Evaluation of a carbohydrate–π interaction in a peptide model system†

Sarah E. Kiehna, Zachary R. Laughrey and Marcey L. Waters*

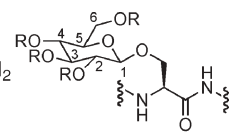
Received (in Austin, TX, USA) 25th July 2007, Accepted 19th August 2007
First published as an Advance Article on the web 28th August 2007
DOI: 10.1039/b711431a

A carbohydrate–π interaction contributes −0.8 kcal mol^{-1} to the stabilization of a β-hairpin peptide.

The binding of carbohydrates by proteins plays many important roles in biology, including recognition of bacterial cell walls, viral infection, and fertilization.[1] Thus, there is significant interest in understanding the driving force for affinity and selectivity in carbohydrate binding in water, both for the purpose of understanding protein–carbohydrate interactions and for the design of carbohydrate receptors.[2] In proteins, a common feature in carbohydrate binding sites is the interaction of the sugar with an aromatic sidechain;[3] for example, all of the galactose-specific lectins exhibit stacking of an aromatic sidechain with the α-face of the sugar.[4] Such carbohydrate–π interactions have been proposed to contribute to carbohydrate recognition and have been investigated through protein mutation studies,[5] NMR,[5c,6] IR,[7] and computationally.[5,8] In addition, aromatic rings are often incorporated into synthetic carbohydrate receptors to assist in binding.[9] Indeed, some have suggested that the carbohydrate–π interaction may be more important than hydrogen bonding in aqueous solution.[9e] However, there is limited experimental data demonstrating the favorable contribution of such a carbohydrate–π interaction to binding.[5c,9a,c,10] Given the challenges of carbohydrate recognition in water, a better understanding of the role of carbohydrate–π interactions in carbohydrate recognition is warranted.

To investigate the efficacy of an isolated carbohydrate–π interaction in aqueous solution, we incorporated tetraacetylglucoserine, Ser(Ac_4Glc), or glucoserine, Ser(Glc), in close proximity to a Trp residue on the face of a β-hairpin peptide (Fig. 1).[11] Ser(Ac_4Glc) was compared to Ser(Glc) to probe the role of desolvation in the carbohydrate–π interaction, as the acetyl groups reduce its hydrogen bonding ability as compared to Ser(Glc), and hence reduce its desolvation cost. We chose a β-hairpin model system since it has been shown to be useful for the investigation of noncovalent interactions in aqueous solution.[12] In this system, we found that the interaction of Trp with Ser(Ac_4Glc), but not with Ser(Glc), is significantly stabilizing. These findings have implications for the role of carbohydrate–π interactions in carbohydrate binding proteins and glycosylated proteins as well as in the design of carbohydrate receptors.

1: Ac-R**W**VTVNGK**S(Ac$_4$Glc)**ILQ-NH_2
2: Ac-R**W**VTVNGK**S(Glc)**ILQ-NH_2
3: Ac-R**W**VTVNGK**S**ILQ-NH_2



R = Ac: S(Ac_4Glc)
R = H: S(Glc)

Fig. 1 Peptide sequences and structure of Ser(R_4Glc).

A 12-residue peptide based on a previously reported system[12d,e] was used for the study of carbohydrate–π interactions (Fig. 1). Trp and Ser(Ac_4Glc), Ser(Glc), or serine (control) were placed in positions 2 and 9, respectively. These positions have been shown to allow for a sidechain–sidechain interaction and were expected to be able to accommodate the large glycosylated amino acid.[12d,e,13] Hairpin formation was promoted *via* an Asn–Gly Type I′ turn,[14] and Lys and Arg residues were included to provide water solubility and prevent aggregation.

NMR spectroscopy of the Trp and glucose sidechains provides insight into the nature and geometry of the interaction. The carbohydrate protons of peptide **1** exhibit considerable upfield shifting relative to random coil chemical shifts (Fig. 2a), indicating that they are in close proximity to the face of the aromatic ring of Trp.[15] Shifting is most significant on the α-face of the glucose moiety (H1, H3, and H5), up to 1.35 ppm at position H5, indicating a favorable interaction between the α-face of the sugar and the face of the aromatic ring.‡ Interestingly, H6 and H6′ were found to be upfield shifted to a similar extent as H3, suggesting that the exocyclic CH_2 group also contributes to the interaction, in a geometry such as that seen in galactose-binding lectins (Fig. 2b).[4] Although the individual acetyl groups could not be assigned, the maximum upfield shifting of those peaks was ⩽0.07 ppm, appreciably less than that of the protons on the α-face of the sugar, indicating that the acetyl groups do not directly contribute to the interaction.

Further support for a favorable carbohydrate–π interaction came from NOEs between Ac_4Glc and Trp in peptide **1**: NOEs were only observed to the protons on the α-face of Ser(Ac_4Glc)

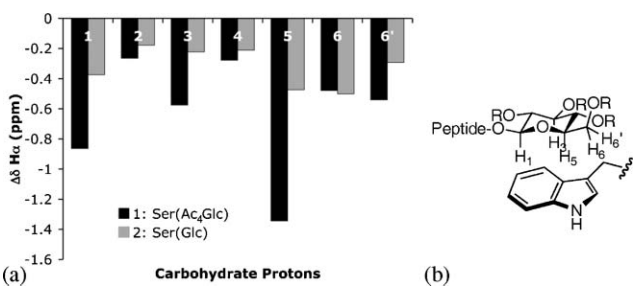


Fig. 2 (a) Upfield shifting of carbohydrate protons in peptides **1** and **2**. See Table 1 for conditions. (b) Proposed geometry between Ac_4Glc and Trp in peptide **1**.

Department of Chemistry, CB 3290, University of North Carolina, Chapel Hill, NC 27599, USA. E-mail: mlwaters@email.unc.edu; Fax: +1 919 962 2388; Tel: +1 919 843 6522

† Electronic supplementary information (ESI) available: Synthetic procedures, NMR assignments, and characterization of all peptides. See DOI: 10.1039/b711431a

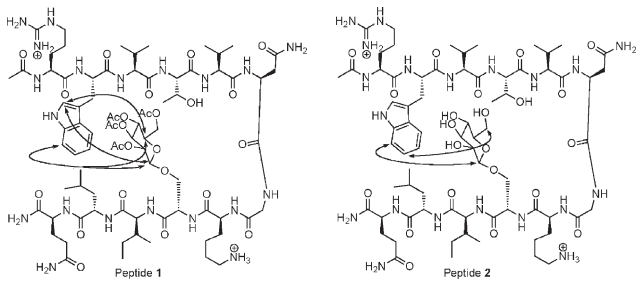


Fig. 3 NOEs present between the carbohydrate side-chain and Trp for **1** and **2**.

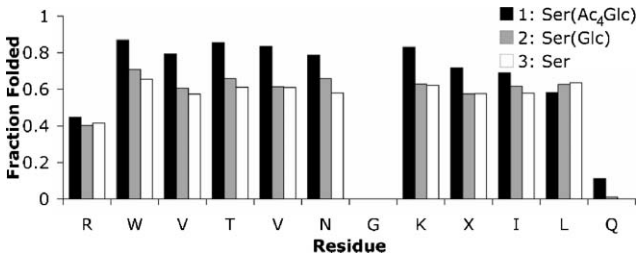


Fig. 4 Fraction folded as determined from H_α chemical shifts. See Table 1 for conditions.

(Fig. 3). A few NOEs were also observed between Trp and the acetyl groups of peptide **1**, but could not be assigned.

The fraction folded was determined from the Gly[12a] and α-proton (H_α)[16] chemical shifts relative to control peptides representing the fully folded and random coil states (see ESI†).[17] Analysis of the NMR data indicates that the incorporation of Ser(Ac_4Glc) in peptide **1** results in a well-folded β-hairpin structure, and that removal of the sugar (peptide **3**) results in a decrease in the hairpin stability (Table 1, Fig. 4).

Quantification of the sidechain–sidechain interaction energy in peptide **1** was achieved through a double mutant cycle (see ESI†), which allows for the measurement of a specific interaction while correcting for any differences in β-sheet propensities of the residues, hydrogen bond strengths, *etc.*, in the mutant peptides.[12a,16] The magnitude of the Ac_4Glc–Trp interaction in peptide **1** was determined to be -0.8 (± 0.1) kcal mol^{-1}. Interestingly, the magnitude of this interaction is greater than that of either a π–π interaction between two Phe sidechains or a cation–π interaction between Trp and Lys or Arg, as measured in the context of a β-hairpin.[12c–f]

From thermal denaturation,[16] folding of **1** was found to be enthalpically favorable ($\Delta H^\circ = -5.9$ kcal mol^{-1}) and entropically unfavorable ($\Delta S^\circ = -16.4$ cal mol^{-1} K^{-1}) with a negative change in heat capacity ($\Delta C_p^\circ = -112$ cal mol^{-1} K^{-1}) (see ESI†). The enthalpic driving force for the folding of peptide **1** is consistent with values determined for protein–oligosaccharide interactions.[5c] Moreover, it is similar to that observed for peptides containing cation–π and π–π interactions,[12c–f] but differs from that of peptides with primarily hydrophobic interactions which are entropically more favorable.[12e] Hence, this suggests contributions from C–H···π and van der Waals to the carbohydrate–π interaction, as has been proposed elsewhere.[5]

To determine the role of solvation on the Trp–carbohydrate interaction, peptide **2**, in which glucose is unprotected, was investigated. Surprisingly, removal of the acetyl groups greatly reduced the sidechain–sidechain interaction, as determined from NMR chemical shifts and NOEs. Little upfield shifting of the glucose was observed (Fig. 2), with the greatest shifts observed at the hydrophobic cluster at H5, H6, and H6′. In addition, fewer NOEs between the Trp and Ser(Glc) sidechains were observed (Fig. 3). Moreover, the hairpin itself was significantly destabilized (Table 1); indeed, peptide **2** is no more stable than the control peptide **3**, in which there is no carbohydrate. The fact that Ser(Glc) does not stabilize the hairpin structure is likely due to the higher desolvation cost of Glc relative to Ac_4Glc: the Clog*P* values for 1-MeGlc and 1-Me-Ac_4Glc (where the Ser was replaced by a methyl group) are -2.42 and 0.93, respectively, indicating that 1-MeGlc is hydrophilic and 1-Me-Ac_4Glc is hydrophobic.§ Hence, this suggests that the desolvation cost for Ser(Glc) is larger than the magnitude of the Trp–Glc interaction, resulting in no net stabilization.

These studies indicate that glucose and Trp interact in an attractive manner *via* the α-face of the sugar and that this interaction provides considerable stabilization to a β-hairpin peptide when the desolvation cost is reduced through protection of the hydroxyl groups. This model system differs from a carbohydrate binding protein in that the sugar must be protected to interact with the aromatic ring. This difference arises from the fact that a protein complex provides specific hydrogen bonds between the sugar and protein which effectively supply the energy to overcome the desolvation penalty. Hence the hydrogen bonds act in concert with the carbohydrate–π interaction to bind the carbohydrate. This is conceptually similar to Lemieux's hydrated polar gate concept for the binding of amphiphilic molecules.[18]

Although numerous synthetic hosts have incorporated aromatic groups as part of the recognition site for carbohydrate binding, to the best of our knowledge, this is the first example quantifying the magnitude of the carbohydrate–π interaction experimentally. Moreover, we demonstrate the ability to utilize a carbohydrate–π interaction to control folding.

In conclusion, these studies provide insight into the role of carbohydrate–π interactions in carbohydrate recognition and demonstrate its utility as a molecular recognition element

Table 1 Fraction folded and ΔG° at 298 K for peptides **1**–**3**[a]

Peptide	Carbohydrate	Fraction folded (Gly)[b]	Fraction folded (H_α)[c]	ΔG°/kcal mol^{-1} [d]
1	Ac_4Glc	0.85	0.83 (0.02)	−1.03
2	Glc	0.65	0.63 (0.02)	−0.37
3	None	0.64	0.60 (0.01)	−0.34

[a] Conditions: 50 mM sodium acetate-*d4*, pD 4.0 (uncorrected) at 298 K, referenced to DSS. [b] Error is ±0.01 based on the NMR chemical shifts. [c] H_α fraction folded was determined from the average of the values for Val 3, Val 5, Lys 8, and Ile 10. The value in parentheses is the standard deviation. [d] ΔG° was determined from the Gly splitting. Error is $\leqslant \pm 0.06$ kcal mol^{-1} based on the error in the fraction folded.

comparable in magnitude to other aromatic interactions. Further studies are underway to explore the scope and driving force of this interaction.

This work was supported in part from the NIH (GM071589 and GM072691), the STC Program of the National Science Foundation under Agreement No. CHE-9876674, and an Alfred P. Sloan Fellowship to MLW.

Notes and references

‡ Selective upfield shifting of the protons on the α-face of a carbohydrate has been observed in an NMR titration with phenol in aqueous solution, although the magnitude of upfield shifting was very small (6–14 Hz). See ref. 8*a*. In addition, upfield shifting of carbohydrate protons has been observed in some host–guest complexes in water (ref. 9*a*) and CD_3CN–CD_3OD (ref. 9*c*).

§ Clog*P* is the calculated octanol–water partition constant. Values were calculated using Chemdraw Ultra 9.0.1.

1 (*a*) T. Feizi, in *Glycoimmunology 2: Advances in Experimental Medicinal and Biology*, ed. J. S. Axford, Plenum, New York, 1998, vol. 435, pp. 51–54; (*b*) T. K. Lindhorst, *Top. Curr. Chem.*, 2002, **218**, 201–235.

2 For a review on biomimetic carbohydrate receptors, see: A. P. Davis and R. S. Wareham, *Angew. Chem., Int. Ed.*, 1999, **38**, 2978–2996.

3 (*a*) S. Elgavish and B. Shaanan, *Trends Biochem. Sci.*, 1997, **22**, 462–467; (*b*) N. K. Vyas, M. N. Vyas and F. A. Quiocho, *Science*, 1988, **242**, 1290–1295.

4 S. F. Poget, G. B. Legge, M. R. Proctor, J. G. Butler, M. Bycroft and R. L. Williams, *J. Mol. Biol.*, 1999, **290**, 867–879.

5 (*a*) M. Muraki, H. Morii and K. Harata, *Protein Pept. Lett.*, 1998, **5**, 193–198; (*b*) M. Muraki, H. Morii and K. Harata, *Protein Eng.*, 2000, **13**, 385–389; (*c*) M. I. Chavez, C. Andreu, P. Vidal, N. Aboitiz, F. Freire, P. Groves, J. L. Asensio, G. Asensio, M. Muraki, F. J. Canada and J. Jimenez-Barbero, *Chem.–Eur. J.*, 2005, **11**, 7060–7074.

6 (*a*) M. D. Fernandez-Alonso, F. J. Canada, J. Jimenez-Barbero and G. Cuevas, *J. Am. Chem. Soc.*, 2005, **127**, 7379–7386; (*b*) G. Giancarlo, D. Potenza, A. Canales, J. Jimenez-Barbero, K. K. Baldridge and A. Bernardi, *J. Am. Chem. Soc.*, 2007, **129**, 2890–2900; (*c*) J. N. Martin, E. M. Munoz, C. Schwergold, F. Souard, J. L. Asensio, J. Jimenez-Barbero, J. Canada and C. Vicent, *J. Am. Chem. Soc.*, 2005, **127**, 9518–9533.

7 J. Screen, E. C. Stanca-Kaposta, D. P. Gamblin, B. Liu, N. A. Macleod, L. C. Snoek, B. G. Davis and J. P. Simons, *Angew. Chem., Int. Ed.*, 2007, **46**, 3644–3648.

8 (*a*) R. Palma, M. E. Himmel and J. W. Brady, *J. Phys. Chem. B*, 2000, **104**, 7228–7234; (*b*) V. Spiwok, P. Lipovova, T. Skalova, E. Buchtelova, J. Hasek and B. Kralova, *Carbohydr. Res.*, 2004, **339**, 2275–2280; (*c*) M. S. Sujatha, Y. U. Sasidhar and P. V. Balaji, *Biochemistry*, 2005, **44**, 8554–8562.

9 (*a*) K. Kobayashi, Y. Asakawa, Y. Kato and Y. Aoyama, *J. Am. Chem. Soc.*, 1992, **114**, 10307–10313; (*b*) J. Otsuki, K. Kobayashi, H. Toi and Y. Aoyama, *Tetrahedron Lett.*, 1993, **34**, 1945–1948; (*c*) S. Anderson, U. Neidlein, V. Gramlich and F. Diederich, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 1596–1600; (*d*) A. Vacca, C. Nativi, M. Cacciarini, R. Pergoli and S. Roelens, *J. Am. Chem. Soc.*, 2004, **126**, 16456–16465; (*e*) M. Mazik, H. Cavga and P. G. Jones, *J. Am. Chem. Soc.*, 2005, **127**, 9045–9052; (*f*) R. D. Hubbard, S. R. Horner and B. L. Miller, *J. Am. Chem. Soc.*, 2001, **123**, 5810–5811; (*g*) E. Klein, M. P. Crump and A. P. Davis, *Angew. Chem., Int. Ed.*, 2005, **44**, 298–302.

10 J. C. Morales and S. Penades, *Angew. Chem., Int. Ed.*, 1998, **37**, 654–657.

11 Synthesis of Ser(Glc) and Ser(Ac_4Glc) was achieved using a modified procedure of Elofsson *et al.* (see ESI†); L. A. Salvador, M. Elofsson and J. Kihlberg, *Tetrahedron*, 1995, **51**, 5643–5656. Deprotection of the sugar was accomplished following the procedure of P. Sjolin, M. Elofsson and J. Kihlberg, *J. Org. Chem.*, 1996, **61**, 560–565.

12 (*a*) M. S. Searle, S. R. Griffiths-Jones and H. Skinner-Smith, *J. Am. Chem. Soc.*, 1999, **121**, 11615–11620; (*b*) M. S. Searle, *J. Chem. Soc., Perkin Trans. 2*, 2000, 1011–1020; (*c*) C. D. Tatko and M. L. Waters, *J. Am. Chem. Soc.*, 2002, **124**, 9372–9373; (*d*) C. D. Tatko and M. L. Waters, *Protein Sci.*, 2003, **12**, 2443–2452; (*e*) C. D. Tatko and M. L. Waters, *J. Am. Chem. Soc.*, 2004, **126**, 2028–2034; (*f*) S. E. Kiehna and M. L. Waters, *Protein Sci.*, 2003, **12**, 2657–2667; (*g*) R. M. Hughes and M. L. Waters, *J. Am. Chem. Soc.*, 2005, **127**, 6518–6519.

13 F. A. Syud, H. E. Stanger and S. H. Gellman, *J. Am. Chem. Soc.*, 2001, **123**, 8667–8677.

14 S. R. Griffiths-Jones, A. J. Maynard, G. J. Sharman and M. S. Searle, *Chem. Commun.*, 1998, 789–790.

15 J. A. Pople, *J. Chem. Phys.*, 1956, **24**, 1111.

16 A. J. Maynard, G. J. Sharman and M. S. Searle, *J. Am. Chem. Soc.*, 1998, **120**, 1996–2007.

17 F. A. Syud, J. F. Espinosa and S. H. Gellman, *J. Am. Chem. Soc.*, 1999, **121**, 11577–11578.

18 R. U. Lemieux, *Acc. Chem. Res.*, 1996, **29**, 373–380.

