JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Structural and Energetic Basis of Carbohydrate−Aromatic Packing Interactions in Proteins

Wentao Chen,[†,‡,§,#] Sebastian Enck,[‡,§,#,▲] Joshua L. Price,[†,‡,△] David L. Powers,[⊥] Evan T. Powers,*[,‡,§] Chi-Huey Wong,*[,‡,§,∥] H. Jane Dyson,*[,¶] and Jeffery W. Kelly*[,†,‡,§]

[†]Department of Molecular and Experimental Medicine, [‡]The Skaggs Institute for Chemical Biology, and [§]Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, United States

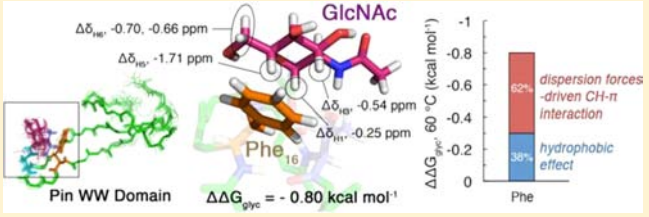
[⊥]Department of Mathematics, Clarkson University, Potsdam, New York 13676, United States

[∥]Genomics Research Center, Academica Sinica, Taipei 115, Taiwan

[¶]Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, United States

Supporting Information

ABSTRACT: Carbohydrate−aromatic interactions mediate many biological processes. However, the structure−energy relationships underpinning direct carbohydrate−aromatic packing interactions in aqueous solution have been difficult to assess experimentally and remain elusive. Here, we determine the structures and folding energetics of chemically synthesized glycoproteins to quantify the contributions of the hydrophobic effect and CH−π interactions to carbohydrate−aromatic packing interactions in proteins. We find that the hydrophobic effect contributes significantly to protein−carbohydrate interactions. Interactions between carbohydrates and aromatic amino acid side chains, however, are supplemented by CH−π interactions. The strengths of experimentally determined carbohydrate CH−π interactions do not correlate with the electrostatic properties of the involved aromatic residues, suggesting that the electrostatic component of CH−π interactions in aqueous solution is small. Thus, tight binding of carbohydrates and aromatic residues is driven by the hydrophobic effect and CH−π interactions featuring a dominating dispersive component.



INTRODUCTION

Carbohydrates and proteins are the two most abundant biopolymers in nature, and binding interactions between them are vital for life.[1,2] Covalently attached carbohydrates influence glycoprotein folding, through both intrinsic protein−carbohydrate interactions[3−6] and extrinsic chaperone-mediated mechanisms.[7] Interactions between carbohydrates and carbohydrate-binding proteins, or lectins, on proximal cells direct many biological processes,[8] especially immune system function.[9] An understanding of the nature of intra- and intermolecular protein−carbohydrate interactions is essential for understanding these processes. Such an understanding would also improve our ability to, for example, engineer cellulases to improve their substrate affinity for biofuel production,[10] design synthetic lectins for carbohydrate sensing,[11] tailor glycomimetic drugs to bind strongly to target proteins,[12] or improve the stability and shelf life of protein-based therapeutics by introducing N-glycosylation sites.[4,13]

Protein−carbohydrate interactions have been studied experimentally using peptide-based[14] and host−guest-based synthetic model systems.[11,15−18] The binding of carbohydrates by lectins and other proteins has also been investigated.[19−22] From these studies, it is clear that protein−carbohydrate interactions are driven by the formation of desolvated hydrogen bonds, the hydrophobic effect, and CH−π interactions.[19,20,23] However, studies aimed at understanding the relative contributions of each of these forces are rare.[14,15] Moreover, such studies are often limited by the difficulty of synthesizing a series of "hosts" for the carbohydrate "guests", each with enough diversity to simultaneously quantify the various contributions to the overall interaction energy. Quantum mechanical studies of diverse carbohydrate−aromatic interactions have successfully quantified the contributions of electrostatics, induction, dispersion, etc. to the interaction energy,[24−27] but such studies must be performed on molecules in the gas phase. The balance of forces in aqueous solution may be different since the net interaction energy for a carbohydrate−aromatic binding event in a protein is the difference between the energies of the protein−carbohydrate interaction and the sum of the protein−solvent and carbohydrate−solvent interactions. To experimentally parse the energy of carbohydrate−aromatic interactions in aqueous solution, we have utilized enhanced aromatic sequons, a class of structural modules that can be placed in proteins to enforce an intramolecular protein−carbohydrate interaction in the folded conformation.

Received: April 23, 2013
Published: June 7, 2013

dx.doi.org/10.1021/ja4040472 | J. Am. Chem. Soc. 2013, 135, 9877−9884

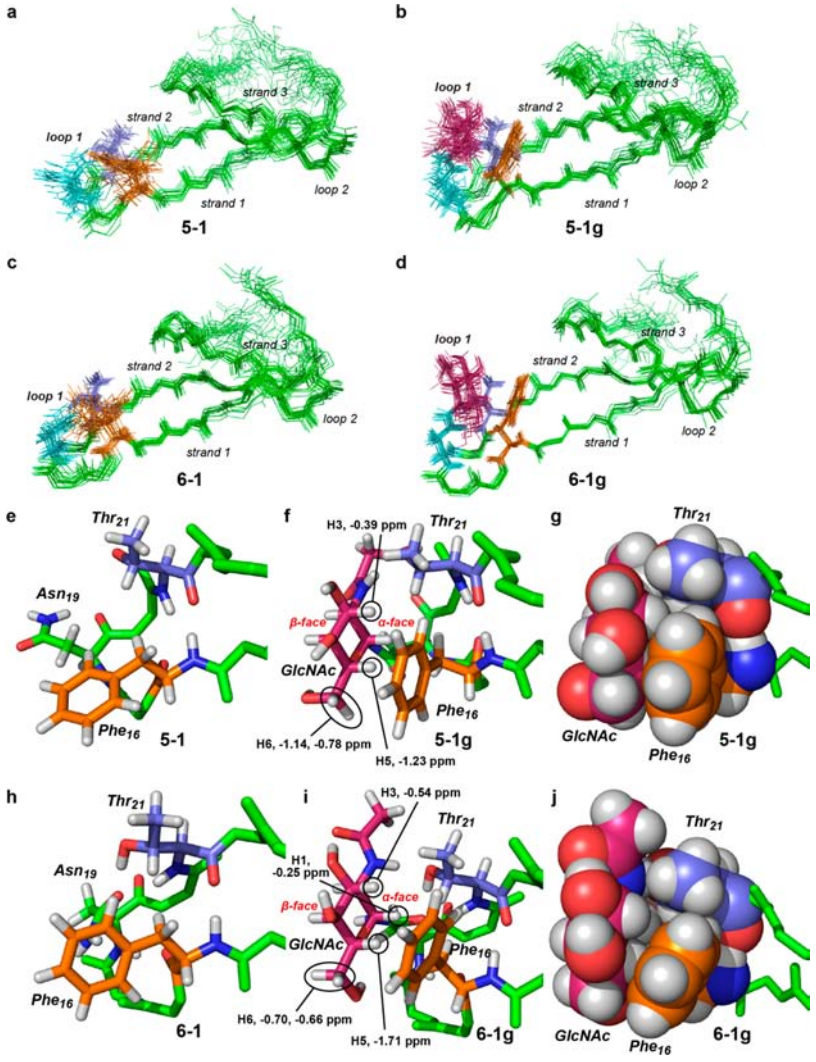


Figure 1. (a) Superposition of the main chains of 20 structures of Pin WW variant **5-1** (PDB code: 2M9E). (b) Superposition of the main chains of 20 structures of Pin WW variant **5-1g**, in which Asn_{19} is glycosylated with a single GlcNAc (PDB code: 2M9F). (c) Same as panel a, but for 19 structures of Pin WW variant **6-1** (PDB code: 2M9I). (d) Same as panel b, but for 17 structures of Pin WW variant **6-1g** (PDB code: 2M9J). (e) Enlarged view of loop 1 from Pin WW variant **5-1**. (f) Enlarged view of loop 1 from Pin WW variant **5-1g**. Chemical shift perturbations of GlcNAc protons ($\Delta\delta$) are shown in ppm. (g) Same as panel f, but with GlcNAc, Phe, and Thr rendered in CPK format. (h) Enlarged view of loop 1 from Pin WW variant **6-1**. (i) Enlarged view of loop 1 from Pin WW variant **6-1g**. Chemical shift perturbations of GlcNAc protons ($\Delta\delta$) are shown in ppm. (j) Same as panel i, but with GlcNAc, Phe, and Thr rendered in CPK format. In all panels, Phe_{16}, Asn_{19}, Thr_{21}, and GlcNAc are shown in gold, cyan, blue, and magenta, respectively. Images of the structures were made using PyMOL.

The typical sequence, or sequon, that the enzyme oligosaccharyltransferase recognizes to attach an oligosaccharide onto the side chain amide nitrogen of Asn is Asn-Xxx-Ser/Thr.[28] Enhanced aromatic sequons are sequences that include an aromatic residue that is two residues N-terminal to a sequon in a type 1 β-turn with a G1 β-bulge (a five-residue enhanced aromatic sequon) or three residues N-terminal to a sequon in a type 2 β-turn within a larger six-residue loop (a six-residue enhanced aromatic sequon).[4,29] Using the β-sheet-rich 34-residue WW domain from the human Pin1 protein (Pin WW) as a scaffold, we previously showed that loop 1 of Pin WW (residues 16−21) can be converted to enhanced aromatic sequons with five (sequence: Phe_{16}-Ala_{18}-Asn_{19}-Gly_{20}-Thr_{21}) or six (sequence: Phe_{16}-Arg_{17}-Ser_{18}-Asn_{19}-Gly_{20}-Thr_{21}) residues, and that glycosylation of Asn_{19} with *N*-acetylglucosamine (GlcNAc) stabilizes the native state in each case.[4,29] In both cases, thermodynamic cycle analysis indicated that most of the glycosylation-associated stabilization is due to interactions between the Phe_{16} aromatic ring and the GlcNAc.[4,29] Thus, enhanced aromatic sequons within Pin WW could be an ideal structural setting for probing protein−carbohydrate molecular recognition. Pin WW can be synthesized chemically, enabling us to incorporate an array of natural and unnatural amino acids into the protein−carbohydrate interaction site (position 16 in the Pin WW sequence) and to use linear free energy relationships[30] to quantify the contributions of different forces to the interaction energy, which can be measured from changes in the folding free energy of Pin WW.

EXPERIMENTAL SECTION

Peptide Synthesis and Thermal Denaturation. Glycosylated and nonglycosylated variants of Pin WW were synthesized by solid-phase peptide synthesis by using a standard Fmoc α-amine protecting group strategy as described previously.[4,29] The sequences of the variants were as follows.

dx.doi.org/10.1021/ja4040472 | J. Am. Chem. Soc. 2013, 135, 9877−9884

Five-residue enhanced aromatic sequon (**5-1**): Lys_6-Leu_7-Pro_8-Pro_9-Gly_{10}-Trp_{11}-Glu_{12}-Lys_{13}-Arg_{14}-Met_{15}-Phe_{16}-Ala_{18}-$\underline{Asn_{19}}$-Gly_{20}-Thr_{21}-Val_{22}-Tyr_{23}-Tyr_{24}-Phe_{25}-Asn_{26}-His_{27}-Ile_{28}-Thr_{29}-Asn_{30}-Ala_{31}-Ser_{32}-Gln_{33}-Phe_{34}-Glu_{35}-Arg_{36}-Pro_{37}-Ser_{38}-Gly_{39}.

Six-residue enhanced aromatic sequon (**6-1**): Lys_6-Leu_7-Pro_8-Pro_9-Gly_{10}-Trp_{11}-Glu_{12}-Lys_{13}-Arg_{14}-Met_{15}-Phe_{16}-Arg_{17}-Ser_{18}-$\underline{Asn_{19}}$-Gly_{20}-Thr_{21}-Val_{22}-Tyr_{23}-Tyr_{24}-Phe_{25}-Asn_{26}-His_{27}-Ile_{28}-Thr_{29}-Asn_{30}-Ala_{31}-Ser_{32}-Gln_{33}-Phe_{34}-Glu_{35}-Arg_{36}-Pro_{37}-Ser_{38}-Gly_{39}.

The amino acid at position 16 was mutated to other amino acids as indicated in the text. Asn_{19} (underlined) was modified with an N-linked *N*-acetyl glucosamine (GlcNAc) residue in variants in which the compound number is followed by the letter "g". Thus, **5-1g** and **6-1g** have an Asn(GlcNAc) residue at position 19. Mass spectral characterization of the Pin WW variants is presented in the Supporting Information.

Thermal denaturation profiles were obtained by monitoring the ellipticity signal at 227 nm from 0.2 to 108.2 °C (for variants of **5-1g**) or from 0.2 to 98.2 °C (for variants of **6-1g**) 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.[4,29] Although the heat capacities of folding ($\Delta C_{p,fold}$) were used as an adjustable parameter in these fits, the error in their estimated values is large, making calculations of folding free energies at temperatures far from the midpoint temperature (T_m) of the thermal denaturation curve unreliable. We compensated for this by deriving a consensus value of the heat capacity of folding for all the glycosylated and non-glycosylated Pin WW variants from the slope of a plot of their enthalpies of folding at T_m (ΔH_m) versus their T_m values (since $d\Delta H_m/dT_m = \Delta C_{p,fold}$; Supporting Information Figure S1).[31,32] This approach is justified by the finding that heat capacities of folding are primarily dependent on protein size, not sequence, and should therefore change little for variants of a particular protein[33,34] and by the observation that the ΔH_m values for all the Pin WW variants in this study fall on the same line (Figure S1). The consensus value for $\Delta C_{p,fold}$ was −0.41 ± 0.02 kcal mol^{-1} K^{-1}, which was used for calculating temperature-dependent free energies for all Pin WW variants.

Determination of NMR Solution Structures of 5-1, 5-1g, 6-1, and 6-1g. ^{1}H 2D NMR spectra were recorded on Bruker DRX600 or DMX750 spectrometers at 293 K. Interproton distance restraints were derived from 2D ^{1}H NOESY spectra acquired with a 100 ms mixing time. To facilitate NMR assignments, we synthesized variants of **5-1**, **5-1g**, **6-1**, and **6-1g**, in which ^{13}C and ^{15}N isotopes were incorporated into loop 1 by solid-phase peptide synthesis, using the following isotopically labeled Fmoc-protected amino acids (Sigma-Aldrich): uniformly ^{13}C- and ^{15}N-labeled Fmoc-Phe-OH, Fmoc-Arg(Pbf)-OH, and Fmoc-Thr(O*t*Bu)-OH; and uniformly ^{15}N-labeled Fmoc-Ala-OH, Fmoc-Asn(Trt)-OH, and Fmoc-Gly-OH. Structures were calculated in AMBER[35] using preliminary coordinates derived from the published structure of the Pin WW domain.[36] Variant primary structures of loop 1 were built into the Pin WW structure for **5-1**, **5-1g**, **6-1**, and **6-1g** using NAB.[37] Details of NMR restraints and structure statistics are included in Table S1. The lowest energy structures (20 each for **5-1** and **5-1g**, 19 for **6-1**, and 17 for **6-1g**) were analyzed using PROCHECK-NMR.[38] The Ramachandran statistics for **5-1** are as follows: most favored regions, 81.3%; additionally allowed regions, 17.1%; generously allowed regions, 0.6%; disallowed regions, 1.0%. For **5-1g**, the corresponding values are 80.0, 18.8, 0.6, and 0.6%. For **6-1**, the Ramachandran statistics are as follows: most favored regions, 71.5%; additionally allowed regions, 25.6%; generously allowed regions, 2.6%; disallowed regions, 0.4%; and for **6-1g**, the corresponding values are 72.2, 23.0, 3.5, and 1.3%. All structures have been deposited in the RCSB Protein Data Bank. Their accession codes are **5-1**, 2M9E; **5-1g**, 2M9F; **6-1**, 2M9I; **6-1g**, 2M9J.

RESULTS

NMR Structures of Pin WW Variants with Enhanced Aromatic Sequons. To characterize the protein−carbohydrate interaction geometry and to ensure that glycosylation does not change the conformation of Pin WW, we determined the NMR solution structures of Pin WW variants with five- (**5-1g**) and six-residue (**6-1g**) enhanced aromatic sequons glycosylated with a single GlcNAc on Asn_{19} and compared these to the structures of their nonglycosylated counterparts (**5-1** and **6-1**, Figure 1 and Table S1). The backbone structures of nonglycosylated **5-1** and glycosylated **5-1g** are very similar (Figure 1a,b), as are those of **6-1** and **6-1g** (Figure 1c,d). The main differences are in the conformations of the side chains near the glycosylation site. In nonglycosylated **5-1** and **6-1**, the Phe_{16} side chain prefers a *trans* conformation ($\chi_1 \approx 180°$), while the Asn_{19} side chain is split between *gauche*$^-$ ($\chi_1 \approx -60°$) and *trans* (Figure 1e,h). In glycosylated **5-1g** and **6-1g**, the Phe_{16} and Asn_{19} side chains rotate into the *gauche*$^-$ ($\chi_1 \approx -60°$) and *gauche*$^+$ ($\chi_1 \approx 60°$) conformations, respectively, bringing the α-face of the GlcNAc into contact with the face of the Phe_{16} aromatic ring (Figure 1f,g and Figure 1i,j).

The planes of the phenyl ring of Phe_{16} and the GlcNAc on Asn_{19} in **5-1g** are almost parallel; the angle between them is 13.9 ± 0.5° (mean ± standard error), and they are separated by 3.97 ± 0.01 Å (Figure 1f,g). The rings are offset by 1.51 ± 0.03 Å, making H5 the closest GlcNAc atom to the Phe_{16} phenyl ring (2.40 ± 0.01 Å above and 0.78 ± 0.02 Å offset from the center of the Phe_{16}; Figure 1f). Consistent with the calculated structure, the H5 resonance is upfield shifted by about 1.2 ppm (relative to a published reference[6]) by the Phe_{16} ring current (Figure 1f). The other protons on the α-face of GlcNAc (H1 and H3) are also shifted but to a lesser degree. The separation between the GlcNAc and Phe_{16} ring planes is slightly greater in **6-1g** than in **5-1g** (4.11 ± 0.02 Å), and the angle between them is greater as well (21.5 ± 0.3°). Thus, despite the slightly greater separation between the rings, H5 of GlcNAc is as close to the Phe_{16} ring in **6-1g** (2.40 ± 0.02 Å) as it is in **5-1g** (Figure 1i,j). Indeed, we find that H5 is also upfield shifted in **6-1g** (by 1.7 ppm) as are the other protons on the α-face of the GlcNAc (H1 and H3) (Figure 1i).

Components of the Intramolecular Protein−Carbohydrate Interaction Energy. The interaction energy for protein−carbohydrate molecular recognition in glycosylated Pin WW variants is equivalent to the native state stabilization energy due to glycosylation of enhanced aromatic sequons, or $\Delta\Delta G_{glyc}$. $\Delta\Delta G_{glyc}$ can be divided into three components. The first is the intrinsic effect of glycosylating Asn_{19} (ΔG_i), which probably reflects the conformational preferences of Asn_{19} versus $Asn(GlcNAc)_{19}$ and the interaction of the GlcNAc with parts of the protein other than position 16. This is the only component of $\Delta\Delta G_{glyc}$ that is likely to be different for Pin WW variants with five- versus six-residue enhanced aromatic sequons because the φ and ψ angles for Asn_{19} are different in these two structural contexts ($\varphi = -80.0 \pm 2.2°$, $\psi = 0.9 \pm 1.4°$ in **5-1g** vs $\varphi = -139.3 \pm 0.9°$, $\psi = -19.7 \pm 0.5°$ in **6-1g**). The second component of $\Delta\Delta G_{glyc}$ is due to hydrophobic burial of the α-face of the GlcNAc (which is largely nonpolar[20,39]) and the side chain at position 16 (ΔG_{phob}). In proteins, such hydrophobic burial energies are generally found to correlate with the water−octanol transfer free energies (ΔG_{tr}, a measure of hydrophobicity)[40] of appropriate analogues of the side chains.[41−43] Thus, ΔG_{phob} should be proportional to ΔG_{tr} for the position

dx.doi.org/10.1021/ja4040472 | J. Am. Chem. Soc. 2013, 135, 9877−9884

Table 1. Data for Pin WW Variants with Five- or Six-Residue Enhanced Aromatic Sequons in Loop 1 with Sequences Xxx_{16}-Ala_{18}-Asn_{19}-Gly_{20}-Thr_{21} or Xxx_{16}-Arg_{17}-Ser_{18}-Asn_{19}-Gly_{20}-Thr_{21}, Respectively, in Which Xxx_{16} Is Varied[a]

				Five-residue enhanced aromatic sequon (energies reported at 60 °C)		Six-residue enhanced aromatic sequon (energies reported at 60 °C)	
Xxx_{16}[b]	ΔG_{tr}[c]	σ_m	σ_p	#[d]	$\Delta\Delta G_{glyc}$	#[d]	$\Delta\Delta G_{glyc}$
Phe	−3.44	0	0	**5-1, 5-1g**	−0.86 ± 0.04	**6-1, 6-1g**	−0.80 ± 0.06
Ala	−1.49	NA	NA	**5-2, 5-2g**	−0.33 ± 0.05	**6-2, 6-2g**	−0.30 ± 0.01
Ser	0.37	NA	NA	**5-3, 5-3g**	−0.21 ± 0.10	**6-3, 6-3g**	0.09 ± 0.07
Thr	−0.10	NA	NA	**5-4, 5-4g**	−0.21 ± 0.05	**6-4, 6-4g**	0.17 ± 0.07
Abu	−1.81	NA	NA	**NA**	NA	**6-5, 6-5g**	−0.17 ± 0.08
Nva	−2.39	NA	NA	**NA**	NA	**6-6, 6-6g**	−0.25 ± 0.07
Leu	−2.84	NA	NA	**5-7, 5-7g**	−0.51 ± 0.04	**6-7, 6-7g**	−0.12 ± 0.05
Met	−1.57	NA	NA	**5-8, 5-8g**	−0.37 ± 0.07	**6-8, 6-8g**	−0.08 ± 0.03
Val	−2.39	NA	NA	**5-9, 5-9g**	−0.42 ± 0.06	**6-9, 6-9g**	−0.25 ± 0.03
Ile	−2.96	NA	NA	**5-10, 5-10g**	−0.28 ± 0.02	**6-10, 6-10g**	−0.46 ± 0.02
Cha	−3.86	NA	NA	**5-11, 5-11g**	−0.63 ± 0.09	**6-11, 6-11g**	−0.14 ± 0.05
Tyr	−2.90	-0.37	0.12	**5-12, 5-12g**	−0.79 ± 0.06	**6-12, 6-12g**	−1.04 ± 0.05
Trp	−2.81	NA	NA	**5-13, 5-13g**	−0.89 ± 0.03	**6-13, 6-13g**	−0.83 ± 0.05
His	−0.05	NA	NA	**5-14, 5-14g**	−0.67 ± 0.03	**6-14, 6-14g**	−0.40 ± 0.06
pNH_2-Phe	−1.95	-0.66	-0.16	**5-15, 5-15g**	−1.16 ± 0.07	**6-15, 6-15g**	−0.76 ± 0.05
*p*OMe-Phe	−3.74	-0.27	0.12	**5-16, 5-16g**	−1.05 ± 0.07	**6-16, 6-16g**	−0.93 ± 0.03
*p*COOH-Phe	−2.24	0.00	-0.10	**5-17, 5-17g**	−0.96 ± 0.01	**6-17, 6-17g**	−0.45 ± 0.04
*p*Cl-Phe	−4.20	0.23	0.37	**5-18, 5-18g**	−1.09 ± 0.04	**6-18, 6-18g**	−0.93 ± 0.04
*p*F-Phe	−3.65	0.06	0.34	**5-19, 5-19g**	−0.90 ± 0.04	**6-19, 6-19g**	−0.78 ± 0.06
*p*CN-Phe	−3.48	0.66	0.56	**5-20, 5-20g**	−1.16 ± 0.05	**6-20, 6-20g**	−0.79 ± 0.03
pNO_2-Phe	−3.05	0.78	0.71	**5-21, 5-21g**	−0.63 ± 0.03	**6-21, 6-21g**	−0.67 ± 0.02
*p*Br-Phe	−4.57	0.23	0.39	**NA**	NA	**6-22, 6-22g**	−0.94 ± 0.07
pCF_3-Phe	−4.69	0.54	0.43	**NA**	NA	**6-23, 6-23g**	−1.00 ± 0.07
3,4,5 F_3-Phe	−4.09	NA	NA	**5-24, 5-24g**	−0.66 ± 0.05	**6-24, 6-24g**	−0.70 ± 0.02

[a]Entries for aromatics and nonaromatics are shaded light orange and light blue, respectively. All energies are reported in kcal mol^{-1} ± standard errors. $\Delta\Delta G_{glyc}$ values are reported at 60 °C (data at other temperatures are in Table S2). The folding free energies ($\Delta G_{fold,glyc}$ and $\Delta G_{fold,nonglyc}$) and other parameters for the individual Pin WW variants on which $\Delta\Delta G_{glyc}$ are based are reported in Table S2. [b]Abu = α-aminobutyric acid; Nva = norvaline; Cha = cyclohexylalanine; 3,4,5-F_3-Phe = 3,4,5-trifluorophenylalanine. Position 16 variants denoted as "*p*X-Phe" are derivatives of Phe with the functional group X in the *para* position of the Phe ring. [c]ΔG_{tr} is the water-to-octanol transfer free energy for an appropriate side chain analogue (Table S2) calculated using Crippen's fragmentation[40] as implemented in ChemBioDraw Ultra 12.0. [d]Pin WW variants with a "**g**" following the number are glycosylated on Asn_{19} with a single GlcNAc.

16 side chains. The third component of $\Delta\Delta G_{glyc}$ derives from the CH−π interaction between the axial hydrogens projecting from the α-face of GlcNAc (especially H5) and the aromatic amino acid side chain at position 16, which is driven by a combination of dispersion and electrostatic forces[44,45] ($\Delta G_{CH-\pi}$). Dispersion forces refer to interactions between instantaneous dipoles and multipoles that arise from complementary polarization of the interacting species' electron charge distributions[46] and as such occur between all species. $\Delta G_{CH-\pi}$ reflects the stronger dispersion interaction of GlcNAc with aromatic amino acids at position 16 compared to nonaromatic amino acids. Electrostatic contributions to $\Delta G_{CH-\pi}$ would be due to interactions between permanent dipoles or multipoles on GlcNAc and the aromatic side chain at position 16 and should therefore be related to the electrostatic properties of the aromatic ring.

Quantifying the Components of the Protein−Carbohydrate Interaction Energy. To quantify the various components of $\Delta\Delta G_{glyc}$, we synthesized 88 glycosylated and nonglycosylated variants of **5-1g** and **6-1g** (44 pairs) with an array of natural and unnatural amino acids at position 16, while keeping the rest of the glycoprotein structure constant. We determined the folding free energy for each variant by thermal denaturation and calculated the change in folding free energy upon glycosylation ($\Delta\Delta G_{glyc}$) by subtracting the folding free energy of each nonglycosylated variant ($\Delta G_{fold,nonglyc}$) from that of the corresponding glycosylated variant ($\Delta G_{fold,glyc}$; i.e., $\Delta\Delta G_{glyc} = \Delta G_{fold,glyc} - \Delta G_{fold,nonglyc}$). The $\Delta\Delta G_{glyc}$ values for these variants at 60 °C are presented in Table 1 (60 °C was chosen because it is close to the average thermal transition midpoints (T_m) of all variants; data at other temperatures, at which the overall conclusions still hold, are reported in Table S2).

Quantifying Energetic Contributions from the Hydrophobic Effect. Linear regression of the $\Delta\Delta G_{glyc}$ values of the **5-1g** variants harboring nonaromatic side chains at position 16

dx.doi.org/10.1021/ja4040472 | J. Am. Chem. Soc. 2013, 135, 9877−9884

versus calculated water−octanol transfer free energies for appropriate analogues of the position 16 side chains yields a moderately good fit (R^2 = 0.69, F = 13.4, p = 0.01) with a slope of 0.08 ± 0.02 and an intercept of −0.21 ± 0.05 kcal mol^{-1} (Figure 2a, yellow circles). The fit is not as good for the

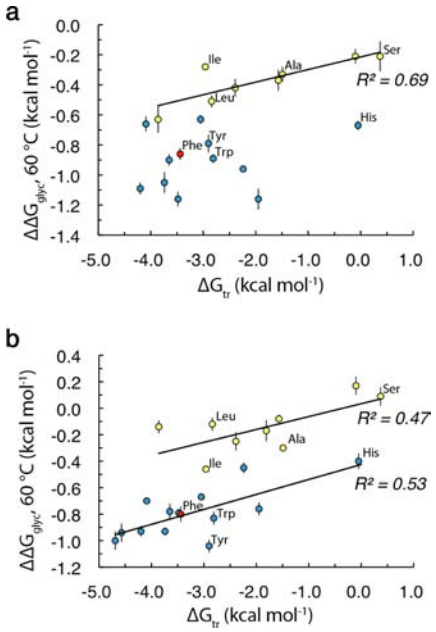


Figure 2. (a) Plot of the values of $\Delta\Delta G_{glyc}$ vs the calculated water-to-octanol transfer free energies (ΔG_{tr}) of the appropriate side chain analogues (Table 1 and Table S2) for Pin WW variants with nonaromatic or aromatic amino acids at position 16 in a five-residue enhanced aromatic sequon (variants **5-1g** through **5-24g**). The values of $\Delta\Delta G_{glyc}$ when nonaromatic amino acids, Phe, and other aromatic amino acids are in position 16 are shown in yellow, red, and blue, respectively. The line of best fit (straight line) has R^2 = 0.69, F = 13.4, p = 0.01, slope = 0.08 ± 0.02, and intercept = −0.21 ± 0.05 kcal mol^{-1}. (b) Same as panel a, except for Pin WW variants with a six-residue enhanced aromatic sequon. The line of best fit (straight line) for nonaromatic residues has R^2 = 0.47, F = 7.0, p = 0.03, slope = 0.10 ± 0.04, and intercept = 0.03 ± 0.08 kcal mol^{-1}. The line of best fit (straight line) for aromatic residues has R^2 = 0.53, F = 13.7, p = 0.003, slope = 0.11 ± 0.03, and intercept = −0.42 ± 0.10 kcal mol^{-1}. All reported errors are standard errors of the coefficients. Selected amino acids are labeled in panels a and b.

corresponding variants of **6-1g** with nonaromatic amino acids at position 16 (Figure 2b, yellow circles, R^2 = 0.47, F = 7.0, p = 0.03). The slope and intercept of the regression line are 0.10 ± 0.04 and 0.03 ± 0.08 kcal mol^{-1}, respectively. Nonetheless, the trends for the variants of **5-1g** and **6-1g** are the same: $\Delta\Delta G_{glyc}$ increasing as ΔG_{tr} increases, with similar slopes. The intercepts indicate the intrinsic effect of glycosylating Asn_{19} in a five- or six-residue enhanced aromatic sequon (ΔG_i), independent of any interactions with the side chain at position 16. Glycosylating the Asn is more stabilizing in five- than in six-residue enhanced aromatic sequons (−0.21 ± 0.05 vs 0.03 ± 0.08 kcal mol^{-1}), consistent with our previous findings.[29]

$\Delta\Delta G_{glyc}$ values of the **5-1g** variants harboring aromatic side chains show no correlation with ΔG_{tr} (Figure 2a, blue and red circles). In contrast, the $\Delta\Delta G_{glyc}$ values of the **6-1g** variants harboring aromatic side chains show a fair correlation with ΔG_{tr} (Figure 2b, blue and red circles, R^2 = 0.53, F = 13.7, p = 0.003) with a slope of 0.11 ± 0.03 and an intercept of −0.42 ± 0.10 kcal mol^{-1}. Thus, three out of the four cases examined yield statistically significant correlations between $\Delta\Delta G_{glyc}$ and ΔG_{tr}, indicating that the hydrophobic effect generally makes a substantial contribution to protein−carbohydrate interactions in both five- and six-residue enhanced aromatic sequons. In the remaining case, aromatic variants of **5-1g**, the combination of experimental error, and the innate variability of the protein−carbohydrate interaction make direct quantification of the hydrophobic effect difficult. However, it is worth noting that the contribution of the hydrophobic effect to the protein−carbohydrate interactions in the nonaromatic variants of **5-1g** and the aromatic and nonaromatic variants of **6-1g** are all similar, as indicated by the similar slopes of their regression lines between $\Delta\Delta G_{glyc}$ versus ΔG_{tr} (0.08 ± 0.02, 0.10 ± 0.04, and 0.11 ± 0.03, respectively; Figure 2a,b). We therefore expect the contribution of the hydrophobic effect to the protein−carbohydrate interactions in aromatic variants of **5-1g** to be similar to that in the other three cases.

The slopes of the correlations between $\Delta\Delta G_{glyc}$ and ΔG_{tr} reflect the strength of the hydrophobic interaction between the α-face of the GlcNAc and the side chains at position 16 relative to transfer from water to octanol. It has been shown that the effects of mutations on protein folding free energies can correlate with ΔG_{tr} with slopes close to 1 for fully buried positions.[41−43] Our smaller slopes, generally ∼0.1 for the cases where there is a demonstrable correlation between $\Delta\Delta G_{glyc}$ and ΔG_{tr}, suggest that burying the position 16 side chain under the α-face of GlcNAc is not as energetically favorable, likely reflecting the difficulty of burying the nonpolar α-face of GlcNAc without the destabilizing side effect of burying adjacent polar hydroxyls.[19,47]

Quantifying the Energetic Contributions from CH−π Interactions. The values of $\Delta\Delta G_{glyc}$ predicted by the regression lines of nonaromatic amino acids in Figure 2a,b and the ΔG_{tr} of the Phe side chain (−3.4 kcal mol^{-1}) are −0.48 ± 0.08 kcal mol^{-1} for **5-1g** and −0.31 ± 0.16 kcal mol^{-1} for **6-1g**. However, the experimental values of $\Delta\Delta G_{glyc}$ for the Phe variants are −0.86 ± 0.04 and −0.80 ± 0.06 kcal mol^{-1}, respectively (red circles; Figure 2a,b), much larger than expected from consideration of the hydrophobic effect alone. The differences between the predicted and measured values (−0.38 ± 0.09 kcal mol^{-1} for **5-1g** and −0.49 ± 0.17 kcal mol^{-1} for **6-1g**) represent the energies resulting from the special nature of the GlcNAc−aromatic side chain interaction in the five- and six-residue enhanced aromatic sequons. The proximity of H5 of GlcNAc and its neighboring axial hydrogens to the aromatic ring of Phe_{16} in both **5-1g** and **6-1g** suggests that those differences are due to CH−π interactions.[44] To further probe the electrostatic contribution to the CH−π interaction, we attempted to fit the $\Delta\Delta G_{glyc}$ values of Phe, Tyr, and additional *para*-substituted position 16 aromatic variants to a Hammett-type linear free energy relationship:

$$\Delta\Delta G_{glyc} = Ax + B$$

where B is a constant, x is a variable that measures a substituent's electronic effect on the aromatic ring, and A is the sensitivity of $\Delta\Delta G_{glyc}$ to x.[48] Of the various x parameters available in the literature, we used Hammett's σ_m and σ_p.[48]

The values of $\Delta\Delta G_{glyc}$ in Table 1 for variants of **5-1g** are fit poorly by the Hammett relationship with σ_m and σ_p (R^2 = 0.04 and 0.07, respectively; Figure 3a,b). The values of $\Delta\Delta G_{glyc}$ for variants of **6-1g** are also poorly fit by the Hammett relationship with σ_m and σ_p (R^2 = 0.14 and 0.01, respectively; Figure 3c,d).

dx.doi.org/10.1021/ja4040472 | J. Am. Chem. Soc. 2013, 135, 9877−9884

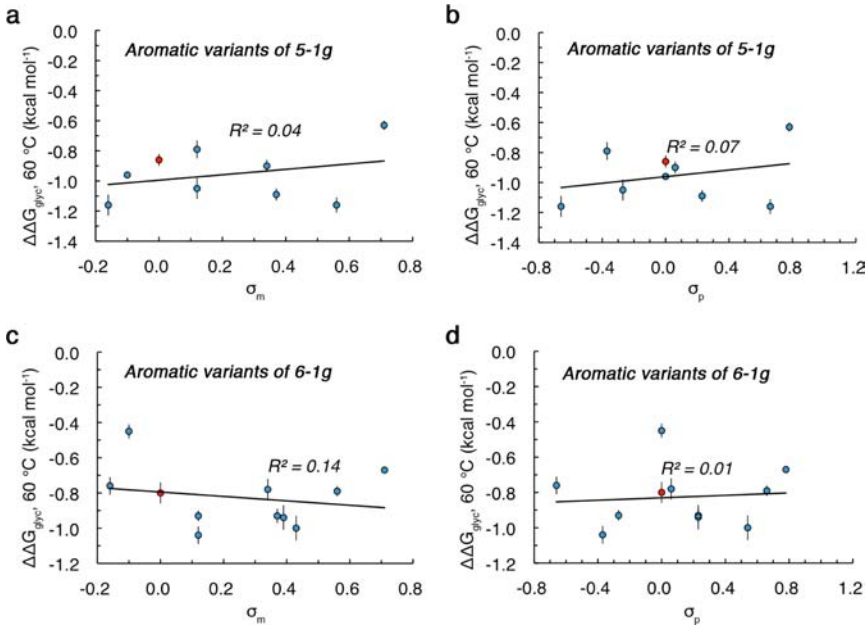


Figure 3. (a,b) Plots of $\Delta\Delta G_{glyc}$ for variants of **5-1g** with *para*-substituted aromatic amino acids at position 16 vs σ_m and σ_p, respectively. The lines of best fit (straight lines) have $R^2 = 0.04$ and 0.07, respectively. (c,d) Plots of $\Delta\Delta G_{glyc}$ for variants of **6-1g** with *para*-substituted aromatic amino acids at position 16 vs σ_m and σ_p, respectively. The lines of best fit (straight lines) have $R^2 = 0.14$ and 0.01, respectively. Values of $\Delta\Delta G_{glyc}$ for **5-1g** and **6-1g** are shown in red.

This lack of dependence of $\Delta\Delta G_{glyc}$ on the electronic properties of the substituents suggests that interactions involving partial charges or permanent dipoles (i.e., electrostatic interactions) are not measurable in either the five- or the six-residue enhanced aromatic sequon.

This result suggests, by a process of elimination, that the carbohydrate−aromatic interaction in the five- and six-residue enhanced aromatic sequons must be driven by dispersion forces.[46] Several quantum mechanical studies of carbohydrate−aromatic and other CH−π interactions in the gas phase have suggested that, while both electrostatics and dispersion forces contribute to these interactions, dispersion forces contribute more than electrostatics.[24−27,44,45] Our results suggest that the importance of dispersion forces in carbohydrate−aromatic interactions persists in aqueous solution. However, the contribution from the electrostatic component to the net energy of CH−π interactions relative to that from the dispersion component may be even smaller in aqueous solution than in the gas phase because any changes in the strength of CH−π electrostatic interactions due to changes in the electron density of the aromatic ring are envisioned to be canceled out by parallel changes in the strength of electrostatic interactions between water and the aromatic ring in the denatured state.

DISCUSSION

Three factors contribute to the stabilizing glycan−protein interactions in five- and six-residue enhanced aromatic sequons, quantified by $\Delta\Delta G_{glyc}$: the intrinsic effect of glycosylating Asn_{19} (ΔG_i); the attenuated hydrophobic effect derived from juxtaposing the side chain at position 16 with the α-face of the GlcNAc (ΔG_{phob}); and the CH−π interaction when an aromatic amino acid is at position 16 ($\Delta G_{CH-\pi}$). The additional stabilization derived from the aromatic ring−GlcNAc interaction is mostly due to dispersion forces in the five- and six-residue enhanced aromatic sequons. The estimated contributions of each of these components to $\Delta\Delta G_{glyc}$ when position 16 is Leu, Phe, or Tyr are illustrated in Figure 4 for the five- and six-residue enhanced aromatic sequons.

The quantification of the structure−energy relationships presented herein for protein−carbohydrate molecular recognition was made possible by applying a powerful tool from classical physical organic chemistry (linear free energy relationships) to an extensive set of protein−carbohydrate interaction energies measured within the structurally well-defined context of enhanced aromatic sequons. There are, however, two caveats to the quantification. First, introducing a mutation into a protein that perturbs an interaction of interest may result in a structural reorganization of the native state. The overall effect of the mutation on the folding free energy is then the sum of the change in the interaction energy due to the mutation and the reorganization energy if there are any structural changes. The reorganization energy is always negative (i.e., thermodynamically favorable), so the effect of the mutation on the folding free energy must be considered a lower bound on the interaction energy. In our case, however, we expect the reorganization energies to be minimal because the loop 1 structures of a given class are quite robust, as illustrated by Figure 5. Alignment of the main chain in loop 1 of the lowest energy calculated structure of **5-1g** (sequence: Met_{15}-Phe_{16}-Ala_{18}-Asn_{19}-Gly_{20}-Thr_{21}-Val_{22}) with that of a Pin WW variant (PDB code: 2F21)[49] with the same loop type (a type 1 β-turn with a G1 β-bulge), but a different sequence (Met_{15}-Ser_{16}-Ala_{18}-Asp_{19}-Gly_{20}-Arg_{21}-Val_{22}) shows that the three amino acid differences have little effect on the loop structure (Figure 5a). Analogously, alignment of the main chain in loop 1 of the lowest energy calculated structure of **6-1g** (sequence: Met_{15}-Phe_{16}-Arg_{17}-Ser_{18}-Asn_{19}-Gly_{20}-Thr_{21}-Val_{22}) with that of wild-type Pin WW (PDB code: 1PIN),[50] which has the same loop type (a type 2 β-turn within a six-residue loop) but a different sequence (Met_{15}-Ser_{16}-Arg_{17}-Ser_{18}-Ser_{19}-Gly_{20}-Arg_{21}-Val_{22}), demonstrates only slight perturbations of the structure (Figure 5b).

dx.doi.org/10.1021/ja4040472 | J. Am. Chem. Soc. 2013, 135, 9877−9884

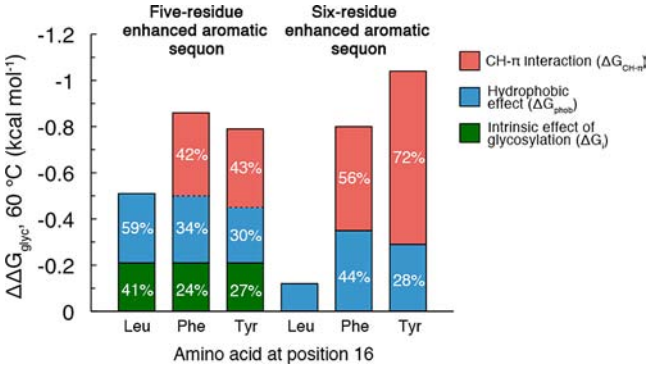


Figure 4. Estimated composition of components of $\Delta\Delta G_{glyc}$ for **5-1g** and **6-1g** variants. Values of $\Delta\Delta G_{glyc}$ for five- and six-residue enhanced aromatic sequons with Leu, Phe, or Tyr at position 16 were broken down into three components: (1) the inherent effect of glycosylating Asn_{19} (ΔG_i); (2) the energy of hydrophobic burial of the Asn_{19} GlcNAc and the position 16 side chain (ΔG_{phob}); and (3) the energy of the CH$-\pi$ interaction between the GlcNAc and aromatic amino acid side chains at position 16 ($\Delta G_{CH-\pi}$). The first component is the intercept of the best-fit lines for nonaromatic amino acids in Figure 2a,b for five- or six-residue enhanced aromatic sequons, respectively. For Leu, the second component is calculated by subtracting the first component from $\Delta\Delta G_{glyc}$. For Phe and Tyr, it is calculated by multiplying their ΔG_{tr} (Table 1) by the slope of the best-fit lines for nonaromatic amino acids in Figure 2a,b for five- or six-residue enhanced aromatic sequons, respectively. The third component, $\Delta G_{CH-\pi}$, is calculated by subtracting the first and second components from $\Delta\Delta G_{glyc}$. Although, as discussed in the main text, experimental error and the innate variability of the protein–carbohydrate interaction energetics prevent precise determination of the contribution of the hydrophobic effect to $\Delta\Delta G_{glyc}$ for aromatic variants of **5-1g**, we expect the hydrophobic effect to be roughly the same as for nonaromatic variants of **5-1g**. This uncertainty is indicated by the dashed line separating the hydrophobic and CH$-\pi$ contributions in the bars for the cases where Phe or Tyr occupy position 16 in the five-residue enhanced aromatic sequon.

Second, we are employing the intramolecular protein–carbohydrate interactions in our Pin WW variants to ascertain features of protein–carbohydrate interactions in general, both intra- and intermolecular, which may have different geometries. The characteristics of protein–carbohydrate interactions are likely to depend on geometry. However, the facial interaction between a carbohydrate and the position 16 amino acid side chain in enhanced aromatic sequons is a common type of protein–carbohydrate interaction, especially for aromatic amino acid side chains.[20,21,51] Thus, the behavior of the protein–carbohydrate interaction in the enhanced aromatic sequon should be representative of a broad class of protein–carbohydrate interactions. Furthermore, while the intramolecular protein–carbohydrate interaction in enhanced aromatic sequons cannot replicate certain aspects of intermolecular protein–carbohydrate interactions, in particular, the effect of the loss of translational entropy, the attractive forces that drive protein–carbohydrate interactions should be the same whether the interaction is intra- or intermolecular.

Figure 4 also shows that, although contributing significantly to protein–carbohydrate packing, the hydrophobic effect cannot drive strong protein–carbohydrate interactions the way it drives protein folding or the binding of other types of ligands; the contributions from the hydrophobic effect are never more than 44% of $\Delta\Delta G_{glyc}$ for **5-1g** or **6-1g**. Thus, tight binding of carbohydrates requires either multivalency[52] or a

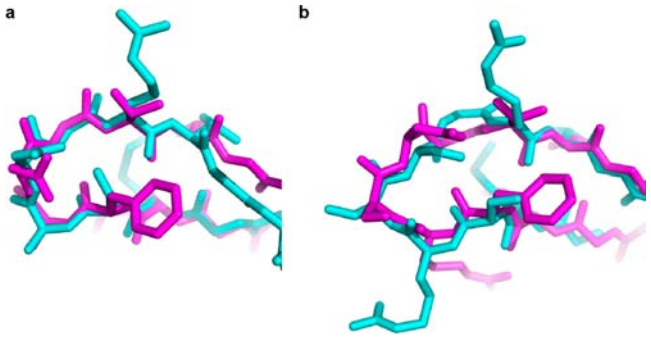


Figure 5. (a) Alignment of the main chain in the loop 1 region from the lowest energy calculated structure of **5-1g** (magenta; sequence: Met_{15}-Phe_{16}-Ala_{18}-Asn(GlcNAc)$_{19}$-Gly_{20}-Thr_{21}-Val_{22}; the GlcNAc is not shown for clarity; PDB code: 2M9F) with that of a Pin WW variant, the structure of which was determined previously by X-ray crystallography (cyan; PDB code: 2F21)[49] that has the same loop type (a type 1 β-turn with a G1 β-bulge) but three different amino acids in its sequence (Met_{15}-Ser_{16}-Ala_{18}-Asp_{19}-Gly_{20}-Arg_{21}-Val_{22}). Main chain RMSD = 0.6 Å for loop 1. (b) Alignment of the main chain in the loop 1 region from the lowest energy calculated structure of **6-1g** (magenta; sequence: Met_{15}-Phe_{16}-Arg_{17}-Ser_{18}-Asn(GlcNAc)$_{19}$-Gly_{20}-Thr_{21}-Val_{22}; the GlcNAc is not shown for clarity; PDB code: 2M9J) with that of wild-type Pin WW (cyan; PDB code: 1PIN),[50] which has the same loop type (a type 2 β-turn within a six-residue loop) but three different amino acids in its sequence (Met_{15}-Ser_{16}-Arg_{17}-Ser_{18}-Ser_{19}-Gly_{20}-Arg_{21}-Val_{22}). Main chain RMSD = 0.9 Å for loop 1. Images of the structures were made, and RMSD values were calculated, using PyMOL.

protein–carbohydrate binding site rich in desolvated hydrogen bonds (the energetics of which were not addressed in this study) and aromatic amino acids.[20,23] The arabinose binding protein of *Escherichia coli* is a naturally occurring example of these strategies.[53]

CONCLUSIONS

We have utilized a robust protein scaffold to study the molecular origin of energetically favorable carbohydrate–aromatic packing interactions in aqueous solution. Both the hydrophobic effect and CH$-\pi$ interactions are found to contribute significantly to the overall interaction energy. The observation that the electronic nature of aromatic groups has little effect in dictating the strength of carbohydrate–aromatic interactions in aqueous solution suggests that the electrostatic component of CH$-\pi$ interaction energies is likely small relative to the dispersion component, possibly for reasons outlined at the end of the Results section.

The structure–energy relationships presented herein provide a framework for understanding the energetic basis of protein–carbohydrate molecular recognition in aqueous solution (as opposed to the computationally more accessible gas phase) and will serve as benchmarks for the molecular mechanics force fields used in simulations of systems with intra- and intermolecular protein–carbohydrate interactions.[54] Improving these force fields will in turn improve our ability to design and optimize protein–carbohydrate interactions, facilitating the development of industrial and therapeutic applications that rely on these interactions.

ASSOCIATED CONTENT

Supporting Information

Plot of ΔH_m vs T_m for the determination of $\Delta C_{p,fold}$ for the Pin WW variants, NMR, and refinement statistics and thermody-

dx.doi.org/10.1021/ja4040472 | J. Am. Chem. Soc. 2013, 135, 9877–9884

namics data from fitting CD-monitored thermal denaturation data of Pin WW variants. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

jkelly@scripps.edu; dyson@scripps.edu; chwong@gate.sinica.edu.tw; epowers@scripps.edu

Present Addresses

▲S.E.: BASF SE, Ludwigshafen D-67056, Germany.

△J.L.P.: Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT 84602.

Author Contributions

#W.C. and S.E. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Gira Bhabha (Scripps) and Dr. Gerard Kroon (Scripps) for help with the NMR experiments, Dr. Pritilekha Deka and Dr. Sangho Park (Scripps), Professor Robert J. Woods, and Dr. B. Lachele Foley (University of Georgia, Complex Carbohydrate Research Center) for help with the structure calculations, and Dr. Sarah R. Hanson, Dr. Elizabeth K. Culyba, and Dr. Amber N. Murray for helpful discussions. This work was supported by the Skaggs Institute for Chemical Biology, the Lita Annenberg Hazen Foundation, and by National Institutes of Health Grants GM051105 (J.W.K.), GM071862 (H.J.D.), and AI072155 (C.-H.W.). J.L.P. was supported in part by NIH postdoctoral fellowship F32 GM086039.

REFERENCES

(1) Lee, Y. C.; Lee, R. T. *Acc. Chem. Res.* **1995**, *28*, 321.

(2) Lindhorst, T. K. In *Carbohydrate-Modifying Biocatalysts*; Grunwald, P., Ed.; Pan Stanford Publishing Pte. Ltd.: Singapore, 2011; p 119.

(3) Chen, M. M.; Bartlett, A. I.; Nerenberg, P. S.; Friel, C. T.; Hackenberger, C. P.; Stultz, C. M.; Radford, S. E.; Imperiali, B. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 22528.

(4) Culyba, E. K.; Price, J. L.; Hanson, S. R.; Dhar, A.; Wong, C. H.; Gruebele, M.; Powers, E. T.; Kelly, J. W. *Science* **2011**, *331*, 571.

(5) Wang, C.; Eufemi, M.; Turano, C.; Giartosio, A. *Biochemistry* **1996**, *35*, 7299.

(6) Wormald, M. R.; Wooten, E. W.; Bazzo, R.; Edge, C. J.; Feinstein, A.; Rademacher, T. W.; Dwek, R. A. *Eur. J. Biochem.* **1991**, *198*, 131.

(7) Helenius, A.; Aebi, M. *Annu. Rev. Biochem.* **2004**, *73*, 1019.

(8) Varki, A.; Etzler, M. E.; Cummings, R. D.; Esko, J. D. In *Essentials of Glycobiology*, 2nd ed.; Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E., Eds.; Cold Spring Harbor Laboratory Press: New York, 2009.

(9) van Kooyk, Y.; Rabinovich, G. A. *Nat. Immunol.* **2008**, *9*, 593.

(10) Beckham, G. T.; Dai, Z.; Matthews, J. F.; Momany, M.; Payne, C. M.; Adney, W. S.; Baker, S. E.; Himmel, M. E. *Curr. Opin. Biotechnol.* **2012**, *23*, 338.

(11) Ke, C.; Destecroix, H.; Crump, M. P.; Davis, A. P. *Nat. Chem.* **2012**, *4*, 718.

(12) Ernst, B.; Magnani, J. L. *Nat. Rev. Drug Discovery* **2009**, *8*, 661.

(13) Sola, R. J.; Griebenow, K. *J. Pharm. Sci.* **2009**, *98*, 1223.

(14) Laughrey, Z. R.; Kiehna, S. E.; Riemen, A. J.; Waters, M. L. *J. Am. Chem. Soc.* **2008**, *130*, 14625.

(15) Barwell, N. P.; Davis, A. P. *J. Org. Chem.* **2011**, *76*, 6548.

(16) Morales, J. C.; Reina, J. J.; Diaz, I.; Avino, A.; Nieto, P. M.; Eritja, R. *Chem.—Eur. J.* **2008**, *14*, 7828.

(17) Terraneo, G.; Potenza, D.; Canales, A.; Jimenez-Barbero, J.; Baldridge, K. K.; Bernardi, A. *J. Am. Chem. Soc.* **2007**, *129*, 2890.

(18) Vandenbussche, S.; Diaz, D.; Fernandez-Alonso, M. C.; Pan, W.; Vincent, S. P.; Cuevas, G.; Canada, F. J.; Jimenez-Barbero, J.; Bartik, K. *Chem.—Eur. J.* **2008**, *14*, 7570.

(19) Lemieux, R. U. *Acc. Chem. Res.* **1996**, *29*, 373.

(20) Quiocho, F. A. *Pure Appl. Chem.* **1989**, *61*, 1293.

(21) Weis, W. I.; Drickamer, K. *Annu. Rev. Biochem.* **1996**, *65*, 441.

(22) Pell, G.; Williamson, M. P.; Walters, C.; Du, H.; Gilbert, H. J.; Bolam, D. N. *Biochemistry* **2003**, *42*, 9316.

(23) Asensio, J. L.; Arda, A.; Canada, F. J.; Jimenez-Barbero, J. *Acc. Chem. Res.* **2013**, *46*, 946.

(24) Fernandez-Alonso, M. D.; Canada, F. J.; Jimenez-Barbero, J.; Cuevas, G. *J. Am. Chem. Soc.* **2005**, *127*, 7379.

(25) Kozmon, S.; Matuska, R.; Spiwok, V.; Koca, J. *Chem.—Eur. J.* **2011**, *17*, 5680.

(26) Raju, R. K.; Ramraj, A.; Vincent, M. A.; Hillier, I. H.; Burton, N. A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6500.

(27) Tsuzuki, S.; Uchimaru, T.; Mikami, M. *J. Phys. Chem. A* **2011**, *115*, 11256.

(28) Kornfeld, R.; Kornfeld, S. *Annu. Rev. Biochem.* **1985**, *54*, 631.

(29) Price, J. L.; Powers, D. L.; Powers, E. T.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 14127.

(30) Wells, P. R. *Chem. Rev.* **1963**, *63*, 171.

(31) Zaiss, K.; Jaenicke, R. *Biochemistry* **1999**, *38*, 4633.

(32) Deechongkit, S.; Kelly, J. W. *J. Am. Chem. Soc.* **2002**, *124*, 4980.

(33) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138.

(34) Robertson, A. D.; Murphy, K. P. *Chem. Rev.* **1997**, *97*, 1251.

(35) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

(36) Kowalski, J. A.; Liu, K.; Kelly, J. W. *Biopolymers* **2002**, *63*, 111.

(37) Macke, T. J.; Case, D. A. *ACS Symp. Ser.* **1998**, *682*, 379.

(38) Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. *J. Biomol. NMR* **1996**, *8*, 477.

(39) Stanca-Kaposta, E. C.; Gamblin, D. P.; Screen, J.; Liu, B.; Snoek, L. C.; Davis, B. G.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4444.

(40) Ghose, A. K.; Crippen, G. M. *J. Chem. Inf. Comp. Sci.* **1987**, *27*, 21.

(41) Bueno, M.; Campos, L. A.; Estrada, J.; Sancho, J. *Protein Sci.* **2006**, *15*, 1858.

(42) Matsumura, M.; Becktel, W. J.; Matthews, B. W. *Nature* **1988**, *334*, 406.

(43) Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. *Biochemistry* **1991**, *30*, 9686.

(44) Nishio, M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 13873.

(45) Salonen, L. M.; Ellermann, M.; Diederich, F. *Angew. Chem., Int. Ed.* **2011**, *50*, 4808.

(46) Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.* **1994**, *94*, 1723.

(47) Chavelas, E. A.; Garcia-Hernandez, E. *Biochem. J.* **2009**, *420*, 239.

(48) Hansch, C.; Leo, A.; Taft, R. W. *Chem. Rev.* **1991**, *91*, 165.

(49) Jager, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10648.

(50) Ranganathan, R.; Lu, K. P.; Hunter, T.; Noel, J. P. *Cell* **1997**, *89*, 875.

(51) Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *Biochem. J.* **2004**, *382*, 769.

(52) Bertozzi, C. R.; Kiessling, L. L. *Science* **2001**, *291*, 2357.

(53) Quiocho, F. A.; Vyas, N. K. *Nature* **1984**, *310*, 381.

(54) Fadda, E.; Woods, R. J. *Drug Discovery Today* **2010**, *15*, 596.