Tetrahedron: *Asymmetry* 20 (2009) 718–722

Contents lists available at ScienceDirect

Tetrahedron: *Asymmetry*

journal homepage: www.elsevier.com/locate/tetasy

Tertahedron: *Asymmetry* Report Number 113

Conformational change and selectivity in explicitly hydrated carbohydrates

John P. Simons [a,*], Benjamin G. Davis [b,*], Emilio J. Cocinero [a], David P. Gamblin [b], E. Cristina Stanca-Kaposta [a,†]

[a] *Department of Chemistry, University of Oxford, Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford OX1 3QZ, UK*
[b] *Department of Chemistry, University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford OX1 3TA, UK*

ARTICLE INFO

Article history:
Received 4 February 2009
Accepted 23 February 2009
Available online 4 May 2009

Special edition of Tetrahedron: Asymmetry in honour of Professor George Fleet's 65th birthday

ABSTRACT

The combination of vibrational spectroscopy, conducted in a supersonic jet expansion, with computation through molecular mechanics, density functional theory (DFT) and ab initio calculation, has provided a new approach to the conformational and structural assignment of carbohydrates and their molecular complexes. This article reviews the new insights it has provided on the regioselectivity and conformational choice in singly and multiply hydrated monosaccharides. It reveals a systematic pattern of conformational preference and binding site selectivity, driven by the provision of optimal, co-operative hydrogen-bonded networks in the hydrated sugars. Water binding is invariably 'focused' around the hydroxymethyl group (when present); the bound water molecules (on multiply hydrated mannose) are located exclusively on its hydrophilic face while the hydrophobic face remains 'dry'; and there is a correlation between the locale of the preferred binding sites and those involved in protein–carbohydrate molecular recognition.

Contents

1. Introduction

Carbohydrate molecules are notoriously flexible and in aqueous or physiological environments their conformational structures can be influenced by interaction with neighbouring ions or molecules and particularly, by explicit hydration.[1–3] This may in turn, influence their selective molecular recognition at protein–carbohydrate receptor sites, thought to involve their preferred, solution conformation(s).[4] Determining and understanding what rules, if any, govern them and characterizing their three-dimensional conformations presents a major challenge. Their NMR spectra are complex and are associated with time averaged, dynamical structures which can complicate the interpretation of NMR measurements.[5–7] Structural determinations based upon nuclear Overhauser effects in solution or residual dipolar coupling in liquid crystal environments,[5,8,9] coupled with molecular dynamics simulations, may not provide unique answers. Molecular dynamics simulations of α1,6-glycosidic conformations in aqueous solution, for example,[10] suggest that intramolecular hydrogen bonding is 'washed out' by the solvent so that steric repulsion provides the principal structural determinant. On the other hand, similar simulations of β1,4- (and β1,3-) linked glycosides identify an important role for inter-molecular hydrogen bonding which creates rigid extended water-bridged structures, while α1,2-, α1,3- and α1,6-linked glycosides adopt flexible and compact conformations.[11]

Recently a new laser-based strategy has been developed, which allows comparisons to be made between the conformational struc-

* Corresponding authors.
E-mail addresses: john.simons@chem.ox.ac.uk (J.P. Simons), ben.davis@chem.ox.ac.uk (B.G. Davis).
† Present address: Insitut für Experimentalphysik, Freie Universität Berlin, Arnimallee 14, 14195, Berlin, Germany.

0957-4166/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tetasy.2009.02.032

tures of isolated carbohydrates and their micro-hydrated complexes determined experimentally, stabilized at low temperatures in the gas phase.[12] The vibrational spectra of individual conformers and hydrated clusters, isolated in a cold molecular beam, can be identified and selected through the depletion in their ground state populations promoted by resonant absorption of tunable IR laser radiation. The depletions are most simply detected via the dip in their UV laser-induced, resonant two-photon ionization signals; when coupled with mass spectrometry this allows size selected molecular complexes to be detected and interrogated spectroscopically.[12]

Characteristic vibrational signatures associated with OH (and NH) stretching modes are particularly informative since they are extremely sensitive to the local hydrogen-bonded environment. When coupled with theory—molecular mechanics, density functional theory (DFT) and ab initio calculations, which provide the 'menus of possibilities from which Nature makes its choice'—the strategy provides direct access to the intrinsic conformational landscape of the carbohydrate and exposes the consequences of its explicit hydration. The level of theory employed is a compromise between the known reliability of the method and the computational cost.

In most cases, since many structures are possible, a typical strategy begins with the generation of a large set of feasible conformers and hydrate structures using a Monte Carlo search procedure implemented from a molecular modelling suite and molecular mechanics force fields, to perform an extensive search of the large number of possible structures. As the size of the studied systems increases, a full search of the potential structures can become very expensive and in these cases the 'sledge–hammer' approach is modified in favour of an intelligent feedback strategy in which the experimental data are used to guide the calculations. Specific spectral features can signal the characteristics of probable structures, for example, groups involved in specific hydrogen bonding can be obtained from the position of their associated vibrational bands in the experimental IR spectrum. Characteristic conserved conformations generate characteristic vibrational signatures to create a 'spectroscopic alphabet' which can aid the analysis of more complex structures; those which do not provide an acceptably close match (at that level) to the experimental spectrum are eliminated from further calculations. In a typical procedure, the relevant structures are first submitted to calculations at the HF level, using a moderate size basis set, usually 6-31G(d) or 6-31+G(d) and the most stable are re-optimized using the DFT method, generally with the Becke-style three-parameter Lee–Yang–Parr hybrid correlation functional, B3LYP which gives reliable results for hydrogen-bonded biomolecules. Relative energies are then calculated using a larger basis set and include electron correlation, typically using the second-order Møller–Plesset perturbation theory method, MP2, and the lowest-lying sub-set of structures, that is, those which might be expected to be significantly populated are submitted for harmonic frequency calculations again using the B3LYP functional and 6-31+G(d) as basis set. These calculations provide the IR vibrational frequencies and intensities and the zero point and free energy corrections; appropriate 'anharmonic' scaling factors are used to compare the calculated spectra with the experimental ones.[13]

Over the last few years, this strategy has been applied to a representative series of α- and β-monosaccharides, including glucose, galactose and mannose, as well as fucose (6-deoxygalactose) and xylose (Fig. 1),[14–16] and has led to a set of predictive rules governing carbohydrate conformational preferences and explicit water-binding site selectivity.[15] A bridge to the condensed phase is also being explored through comparisons between their vibrational Raman optical activity (ROA) spectra measured in aqueous solution, and ab initio simulations based upon a weighted 'basis set' of the explicitly hydrated conformations determined in the gas phase. Early results suggest that the solution phase conformations are strongly influenced by explicit hydration.[17] Herein we report a brief review of some of these developments, illustrated by recent investigations of hydrated monosaccharides.

2. Hydration, conformation and selectivity—the rules of the game

Key factors controlling conformational preference and site selectivity in hydrated monosaccharides include the flexibility of their exocyclic hydroxymethyl groups (in glucose, galactose and mannose); their anomeric configuration; and the relative orientations (axial vs equatorial) of their hydroxyl groups. These factors can operate separately or collectively, adapting the carbohydrate conformation and configuration to optimize the sequence of intra- and inter-molecular hydrogen-bonded interactions in the hydrated complex. The first (and second) bound water molecules are invariably located near the hydroxymethyl group, inserting into the weakest pre-existing hydrogen bond in the bare molecule. In many cases, the carbohydrate conformation changes in order to achieve this. An illustration of this is shown in Figure 2 which presents the experimental and computed IR spectra of phenyl β-D-glucopyranoside and its monohydrate, recorded in the gas phase.[14] The preferred conformation of the hydroxymethyl group changes from G+g− to G−g+ (*gg* to *gt*) while the orientation of the peripheral OH groups switches from counter-clockwise, 'cc' to clockwise, 'c'.

Figure 3 presents the preferred conformations and water-binding sites in the series of hydrated monosaccharides, α- and β-glucose (Glc), galactose (Gal) and mannose (Man). When the OH4 group is oriented equatorially, as in Glc and Man, each of the hydrated sugars either adopts (α,βGlc, βMan), or retains (αMan) a cG−g+ conformation, to provide an optimized structure for reception of the bound water molecule at the (4,6) hydroxymethyl site,

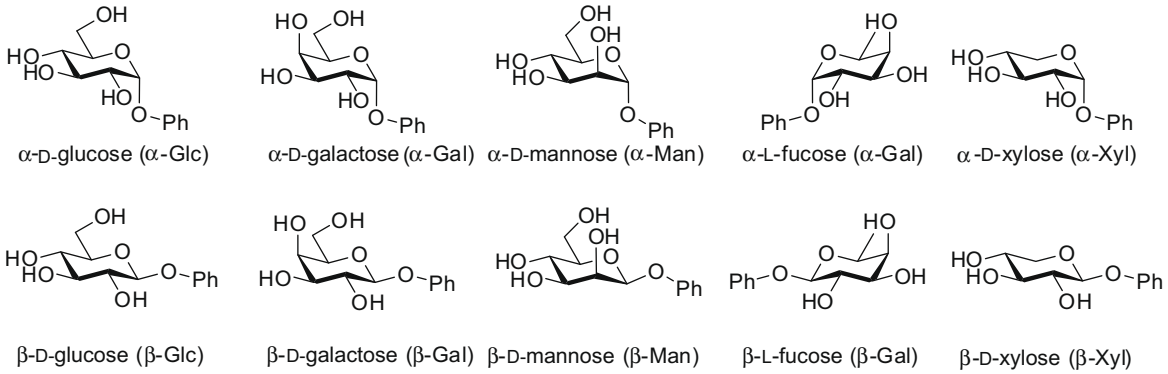

Figure 1. Schematic structures of glucose, galactose, mannose, fucose and xylose monosaccharides: R = methyl, phenyl.

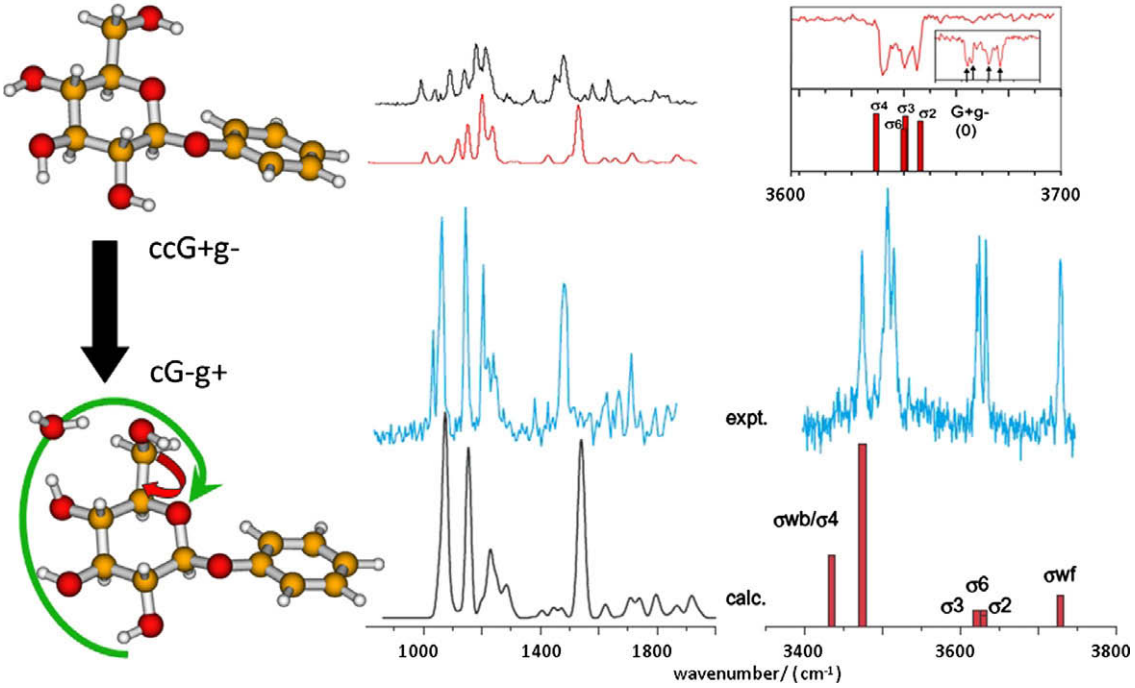

Figure 2. Experimental (upper) and computed (lower) IR spectra of the most populated conformers of phenyl β-D-glucopyranoside and its monohydrate. Note the change in conformation promoted by explicit hydration, which allows selective insertion of the bound water molecule at the (4,6) site.

the weakest link in the original ccG+g– configuration, and maximize the length of the co-operatively hydrogen-bonded chain, OH2→OH3→OH4→*W*→OH6→O1. In Gal the OH4 group is oriented axially, facilitating strong hydrogen bonding between OH4 and OH6 (in the cG–g+ conformer) or between OH4 and OH3 (in the ccG+g– conformer), and in both anomers, the water molecule inserts on the other side of the hydroxymethyl group, at the (6,5) site.

The hydroxymethyl group clearly provides the favoured binding site in the singly hydrated complexes of the three monosaccharides, Glc, Gal and Man.[15] When the hydroxymethyl is replaced by a hydrogen atom or a methyl group with no change in the relative configuration, for example, in xylose[16] rather than glucose or in fucose[18] rather than galactose, the bound water molecule shifts to a site further around the pyranose ring. The binding is still selective however, 'unpicking' the weakest link(s) in the pre-existing (cc) $(OH–O)_n$ chain. In the fucosides (6-deoxy Gal) this corresponds to the (3,2) site; in the xylosides however, it also depends on the anomeric configuration. The (3,2) site is again preferentially selected in the α-anomer but in the β-anomer, the water molecule binds at the (2,1) site, where the OH2→O1(equ)→equ rather than equ→axial) interaction is noticeably weaker. This is very clearly seen in the set of characteristic vibrational signatures shown in Figure 4. In each of the three β-anomers the band associated with the OH2 mode, σ2, is located at the highest wavenumber; in the α-anomers however, it lies at the lowest wavenumber, displaced by the enhanced hydrogen bonding between OH2 and O1. It is tempting to associate explicit hydration at the (2,1) site in the β-anomer of xylose with its preferential population in aqueous solution.

3. Multiple hydration and molecular recognition

Binding site selectivity continues to operate in multiply hydrated carbohydrates. In the doubly hydrated phenyl β-D-glucopyranoside, for example, the binding continues to be focused around the hydroxymethyl group with one molecule occupying the (4,6) site as before and the other, the (6,5) site, to extend the clockwise $(OH–O)_n$ chain even further, see Figure 5a. The energy gain provided by this co-operative chain is enough to persuade the water

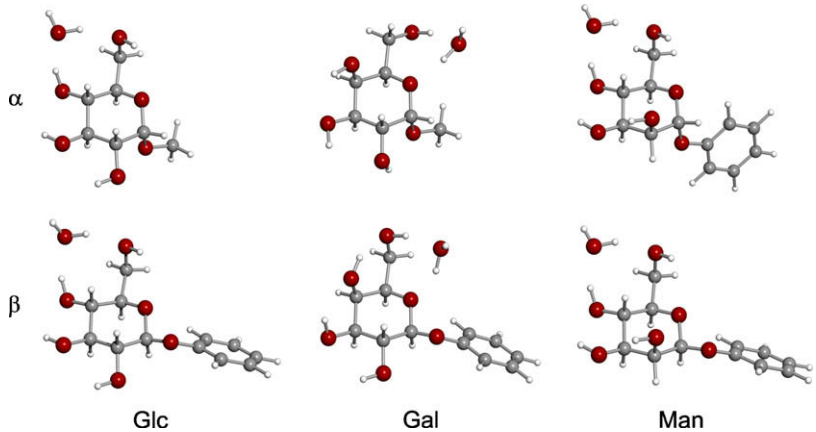

Figure 3. The preferred structures of the singly hydrated α- and β-anomers of phenyl (or methyl) glucoside, galactoside and mannoside.

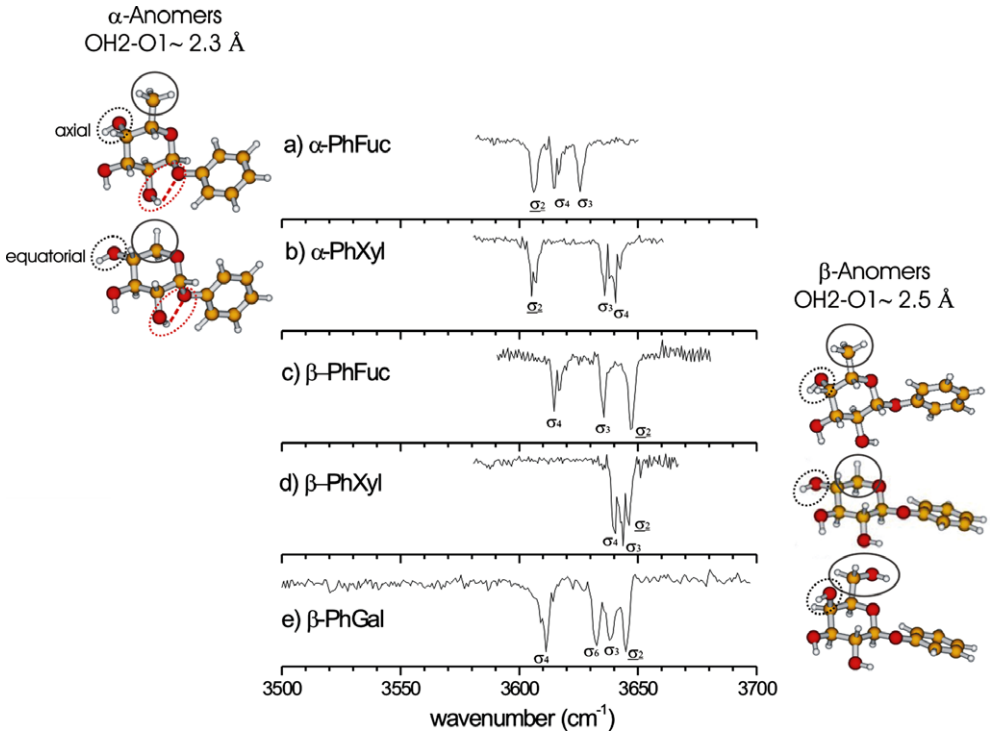

Figure 4. 'Vibrational signatures' of the α-and β-anomers of phenyl fucopyranoside and xylopyranoside (PhFuc, PhXyl) and the β-anomer of phenyl galactopyranoside (PhGal). Note the relative positions of the band σ2, associated with the OH2 stretch mode: in the α-anomers, where the OH2–O1 bonding is stronger, the band is displaced towards lower wavenumber. Galactose and fucose on the one hand, and xylose on the other, also display a similar alternation in σ4, reflecting the change in orientation of OH4 from axial to equatorial and a consequent increase in the OH4–OH3 (hydrogen-bonded) distance. Each of the four OH bands in galactose, σ2–σ6, is clearly resolved and σ6 is clearly 'missing' from the IR spectra of fucose and xylose, both of which lack a C—6 OH group.

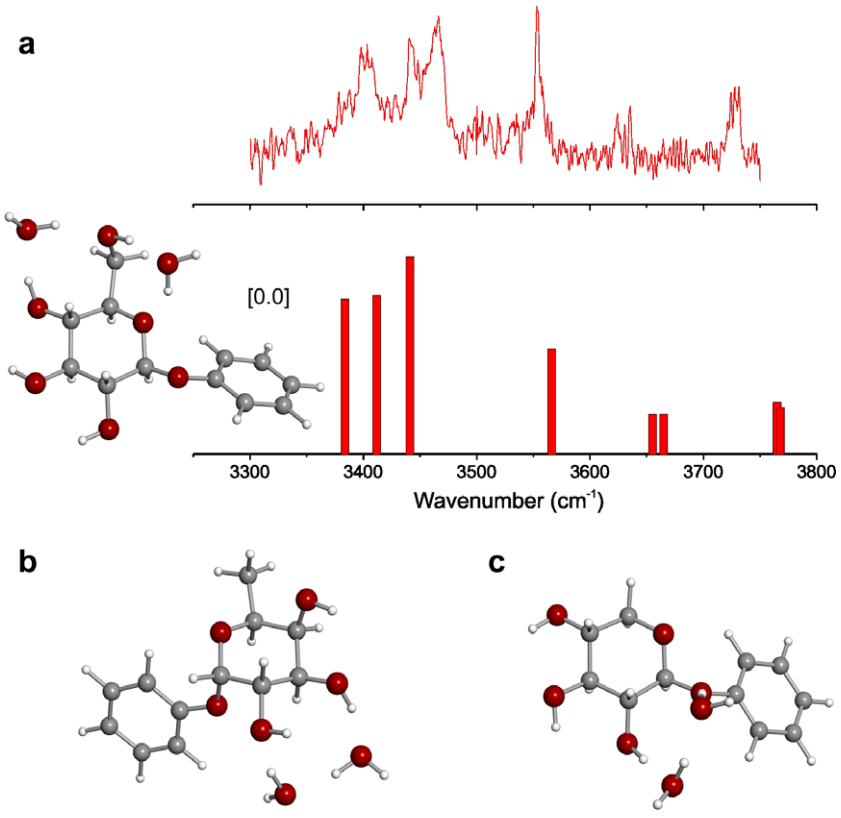

Figure 5. (a) Experimental and computed IR spectra of doubly hydrated phenyl β-D-glucopyranoside; (b) computed (MP2/6-311++G**//B3LYP/6-311+G*) lowest energy structure of doubly hydrated phenyl α-L-fucopyranoside—which accommodates a water dimer bound at the (3,2) site and phenyl β-D-xylopyranoside where the dimer is bound at the (2,1) site (see text).

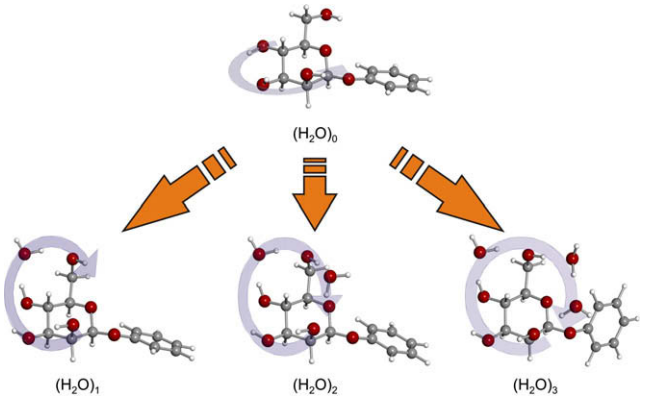

Figure 6. Conformational structures of phenyl β-D-mannopyranoside and its multiply hydrated complexes.

molecules to bind separately to the carbohydrate, rather than to each other.

On the other hand, when there is no C—6 hydroxyl group, for example, in fucose, rather than binding the two water molecules separately (on each side of the hydroxymethyl group) they now prefer to bind as a water dimer, again inserting at the (3,2) site and extending the counter-clockwise hydrogen-bonded chain OH4→OH3→*W1*→*W2*→OH2. Doubly hydrated xylose behaves similarly though in the β-anomer, once again it is the (2,1) site that is favoured. The computed structures and relative energies of the two lowest energy structures of doubly hydrated phenyl α-L-fucopyranoside and β-D-xylopyranoside are shown in Figure 5b.

A particularly striking example of selective binding is provided by multiply hydrated mannose. The preferred structures of singly, doubly and triply hydrated phenyl β-D-mannopyranoside are shown in Figure 6. Like glucose, the two bound water molecules in the dihydrate insert separately on each side of the hydroxymethyl group, the first occupying the (4,6) site favoured in the monohydrate and the second bridging across the pyranose ring to the axial OH2 group. This creates an extended, clockwise 'virtuous circle' of hydrogen bonding, OH2→OH3→OH4→*W1*→OH6→*W2*→OH2; the hydroxymethyl group continues to adopt the G−g+ conformation. The structure adopted by the trihydrate is very similar; the first water molecule occupies the (4,6) site while the second *and* the third insert between OH6 and OH2 as a water *dimer*, to conserve the 'virtuous cycle', OH2→OH3→OH4→*W1*→OH6→*W2*→*W3*→OH2 and retain the G−g+ conformation of the hydroxymethyl group. The side-on view of the trihydrate, shown in Figure 7a provides a striking illustration of the amphiphilic character of the carbohydrate. Its hydrophobic, α-apolar face remains 'untouched' and the water molecules are all located exclusively on the hydrophilic, polar face, with their orientation dictated by the (perturbed) conformation of the carbohydrate to which they are attached; the hydrophobic face remains 'dry'. The bound water molecule acts like a 'spy' surveying the electrostatic topography of the carbohydrate and seeking out the favoured binding site(s).

As well as the binding tending to be focused around the hydroxymethyl group in water–carbohydrate complexes, X-ray crystal structures of simple protein–carbohydrate complexes often identify the hydroxymethyl group as a favoured, though not exclusive recognition point with monosaccharide ligands bound at their O4–O6 or O6–O5 sites. Which carbohydrate conformation does the protein recognize? One possibility would be the conformation most favoured in the aqueous environment;[4] indeed the presence of explicitly bound, conserved water molecules, also commonly involved in protein–carbohydrate complexes, may be because their presence maintains a favoured ligand conformation as well as provides a stronger overall protein–carbohydrate interaction.[19] The

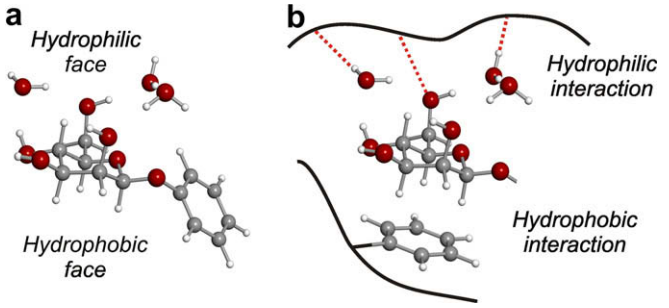

Figure 7. (a) Localised structure of triply hydrated phenyl β-D-mannose and (b) a schematic representation of its potential for interacting simultaneously with polar and non-polar side chains in a stacked protein–carbohydrate complex.

question then arises as to whether the recognition motif is that of the carbohydrate or its solvation shell. Consider again, the hydrophilic and hydrophobic faces displayed in the multiply hydrated phenyl β-D-mannoside. The 'dry' hydrophobic face is ready for stacking interactions with aromatic residues,[4] see Figure 7b but the hydrophilic face is crowned by an ordered array of water molecules with a structure directed by the conformation of the underlying carbohydrate—which may itself be altered by hydration!

Acknowledgements

The research reviewed in this article was founded on Professor George Fleet's dictum: 'With these systems whatever relates to shape is good to learn'. It has been an exciting learning curve to follow and we thank him for the inspiration he provided. We also thank Professor Rebecca Jockusch and Drs Romano Kroemer, Francis Talbot, Pierre Çarçabal and Isabel Hünig who all played key roles in its early development and EPSRC, the Leverhulme Trust (Grant F/08788G), the Spanish Ministry of Education and Science (EJC), the Oxford Supercomputing Centre and the STFC Laser Loan Pool for the research support they have provided.

References

1. Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G. J. *J. Am. Chem. Soc.* **2001**, *123*, 12238–12247.
2. Dashnau, J. L.; Sharp, K. A.; Vanderkooi, J. M. *J. Phys. Chem. B* **2005**, *109*, 24152–24159.
3. Almond, A.; Petersen, B. O.; Duus, J. Ø. *Biochemistry* **2004**, *43*, 5853–5863.
4. Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *Biochem. J.* **2004**, *382*, 769–781.
5. Wormald, M. R.; Petrescu, A. J.; Pao, Y.-L.; Glithero, A.; Elliott, T.; Dwek, R. A. *Chem. Rev.* **2002**, *102*, 371–386.
6. Kräutler, V.; Müller, M.; Hünenberger, P. H. *Carbohydr. Res.* **2007**, *342*, 2097–2124.
7. Coskuner, O. *J. Chem. Phys.* **2007**, *127*, 015101.
8. Imberty, A.; Perez, S. *Chem. Rev.* **2000**, *100*, 4567–4588.
9. (a) Tian, F.; Al-Hashimi, H. M.; Craighead, J. L.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *91*, 485–492; (b) Martin-Pastor, M.; Canales-Mayordomo, A.; Jimenez-Barbero, J. *J. Biomol. NMR* **2003**, *26*, 345–353.
10. Kirschner, K. N.; Woods, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10541–10545.
11. Almond, A. *Carbohydr. Res.* **2005**, *340*, 907–920.
12. Simons, J. P.; Jockusch, R. A.; Çarçabal, P.; Hünig, I.; Kroemer, R. T.; Macleod, N. A.; Snoek, L. C. *Int. Rev. Phys. Chem.* **2005**, *24*, 489–531.
13. Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683–11700.
14. Çarçabal, P.; Jockusch, R. A.; Hünig, I.; Snoek, L. C.; Kroemer, R. T.; Davis, B. G.; Gamblin, D. P.; Compagnon, I.; Oomens, J.; Simons, J. P. *J. Am. Chem. Soc.* **2005**, *127*, 11414–11425.
15. Cocinero, E. J.; Stanca-Kaposta, E. C.; Scanlan, E. M.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *Chem. Eur. J.* **2008**, *14*, 8947–8955.
16. Hünig, I.; Painter, A. J.; Jockusch, R. A.; Çarçabal, P.; Marzluff, E. M.; Snoek, L. C.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2474–2480.
17. Macleod, N. A.; Johannessen, C.; Hecht, L.; Barron, L. D.; Simons, J. P. *Intl. J. Mass Spectrom.* **2006**, *253*, 193–200.
18. Çarçabal, P.; Patsias, T.; Hünig, I.; Liu, B.; Kaposta, C.; Snoek, L. C.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 129–136.
19. Elgavish, S.; Shaanan, B. *J. Mol. Biol.* **1998**, *277*, 917–932.