Fujii A, Hayashi H, Park JW, Kazama T, Mikami N, Tsuzuki S. Phys Chem Chem Phys. 2011; 13:14131. [PubMed: 21566845]

(51). O'Neal HE, Benson SW. J. Chem. Eng. Data. 1970; 15:266.

(52). Page MI, Jencks WP. Proc. Natl. Acad. Sci. U. S. A. 1971; 68:1678. [PubMed: 5288752]

(53). Roldos V, Canada FJ, Jimenez-Barbero J. Chembiochem. 2011; 12:990. [PubMed: 21500331]

(54). Case DA. Curr. Opin. Struct. Biol. 1998; 8:624. [PubMed: 9818268]

(55). Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. Glycobiology. 2004; 14:103. [PubMed: 14514716]

(56). Bock K, Duus JO. J. Carbohydr. Chem. 1994; 13:513.

(57). Mayorkas N, Rudic S, Cocinero EJ, Davis BG, Simons JP. Phys Chem Chem Phys. 2011; 13:18671. [PubMed: 21952414]

(58). Jin L, Simons JP, Gerber RB. J. Phys. Chem. A. 2012; 116:11088. [PubMed: 22559823]

(59). Hudson KL, Bartlett GJ, Diehl RC, Agirre J, Gallagher T, Kiessling LL, Woolfson DN. J. Am. Chem. Soc. 2015; 137:15152. [PubMed: 26561965]

(60). Gao J, Bosco DA, Powers ET, Kelly JW. Nat. Struct. Mol. Biol. 2009; 16:684. [PubMed: 19525973]

(61). Albeck S, Unger R, Schreiber G. J. Mol. Biol. 2000; 298:503. [PubMed: 10772866]

(62). Pace CN, Scholtz JM, Grimsley GR. FEBS Lett. 2014; 588:2177. [PubMed: 24846139]

(63). Asensio JL, Arda A, Canada FJ, Jimenez-Barbero J. Acc. Chem. Res. 2013; 46:946. [PubMed: 22704792]

(64). Nishio M, Umezawa Y, Fantini J, Weiss MS, Chakrabarti P. Phys Chem Chem Phys. 2014; 16:12648. [PubMed: 24836323]

(65). Quiocho FA. Pure Appl. Chem. 1989; 61:1293.

(66). Weis WI, Drickamer K. Annu. Rev. Biochem. 1996; 65:441. [PubMed: 8811186]

(67). Gabius HJ, Andre S, Jimenez-Barbero J, Romero A, Solis D. Trends Biochem. Sci. 2011; 36:298. [PubMed: 21458998]

(68). Luis AS, Venditto I, Temple MJ, Rogowski A, Basle A, Xue J, Knox JP, Prates JA, Ferreira LM, Fontes CM, Najmudin S, Gilbert HJ. J. Biol. Chem. 2013; 288:4799. [PubMed: 23229556]

(69). Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Biochem. J. 2004; 382:769. [PubMed: 15214846]

(70). Abbott DW, van Bueren AL. Curr. Opin. Struct. Biol. 2014; 28:32. [PubMed: 25108190]

(71). Knott BC, Crowley MF, Himmel ME, Stahlberg J, Beckham GT. J. Am. Chem. Soc. 2014; 136:8810. [PubMed: 24869982]

(72). Payne CM, Bomble YJ, Taylor CB, McCabe C, Himmel ME, Crowley MF, Beckham GT. J. Biol. Chem. 2011; 286:41028. [PubMed: 21965672]

(73). Barb AW, Borgert AJ, Liu M, Barany G, Live D. Methods Enzymol. 2010; 478:365. [PubMed: 20816490]

(74). Jimenez-Moreno E, Jimenez-Ose ´s, G. Gomez AM, Santana AG, Corzana F, Bastida A, Jimenez-Barberodef J, Asensio JL. Chem. Sci. 2015; 6:6076.

(75). Jimenez-Moreno E, Gomez AM, Bastida A, Corzana F, Jimenez-Oses G, Jimenez-Barbero J, Asensio JL. Angew. Chem., Int. Ed. 2015; 54:4344.

(76). Pearson RG. J. Am. Chem. Soc. 1963; 85:3533.

(77). Stone AJ. Chem. Phys. Lett. 1993; 211:101.

(78). Apweiler R, Hermjakob H, Sharon N. Biochim. Biophys. Acta. 1999; 1473:4. [PubMed: 10580125]

(79). Westheimer FH. Science. 1987; 235:1173. [PubMed: 2434996]

(80). Eschenmoser A. Angew. Chem., Int. Ed. 2011; 50:12412.

(81). Hang I, Lin CW, Grant OC, Fleurkens S, Villiger TK, Soos M, Morbidelli M, Woods RJ, Gauss R, Aebi M. Glycobiology. 2015; 25:1335. [PubMed: 26240167]

(82). Thaysen-Andersen M, Packer NH. Glycobiology. 2012; 22:1440. [PubMed: 22798492]

(83). Macke TJ, Case DA. ACS Sym. Ser. 1998; 682:379.

(84). Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Proteins. 2010; 78:1950. [PubMed: 20408171]

(85). Li DW, Bruschweiler R. Angew. Chem., Int. Ed. 2010; 49:6778.

(86). Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Proteins. 2006; 65:712. [PubMed: 16981200]

(87). Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, Woods RJ. J. Comput. Chem. 2008; 29:622. [PubMed: 17849372]

(88). Case, DA.; Darden, TA.; Cheatham, CL., III; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Walker, RC.; Zhang, W.; Merz, KM.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, KF.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, SR.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M-J.; Cui, G.; Roe, DR.; Mathews, DH.; Seetin, MG.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, PA. AMBER 11. University of California; San Francisco: 2010.

(89). Pellegrini E, Field MJ. J. Phys. Chem. A. 2002; 106:1316.

(90). Tsui V, Case DA. Biopolymers. 2000; 56:275. [PubMed: 11754341]

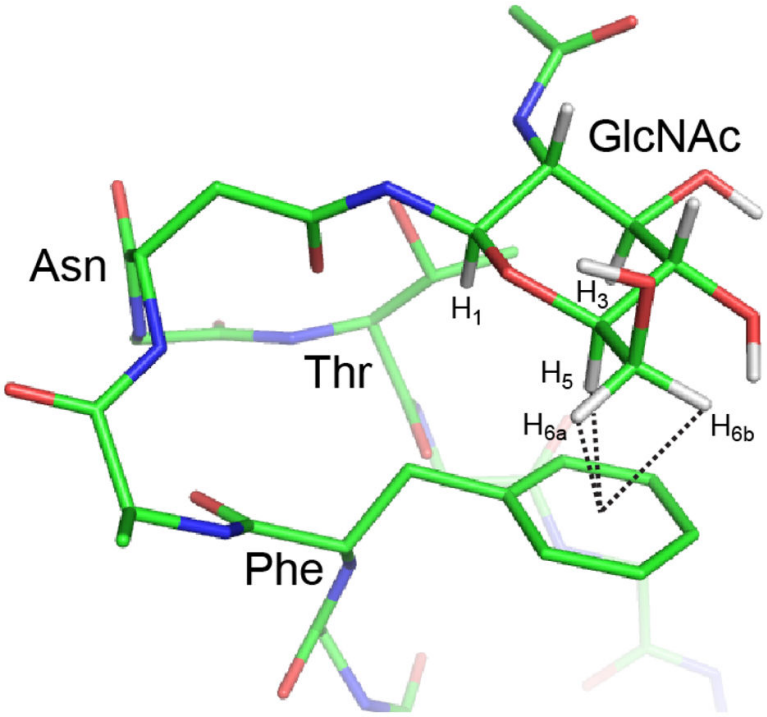


Figure 1.

A five-residue enhanced aromatic sequon in the context of the WW domain from the human protein Pin1 (PDB ID: 2M9F). Hydrogen atoms are omitted for clarity except for those on the pyranose ring of GlcNAc. Dashed lines indicate CH–π interactions.

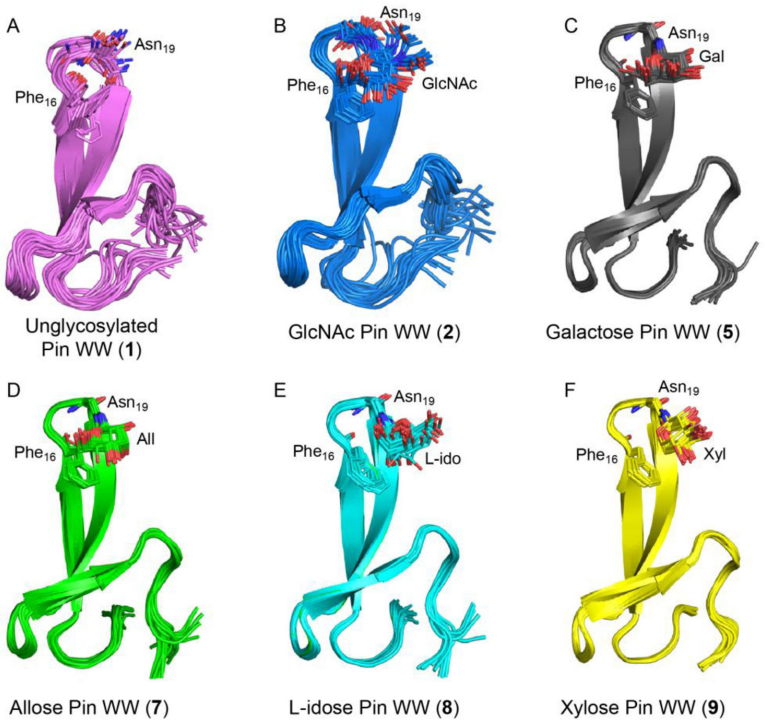


Figure 2.

Overlay of the twenty lowest-energy three-dimensional NMR-based structures of Pin WW N-glycovariants. (A) Unglycosylated Pin WW variant **1** (PDB ID: 2M9E). (B) GlcNAc PinWW glycovariant **2** (PDB ID: 2M9F). (C) Galactose (Gal) Pin WW glycovariant **5** (PDB ID: 2NC4). (D) Allose (All) Pin WW glycovariant **7** (PDB ID: 2NC3). (E) L-idose (L-Ido) Pin WW glycovariant **8** (PDB ID: 2NC6). (F) Xylose Pin WW glycovariant **9** (PDB ID: 2NC5). Note the overall similarity of the three-stranded β-sheet structure independent of the monosaccharide attached to Asn_{19}.

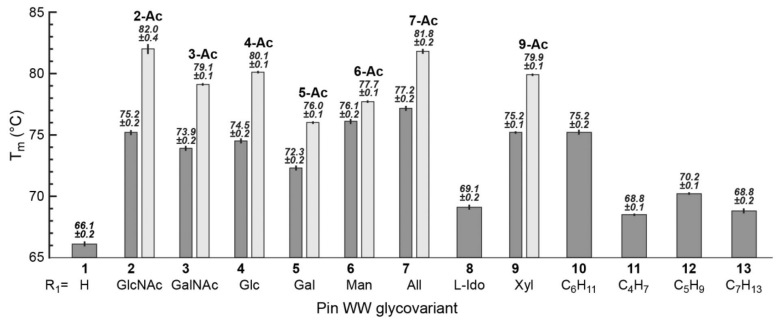


Figure 3.

Thermal denaturation midpoints of Pin WW glycovariants. Dark gray bars are for unprotected glycovariants. Light gray bars are for peracetylated glycovariants. Data are reported as means and standard errors from triplicate measurements (except glycovariant **1**, which was measured in duplicate). See Table S1 for detailed thermodynamic data.

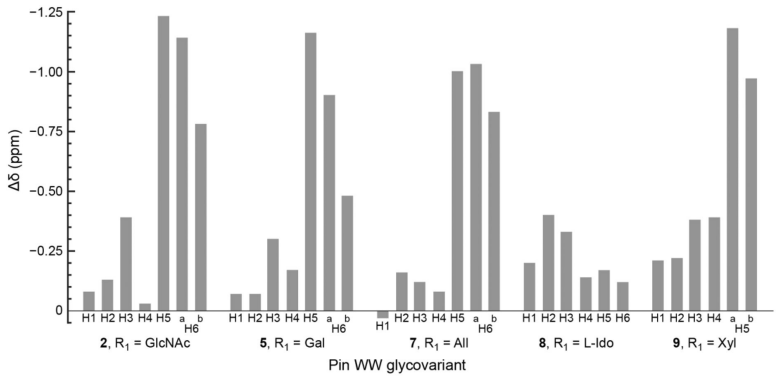


Figure 4.

Carbohydrate ring proton chemical shift perturbations in Pin WW glycovariants relative to the same carbohydrate attached to the free asparagine amino acid. Note the inverted scale on the y-axis. Negative numbers indicate upfield shifts; positive numbers indicate downfield shifts. The diastereotopic H6 protons in L-idose in Pin WW glycovariant **8** are represented by a single bar because they are not distinguishable. The data for Pin WW glycovariant **2** (R_1 = GlcNAc) were previously reported in ref. 14.

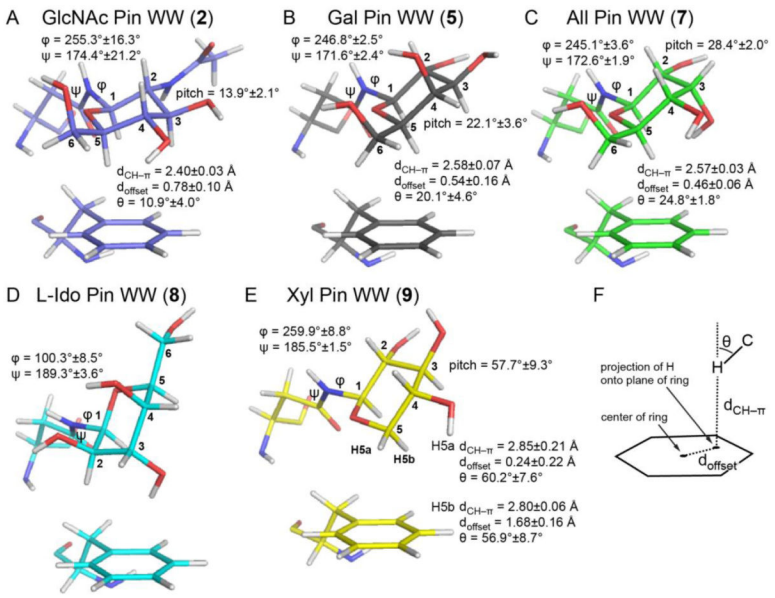


Figure 5.

Carbohydrate-aromatic interaction modes in the NMR solution structures of (A) GlcNAc Pin WW glycovariant **2** (PDB ID: 2M9F); (B) galactose (Gal) Pin WW glycovariant **5** (PDB ID: 2NC4); (C) allose (All) Pin WW glycovariant **7** (PDB ID: 2NC3); (D) L-idose (L-Ido) Pin WW glycovariant **8** (PDB ID: 2NC6); (E) xylose Pin WW glycovariant **9** (PDB ID: 2NC5). For each structure, we indicate the ϕ and ψ dihedral angles of the glycosidic linkage between Asn and the monosaccharides. For **2**, **5**, **7**, and **9**, we indicate the angle between the pyranose and phenyl rings (“pitch”); the distance between the H and the plane of the aromatic ring in the CH–π interaction (“$d_{CH-\pi}$”); the distance from the center of the aromatic ring to the projection of H5 onto the plane of the aromatic ring (“d_{offset}”); and the angle by which the C–H bond vector deviates from being perpendicular to the plane of the aromatic ring (“θ”; see ref. 59). (F) An illustration defining the aforementioned quantities.

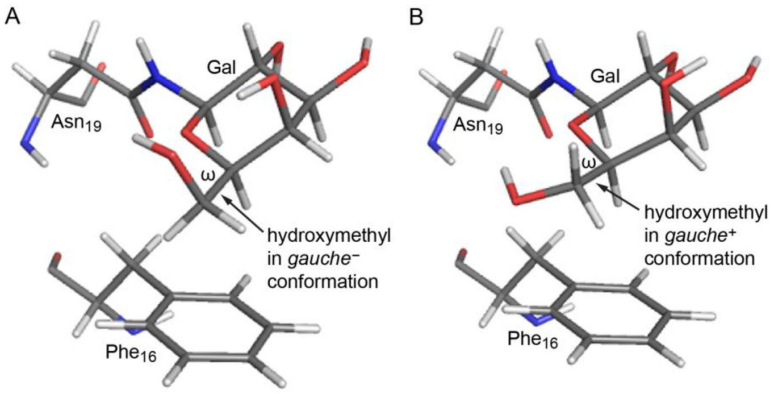


Figure 6.

Two structures from the twenty lowest energy structures of the galactose Pin WW glycovariant **5** showing the exocyclic hydroxymethyl torsion (ω) in the (A) *gauche*$^{-}$ and (B) *gauche*$^{+}$ conformations.

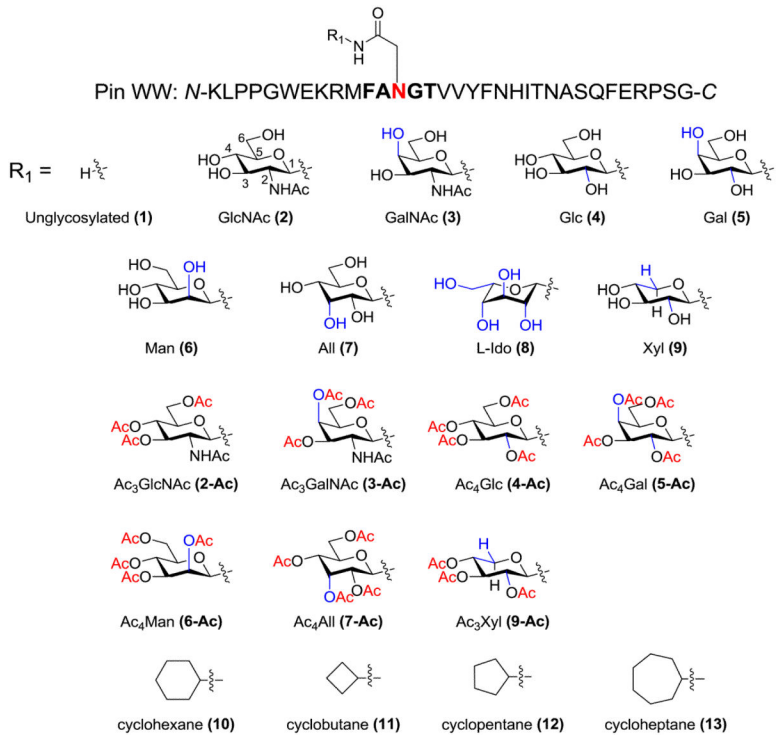


Chart 1.

Top: The sequence of Pin WW used in these studies. The residues that constitute the loop 1 enhanced aromatic sequon are shown in bold and the asparagine (Asn_{19}) that is glycosylated is shown in red. Bottom: Sugar and cycloalkane substructures incorporated as R_1 groups attached to the side chain amide nitrogen of Asn_{19}. Hydroxyl groups with stereochemistry that differs from that in glucose are highlighted in blue. Acetyl groups are highlighted in red.

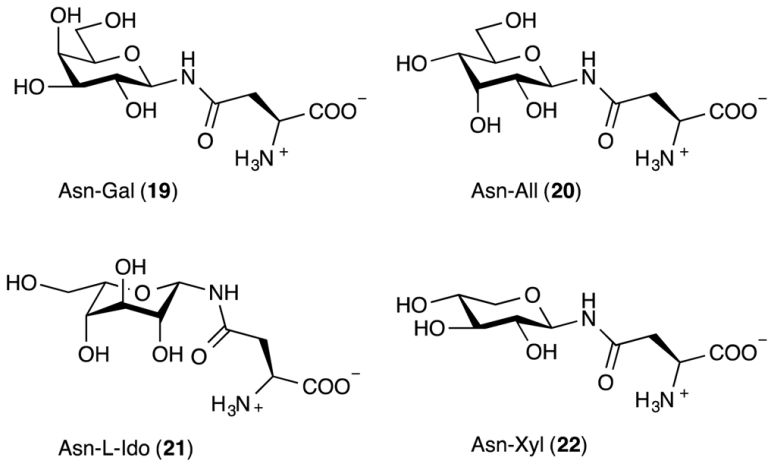
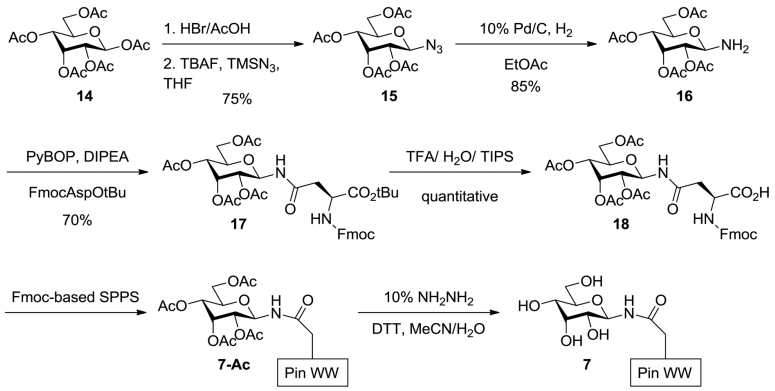


Chart 2.

Asn-monosaccharides **19–22** synthesized to determine the unperturbed chemical shifts of the sugar protons.



Scheme 1.

Representative synthesis of a protected glycosylated asparagine derivative (in this case, allosyl asparagine, **18**) and its incorporation by solid phase peptide synthesis (SPPS) into Pin WW.