Hydrogen bonding vs. dispersive interactions: Carbohydrate-*p*-Cresol complexes

E. C. Stanca-Kaposta, Z. Su and J. P. Simon
University of Oxford, Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford OX1 3QZ, UK

P. Hurtado, D. Gamblin and B. Davis
University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford OX1 3TA, UK

Contact | **john.simons@chem.ox.ac.uk**

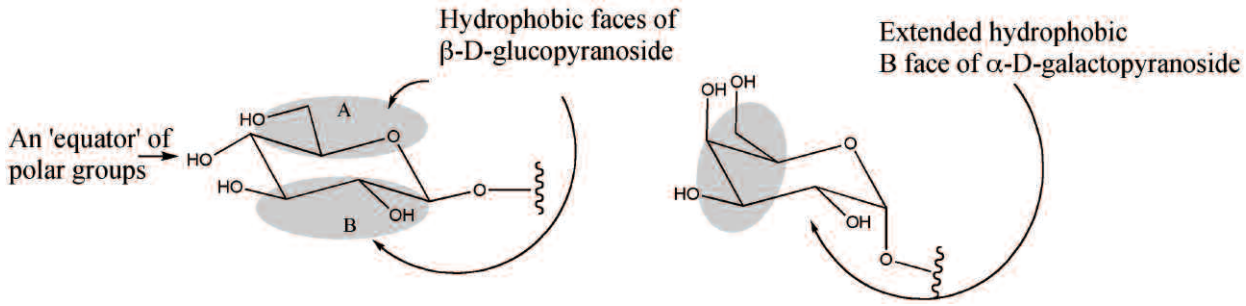

Figure 1. Hydrophobic patches of β-D-glucopyranoside and α-D-galactopyranoside.

Introduction

The selective binding of carbohydrates at protein molecular recognition sites can be mediated by hydrogen bonding and dispersive van der Waals interactions, promoted through their hydrophilic, polar OH groups or by the 'hydrophobic patches' present on one or both sides of the pyranose ring (figure 1)[1,2,3]. In glucose for example, the plane formed by the equatorial orientation of the hydroxyl groups separates two hydrophobic faces located above and below the pyranose ring, referred to as faces A and B: some proteins contain aromatic carbohydrate binding sites which can engage both of these faces of the pyranose ring, creating a sandwich or 'sugar tongs' structure[4]. Carbohydrate-protein interactions, especially with the aromatic residues are of central importance in a wide range of biological processes ranging from cell growth, adhesion and death to the enzymatic recycling of photosynthetically generated plant cell-wall polysaccharides[5,6,7].

The non-covalent interactions between carbohydrates and aromatic residues in proteins (predominantly, tyrosine and tryptophan) which can involve OH-π and dispersive 'CH-π' interactions, are now known to lead to stable structures in the gas phase[8,9], *without* the intervention of the 'hydrophobic interactions' proposed in aqueous solution[10], although this does not exclude their involvement in solution. Gas phase spectroscopic investigations of carbohydrate-aromatic complexes provide a direct means of obtaining information on the fundamentals of sugar-aromatic interactions without the influence of the environment (solvent or crystal). A systematic variation in the choice of sugar molecule, e.g. glucose, galactose and fucose (figure 2), allows an exploration of the response to changes in the disposition of the OH groups, the 'shapes of the apolar patches' and the balance between H-bonded and dispersive interactions.

This strategy was applied initially to a study of carbohydrate-toluene (truncated phenylalanine) complexes generated under supersonic expansion conditions and characterized by mass-selected vibrational spectroscopy[5,6]. Comparisons between the recorded or computed infrared (IR) spectra of the bound and free carbohydrates revealed the contributions made by dispersive, 'CH-π' and in some cases, specific OH-π H-bonded interactions. The strategy exploited the extraordinary sensitivity of their vibrational signatures to the local, intra- and intermolecular hydrogen-bonded environments of their OH groups. The strength of

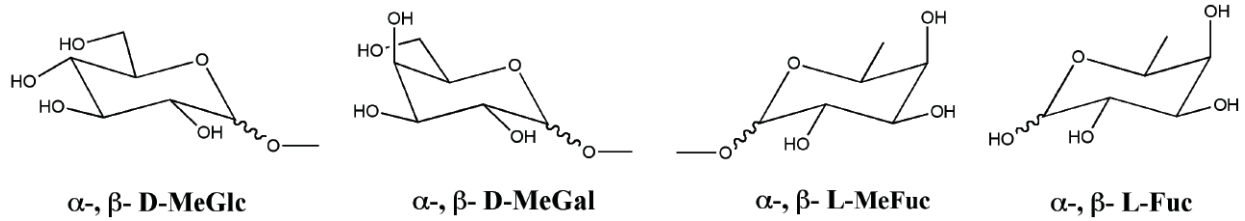

Figure 2. Structures of the monosaccharide derivatives chosen for this study: (Me = methyl, Glc, Gal and Fuc = gluco-, galacto- and fuco-pyranoside).

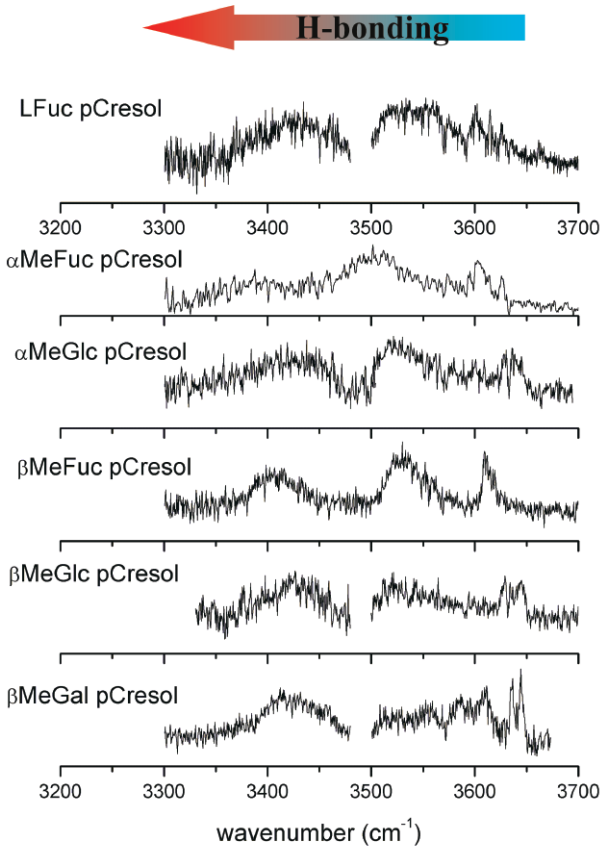

Figure 3. Experimental IRID spectra of carbohydrate-*p*-cresol complexes isolated in the gas phase.

H-bonded interactions was signalled by the degree of displacement of their associated infrared bands towards lower wavenumber. The spectra also reflected the absence of any significant conformational or structural changes imposed on the bound carbohydrate. The strategy has now been extended to address carbohydrate interactions with *p*-cresol (*p*-hydroxy toluene), simulating the tyrosine residues often found in this type of interaction. The complexes were generated in the supersonic expansion of a molecular beam by evaporating the sugars from a heated oven and co-expanding the vapour with *p*-cresol seeded in the Ar carrier gas. Their structures were probed by IR-UV double resonance spectroscopic techniques using an IR laser as the 'pump', and the frequency-doubled output of a YAG-pumped dye laser as the 'probe', inducing resonant two-photon ionisation via the $S_0 \rightarrow S_1$ electronic transition (in the 265-285 nm region) of the bound *p*-cresol.

Figure 3 shows the infrared ion depletion (IRID) spectrum of each of the systems studied: similarities between them are clear. They all present broad, intense bands shifted to frequencies below 3600 cm^{-1}, indicating strong H-bonding. A band centred ~3420 cm^{-1} appears in every case and the appearance of the same feature in the corresponding spectrum of the dimer, $(p\text{-cresol})_2$ (not shown) suggests its association with a complex bound through OH(*p*-cresol) → OH(carbohydrate) hydrogen bonding, while the broad feature located between ~3500 and 3550 cm^{-1} suggests a further strongly H-bonded OH((carbohydrate) → O interaction. The slightly displaced sharp features at higher wavenumbers, above 3600 cm^{-1}, were associated with the 'spectator' OH groups in the carbohydrate, linked through weak, co-operative H-bonds.

Discussion

Quantitative structural assignments of the experimental spectra depend upon the availability of accurate quantum chemical calculations. For illustration figure 4 compares the IRID spectrum of α-MeFuc-*p*-cresol with a series of vibrational spectra computed for its lowest energy structures[11] using density functional theoretical theory (DFT), coupled with *ab initio* calculations using Møller-Plesset perturbation theory, to estimate their relative

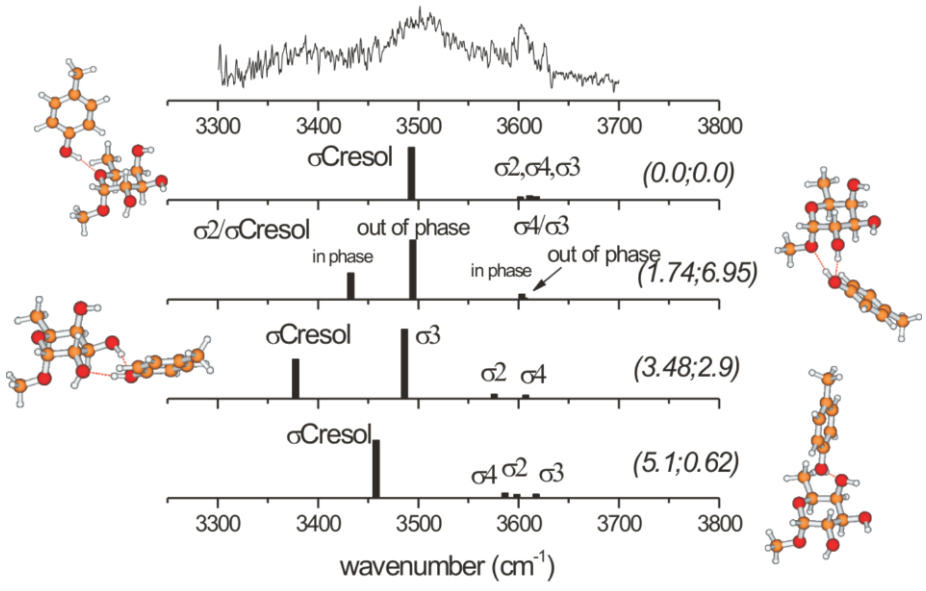

Figure 4. IRID spectrum of the complex of α-methyl fucopyranoside and *p*-cresol compared to the four lowest energy conformers predicted using the B3LYP/6-31+G* // MP2/6-311++G level of theory. Zero point and free energy corrected relative energies are given in brackets.**

energies. The best fit between the experimental spectrum and the predicted frequencies is observed for the third conformer. It is associated with the insertion structure, OH3 (Fuc) → OH (*p*-cresol) → OH2 (Fuc). The most red-shifted band at ~3380 cm^{-1} is associated with OH (*p*-cresol) and the next most displaced band, centred at ~3500 cm^{-1}, is associated with OH3 (Fuc); the remaining weak bands are associated with the 'spectator groups', OH2 and OH4. Although, not the lowest energy conformer, it lies only (~3 kJ mol^{-1} higher in energy then the minimum; since DFT calculations do not reflect the contribution of dispersion which will 'fine tune' the predicted structures and in consequence their relative energies, this is not unexpected.

Assignments can also be aided by comparisons with *experimental* data on similar systems, including the uncomplexed carbohydrates and their bimolecular complexes with toluene or water. Figure 5 for example, compares the IRID spectra of α- and β-MeFuc bound to *p*-cresol with those of their monohydrates, α- and β-PhFuc·H_2O. The IRID spectra of the *p*-cresol and monohydrate complexes are very similar, reproducing well the shifts of the H-bonded OH groups and suggesting similar interactions and hence, similar structures. Indeed, this similarity between the two types of complexes was observed for all the *p*-cresol complexes, suggesting a preference for hydrogen bonded structures similar to those of the hydrated carbohydrates, where water is acting as a spy searching for the most weakly hydrogen bonded OH group in the target. In the singly hydrated complex of α-PhFuc the water molecule inserts between OH3 and OH2, exactly as predicted for the best fit structure of α-MeFuc-*p*-Cresol.

Conclusions

Complexes between carbohydrates and *p*-cresol, a model system for tyrosine, can be isolated in the gas phase but in contrast to the corresponding carbohydrate-toluene complexes their interaction is dominated by hydrogen bonding rather than dispersion. The interaction with *p*-cresol is selective, creating structures that appear to be very similar to these of their singly hydrated complexes. The OH group of *p*-cresol inserts into the weakest hydrogen bond of the carbohydrate, extending and strengthening the co-operative H-bonded OH chain. In 'real life' the environment around a protein-bond carbohydrate will of course, be very different from the gas phase, including the surrounding protein coil and water molecules as well. Stacking interactions between the apolar face(s) of the carbohydrate and tyrosine residues could then be favoured by complementary H-bonded interaction

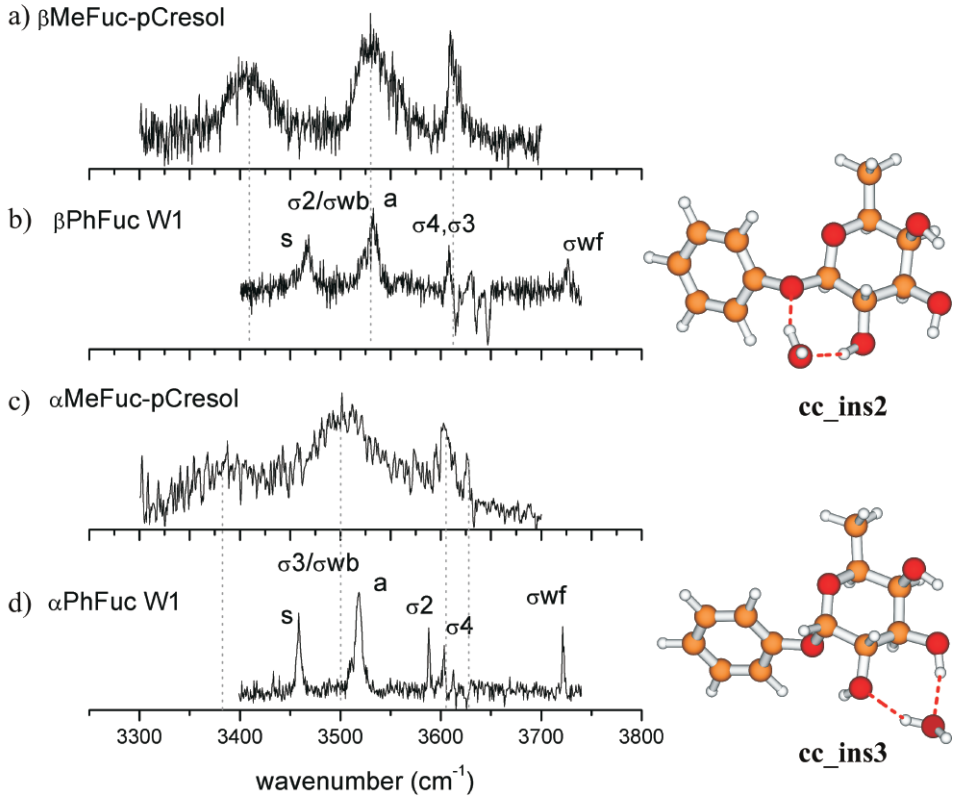

Figure 5. IRID spectra of (a) and (c), β- and α-MeFuc-*p*-Cresol and (b) and (d), β- and α-PhFuc.H_2O with the assignments based on previous studies. The dotted lines indicate the positions of the OH bands in β-MeFuc-*p*-Cresol. [Me=methyl, Ph=phenyl and Fuc= fucopyranoside; σwb and σwf indicate the vibrational frequencies associated with the H-bonded and 'free' OH groups of bound water molecules; 'cc' ('c') indicates a counter-clockwise (clockwise) orientation of the peripheral OH groups, OH4→OH3→OH2→O1 (O1→OH2→OH3→OH4). In the case of the hydrated structures the insertion position of the water molecule is indicated by adding 'ins(position)' to the bare molecule's nomenclature, e.g., ins3 indicates a water molecule inserted between OH3 (acting as an H-bond donor) and OH2 (the acceptor).

of the tyrosine OH with a different part of the protein or a neighbouring water molecule, leaving the aromatic ring free for dispersive 'CH-π' interactions. Higher level calculations, which include dispersion, should lead to a better understanding of the interactions; the possibility of a folding of the aromatic ring over the pyranose ring of the carbohydrate cannot be excluded. From a theoretical point of view however, this field is still quite young and it requires more development and testing to offer methods that are more exact and especially, less time consuming.

Acknowledgments

This research has received support from the STFC laser loan pool, the EPSRC (Grant GR/T26542), the Leverhulme Trust (Grant F/08788G), NSERC Canada (Z.S) and the Oxford Supercomputing Centre.

References

1. "Structural basis of lectin-carbohydrate recognition", W. I. Weis and K. Drickamer, *Ann. Rev. Biochem.*, **65**, 441 (1996).
2. "Glycobiology: Toward Understanding the Function of Sugars", R. A. Dwek, *Chem. Rev.,* **96**, 683 (1996).
3. "Biological roles of oligosaccharides: all of the theories are correct", A. Varki, *Glycobiology*, **3**, 97 (1993).
4. "Carbohydrate-binding modules: fine-tuning polysaccharide recognition", A. B. Boraston, D. N. Bolam, H. J. Gilbert and G. J. Davies, *Biochem. J.*, **382**, 769-781 (2004); "The Structure of Barley α-Amylase Isozyme 1 Reveals a Novel Role of Domain C in Substrate Recognition and Binding: A Pair of Sugar Tongs", X. Robert, R. Haser, T. E. Gottschalk, F. Ratajczak, H. Driguez, B. Svensson and N. Aghajari, *Structure*, **11**, 973-984 (2003).
5. "Structural basis of lectin-carbohydrate recognition", W. I. Weis and K. Drickamer, *Ann. Rev. Biochem.*, **65**, 441 (1996).
6. "Glycobiology: Toward Understanding the Function of Sugars", R. A. Dwek, *Chem. Rev.*, **96**, 683 (1996).
7. "Biological roles of oligosaccharides: all of the theories are correct", A. Varki, *Glycobiology*, **3**, 97 (1993).
8. "IR-Spectral Signatures of Aromatic-Sugar Complexes: Probing Carbohydrate-Protein Interactions", 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.*, **46**, 3644-3648 (2007).
9. "Carbohydrate molecular recognition: a spectroscopic investigation of carbohydrate-aromatic interactions.", E. C. Stanca-Kaposta, D. P. Gamblin, J. Screen, B. Liu, L. C. Snoek, B. G. Davis and J. P. Simons, *Phys. Chem. Chem. Phys.*, **9**, 4444 (2007).
10. "How water provides the impetus for molecular recognition in aqueous solution.", R. U. Lemieux, *Acc. Chem. Res.*, **29**, 373 (1996).
11. M. J. Frisch, *et al*. *Gaussian 03*, Revision B.03; Gaussian, Inc.: Pittsburgh, PA, 2003.