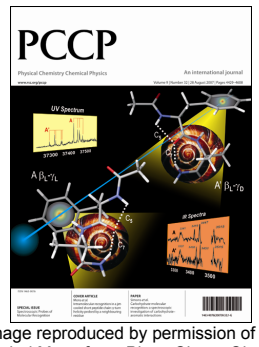

Image reproduced by permission of Dr Michel Mons from *Phys. Chem. Chem. Phys.*, 2007, **9**, 4491

This paper is published as part of a **PCCP** themed issue on

Spectroscopic probes of molecular recognition

Guest edited by Martin Suhm (Universität Göttingen)

Published in issue 32, 2007 of PCCP

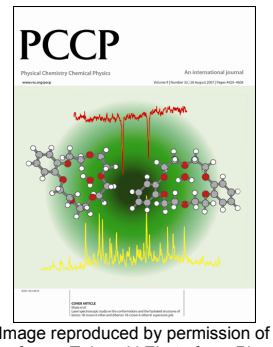

Image reproduced by permission of Professor Takayuki Ebata from *Phys. Chem. Chem. Phys.*, 2007, **9**, 4452

Other papers in this issue:

Carbohydrate molecular recognition: a spectroscopic investigation of carbohydrate–aromatic interactions

E. Cristina Stanca-Kaposta,[a] David P. Gamblin,[b] James Screen,[a] Bo Liu,[a] Lavina C. Snoek,[a] Benjamin G. Davis[b] and John P. Simons*[a]

Received 29th March 2007, Accepted 25th April 2007
First published as an Advance Article on the web 18th May 2007
DOI: 10.1039/b704792d

The physical basis of carbohydrate molecular recognition at aromatic protein binding sites is explored by creating molecular complexes between a series of selected monosaccharides and toluene (as a truncated model for phenylalanine). They are formed at low temperatures under molecular beam conditions, and detected and characterized through mass-selected, infrared ion depletion spectroscopy—a strategy which exploits the extraordinary sensitivity of their vibrational signatures to the local hydrogen-bonded environment of their OH groups. The trial set of carbohydrates, α- and β-anomers of glucose, galactose and fucose, reflects ligand fragments in naturally occurring protein–carbohydrate complexes and also allows an investigation of the effect of systematic structural changes, including the shape and extent of 'apolar' patches on the pyranose ring, removal of the OH on the exocyclic hydroxymethyl group, and removal of the aglycon. Bound complexes invariably form, establishing the general existence of *intrinsic* intermolecular potential minima. In most of the cases explored, comparison between recorded and computed vibrational spectra of the bound and free carbohydrates in the absence of solvent water molecules reveal that dispersion forces involving CH–π interactions, which promote little if any distortion of the bound carbohydrate, predominate although complexes bound through specific OH–π hydrogen-bonded interactions have also been identified. Since the complexes form at low temperatures in the absence of water, entropic contributions associated with the reorganization of surrounding water molecules, the essence of the proposed 'hydrophobic interaction', cannot contribute and other modes of binding drive the recognition of sugars by aromatic residues. Excitingly, some of the proposed structures mirror those found in naturally occurring protein–carbohydrate binding sites.

1. Introduction

The molecular interactions involved in the recognition of carbohydrates by proteins mediate an extraordinarily wide range of biological processes ranging from cell growth, adhesion and death[1–3] to the enzymatic recycling of photosynthetically generated plant cell-wall polysaccharides, such as cellulose and xylan—the most abundant polymers on Earth.[4,5] Non-covalent molecular interactions at protein–carbohydrate binding sites may involve contributions from electrostatic forces such as those generated by coordinated Ca^{2+} ions; or specific hydrogen bonding to polar side chains; or dispersion forces generated by van der Waals interactions.[1–3,6–9] The latter typically involve aromatic residues, notably tyrosine and tryptophan, which often provide platforms for stacking interactions with so-called 'apolar patches' on the pyranose ring and provide recognition sites exploited by Nature.[5,10] In glucose for example, a circumference of hydrophilic, equatorially oriented hydroxyl groups separates two axial apolar hydrophobic faces, located above and below the pyranose ring—the A and B faces (see Fig. 1). Some proteins also contain aromatic carbohydrate binding sites which engage both faces of a pyranose ring, creating a sandwich or so-called 'sugar tong' *motif*[5]—notably glycosyl hydrolases responsible for the degradation of plant cell wall carbohydrates. The involvement of aromatic residues in carbohydrate recognition at protein binding sites is widespread.[5] Sugar-aromatic recognition sites are therefore of primary importance in nature but the fundamentals of these interactions are still not well understood.

The local structures at carbohydrate–protein binding sites have been revealed primarily through X-ray crystallographic or NMR spectroscopic investigations but identification of their structures does not establish the detailed nature of the interactions which create them. Binding to aromatic residues can include dispersion interactions (CH–π bonding), or hydrogen bonding (OH–π), or a finely balanced complement of both;[7] there are arguments which suggest they might also be driven by entropic effects associated with the hydrophobic exclusion of water.[11] Quantitative assessments of the electronic (enthalpic) contributions have been approached through theoretical investigation, using force field, hybrid density functional (DFT) and *ab initio* computation to determine the preferred structures and binding energies of model

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

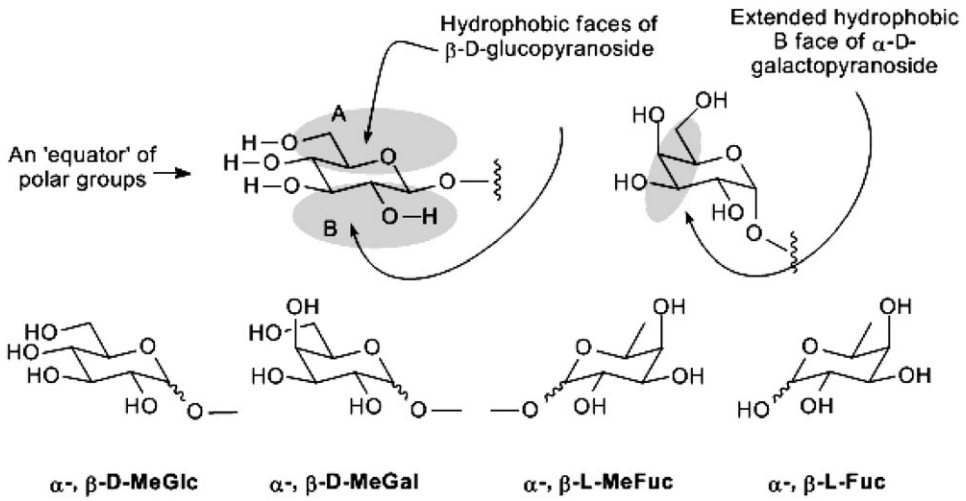

Fig. 1 Carbohydrate structures revealing the apolar patches in carbohydrates exemplified on β-D-glucopyranoside and α-D-galactopyranoside (top) and the monosaccharide derivatives chosen for this study (bottom).

bimolecular carbohydrate-aromatic complexes.[7,9,12,13] These calculations, generally based upon crystallographic protein–carbohydrate stacking structures collected in the protein structure database[14] do reassuringly, indicate their association with attractive carbohydrate-aromatic interactions (and hence negative enthalpy changes) generated through dispersive (CH–π) bonding and hydrogen (OH–π) bonding.[12] This approach has recently been extended to characterize some intramolecular complexes formed within model carbohydrate-aromatic molecular scaffolds and probed *via* NMR NOESY spectroscopy in aqueous solution.[15]

Very recently, the existence of such attractive interactions has been established unequivocally through the experimental creation of monosaccharide-aromatic molecular complexes at low temperatures under molecular beam conditions, detected and characterized through mass-selected vibrational spectroscopy.[16] 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—in the absence of solvent water molecules. They also reflected the incidence, or absence of any significant conformational or structural changes imposed on the bound carbohydrate. The strategy exploits the extraordinary sensitivity of their vibrational signatures to the local, intra- and intermolecular hydrogen-bonded environment of their OH groups.[17] The presence and strength, or absence of H-bonded interactions is signaled by the degree of displacement, or not of their associated infrared bands towards lower wave number.

The present article continues the exploration of intermolecular bonding, selectivity and structural change in carbohydrate-aromatic complexes, through comparisons between the experimental IR spectra of α- and β-anomers of methyl D-gluco- and galacto-pyranoside (MeGlc and MeGal); methyl L-fuco-pyranoside (MeFuc); and L-fucose itself (Fuc), see Fig. 1. The experimental spectra have also been compared with those of the lowest lying conformers of the uncomplexed carbohydrates. Toluene, representing a truncated phenylalanine protein residue, was selected as the aromatic partner. The sequence of monosaccharides, selected to reflect naturally occurring protein–carbohydrate complexes, was also chosen to explore systematically, the effect of structural changes. These include the shape and extent of the apolar patches on the pyranose ring, affected by axial ↔ equatorial configurational changes at the C1 (α ↔ β anomers) and C4 (MeGlc ↔ MeGal, MeFuc) sites; removal of the OH group at the exocyclic carbon site, C6 (MeGal ↔ MeFuc); and removal of the aglycon (MeFuc → Fuc).

2. Experimental

2.1 Spectroscopy and complex formation

A detailed description of the molecular beam experimental method has been published previously.[18] The carbohydrate samples were vaporized at temperatures in the range 70–140 °C, dependent upon the sugar, from an oven mounted directly in front of the nozzle of a pulsed valve (General Valve, 0.8 mm orifice) into an argon free jet expansion (stagnation pressure, 3–6 bar). The aromatic complexes were formed and stabilized in the high collision frequency region of the expanding jet by seeding the argon with toluene vapour (1–3%). The free jet expansion passed through a 3 mm skimmer to form a collimated molecular beam which was subsequently crossed by one, or two tunable laser beams in the extraction region of a linear time of flight mass spectrometer (Jordan).

The frequency-doubled output of a pulsed Nd:YAG-pumped dye laser (Spectron 810/LambdaPhysik FL2002, 0.5 mJ/pulse UV, operating at 10 Hz), excited resonant two photon ionization (R2PI) spectra of each complex, which were recorded in the parent ion mass channel, the toluene partner providing the UV chromophore; the mass spectra also revealed the formation of toluene polymer ions. Conformer specific UV spectra were recorded using IR-UV hole-burning; R2PI and IR-UV hole-burn spectra associated with the complex α-MeFuc · toluene are shown, for illustration in Fig. 2. (Although these were well resolved this was not always the case, primarily because of saturation). Corresponding IR

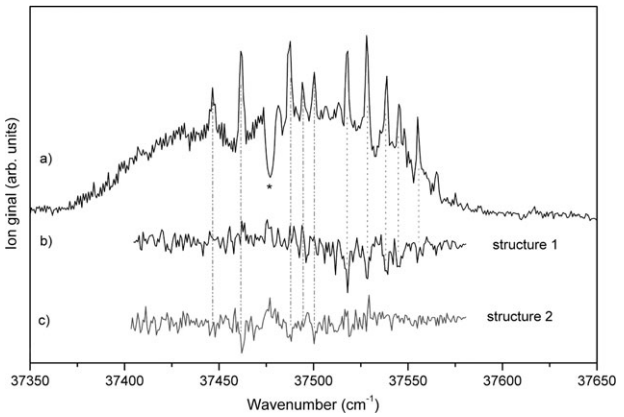

Fig. 2 R2PI (a) and IR-UV hole burn (b, c) spectra of the complex between α-MeFuc and toluene indicating the presence of two different structures. The IR-UV hole burn spectra (b and c) were recorded with the IR burn laser tuned to 3609.48 cm^{-1} and 3618 cm^{-1}, respectively. The dotted and dash-dotted lines indicate the bands in the R2PI spectrum which correspond to one or other of the two contributing complex structures. The dip in the spectrum at 37 477 cm^{-1}, labeled by an asterisk is due to UV absorption by the origin band of uncomplexed toluene.

spectra were recorded through IR ion-dip (IRID) measurements conducted in the range 3400–3800 cm^{-1};[18] the ion dip depletions were generally $\geq 50\%$, more than enough to compensate for the low R2PI signals recorded in the parent ion mass channels. The IR radiation (2–4 mJ/pulse, $\sim$10 ns duration) was generated from the output of a Nd:YAG-pumped IR dye laser (Continuum 8010/ND6000) by difference frequency generation in a lithium niobate crystal. A delay $\sim$150 ns between the IR pump and UV probe laser pulses restricted measurements to the ground electronic state only.

2.2 Computation

Quantum chemical calculations of the structures of the lowest-lying conformers of the uncomplexed methyl pyranosides were performed (following a random MM conformational search[20]) using the Gaussian 03 package.[19] Initial structures were based on those of the corresponding phenyl tagged carbohydrates determined previously.[17] Geometry optimizations and calculations of harmonic vibrational frequencies and intensities were conducted using density functional theory (the B3LYP functional) with a 6-31 + G* basis set. Single point MP2 calculations with a larger basis set (6-311 + + G**) provided relative energies. For comparison with observation, the OH stretch harmonic frequencies were scaled by the factor 0.9734, which provides excellent agreement with experimental data for carbohydrates.[18] Zero-point and free energy corrections to the electronic energy were calculated using the harmonic frequencies.

2.3 Synthesis

Methyl α-L-fucopyranoside was obtained in near quantitative fashion following a standard Fischer glycosylation using a methanolic solution of HCl, generated from the addition of acetyl chloride to methanol. Methyl β-L-fucopyranoside was synthesised from a peracetylated thiofucoside which was activated with methyl triflate in the presence of anhydrous methanol and the proton scavenger 2,6-di-*tert*-butyl-4-methyl-pyridine (DTBMP). The use of the acetyl protection ensured the desired 1,2-*trans* stereochemistry and subsequent deacetylation gave the desired β-fucoside in excellent yield. The remaining mono-saccharides were used as supplied by Fluka.

3. Results and discussion

3.1 MeGlc

The IRID spectra of the β- and α-MeGlc · toluene complexes are presented in Figs. 3 and 4, where they can be compared

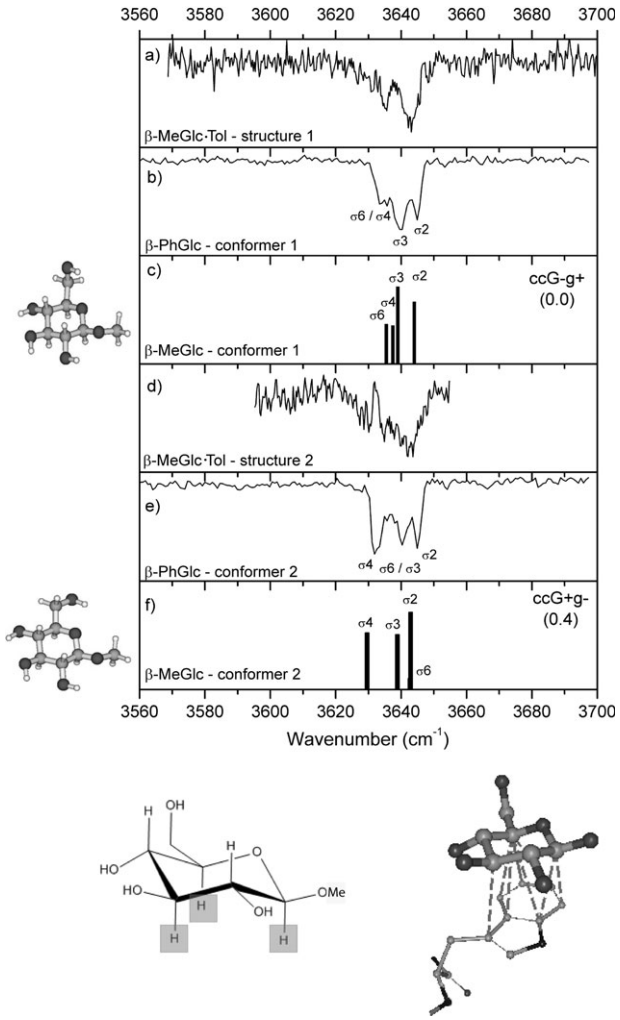

Fig. 3 (a) and (d): IRID spectra associated with two β-MeGlc · toluene complex structures. (b) and (e): IRID spectra and peak assignments, based on previous calculations,[21] of the two lowest lying conformers of uncomplexed β-PhGlc. (c) and (f): predicted vibrational frequencies, assignments and calculated structures for the two lowest-lying conformers of uncomplexed β-MeGlc. Note the counter-clockwise orientation of the co-operative hydrogen bonds. Relative energies in kJ mol^{-1} are shown in brackets. Also shown: a suggested 'three point landing surface' for β-MeGlc bound to toluene *via* a CH1,3,5–π interaction together with the local structure at a tryptophan binding site in the protein cellobiohydrolase I.[22] The dashed lines indicate the hydrophobic interactions between the sugar (CH1,3,5) and the aromatic residue.

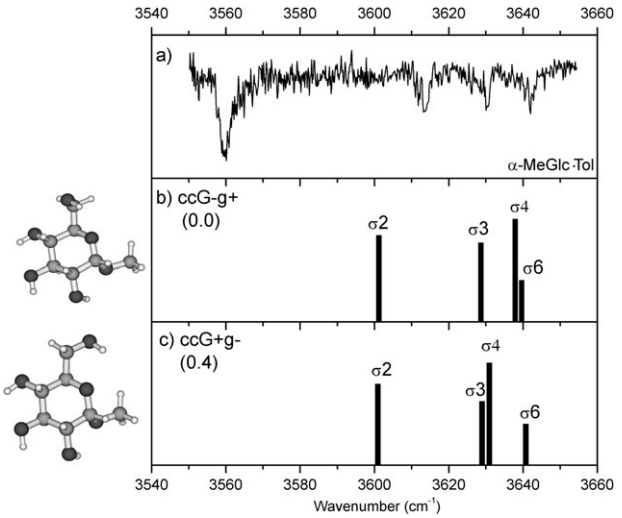

Fig. 4 (a) IRID spectrum of α-MeGlc · toluene and (b), (c) predicted IR spectra and assignments of the two lowest lying conformers of uncomplexed α-MeGlc. Corresponding structures indicating the counter-clockwise orientation of the co-operative hydrogen bonds similar to β-MeGlc are shown as well. Relative energies in kJ mol^{-1} are indicated in brackets.

with the computed spectra of the two lowest-lying, near iso-energetic conformers of the free mono-saccharides, α- and β-MeGlc (ccG-g+) and (ccG + g−)† and also, those observed experimentally for the corresponding phenyl pyranoside, β-PhGlc. The spectra display some striking differences and some striking similarities. Spectral hole-burning measurements for the β-anomeric complex identify two distinct structures but their associated IRID spectra are virtually identical to those computed for the two lowest-lying conformers of the uncomplexed carbohydrate, β-MeGlc and also those observed experimentally for β-PhGlc.[21] The absence of any significant spectral shifts excludes any contribution from OH–π hydrogen bonding and the β-MeGlc · toluene complexes must be bound by dispersive, CH–π-interactions. Furthermore, the close similarity between their IR spectra and those of the two lowest lying conformers of the free carbohydrate implies an absence of any significant distortion of their structure. This would be consistent with weak binding either to the A face *via* the axially oriented H2 and H4 atoms and possibly an H6 atom on the exocyclic CH_2OH group, or alternatively to the B face *via* the axially oriented H1, H3 or H5 atoms of the pyranose ring depicted by the cartoon in Fig. 3—and displayed for example, in the 3D structure of β-MeGlc at a tryptophan binding site in its complex with cellobiohydrolase from *Trichoderma reesei*,[22] also shown in Fig. 3.

The IR signature of the α-MeGlc · toluene complex is completely different from that of the β-complex. Only one structure is seen to be populated and its IRID spectrum is far more dispersed, displaying an intense and strongly displaced band at 3560 cm^{-1} and three bands located between 3613 cm^{-1} and 3642 cm^{-1}. The spectrum does not resemble either of the two lowest-lying conformers of the uncomplexed anomer, each of which presents an OH band located at ∼3600 cm^{-1} and a cluster of bands lying between 3625 cm^{-1} and 3645 cm^{-1}, associated with the H-bonded counter-clockwise orientation, OH4 → OH3 → OH2 → O1, see the structures in Fig. 4. The calculated band shifted to 3600 cm^{-1}, corresponding to $\sigma 2$, reflects the characteristic vibrational signature associated with the OH2 (equatorial) → O1 (axial) bonding in the α-anomers.[18] Strikingly, this band does not appear in the experimental spectrum of the complex.

The intense band at 3560 cm^{-1} indicates a specific, OH–π hydrogen bond—but which of the OH groups is involved? Two co-operatively bonded scenarios can be considered. In the first, the counter-clockwise orientation of the peripheral OH groups is retained to create the chain OH4 → OH3 → OH2 → π. The strongly displaced band at 3560 cm^{-1} would then correspond to the 'missing' OH2 band at 3600 cm^{-1} and the bands at 3613 cm^{-1} and 3630 cm^{-1} could be identified with OH3 and OH4, both displaced to lower wavenumber through co-operativity. Alternatively, if the counter-clockwise orientation were reversed in the complex, as it is, for example, in the mono-hydrated complex of β-PhGlc,[23] to create the chain sequence OH2 → OH3 → OH4 → π the strongly displaced band would then correspond to OH4. The band associated with OH2, released from its bonding to O1, would then shift from ∼3600 cm^{-1} towards higher wavenumber, because of a weaker acquired (equ → equ) interaction with OH3. Thus the issue of identity remains open although a comparison between the IR signatures of the complexes of α-MeGlc and α-MeGal, see below, provides evidence favouring OH4. It should be possible to resolve the issue unequivocally through ^{18}O labeling, either of OH2 or OH4. In either case, the most strongly shifted vibration would correspond to the final OH → π stage of a co-operatively linked H-bonded chain. Indeed there are strong analogies between the spectral shifts associated with OH → π bonding that have been observed in $(CH_3OH)_{n=1-3}$ · benzene clusters[24] and those observed here. As the length of the co-operatively bonded linear $(CH_3OH)_n$ chain increased, the wavenumber of the IR band associated with the terminal π-bonded OH decreased from 3681 cm^{-1} (n = 0) to 3639 cm^{-1} (n = 1) to 3605 cm^{-1} (n = 2) to 3589 cm^{-1} (n = 3).[24] In the current carbohydrate complex, albeit with toluene rather than benzene, an even greater shift could be anticipated consistent with the appearance of a band at 3560 cm^{-1}.

One question remains: why is the behaviour of the α- and β-anomers so different? The switch in the orientation of the anomeric OCH_3 group from equatorial (β) to axial (α) evidently alters the character of the most favoured molecular interaction in the isolated molecular complex from dispersive to H-bonded. In the α-anomer, the axial anomeric OCH_3 group, directed towards the B-face, and the equatorial orientation of CH1 would both reduce the 'attractiveness' of binding through multiple CH → π interactions on the B-face by simply reducing the number of appropriately oriented C–H groups. Perhaps it is this that tips the balance in favour of the alternative interaction provided by co-operatively enhanced, OH–π bonding.

† The designations, ccG-g+, … , indicate the conformation of the carbohydrate: briefly, 'cc' indicates a counter-clockwise orientation of the peripheral OH groups, OH4 → OH3 → OH2 → O1 and G-g+ … indicates the *gauche* orientation of the exocyclic hydroxymethyl group and its terminal OH6 group, respectively.[17]

3.2 MeGal and MeFuc

The switch from MeGlc to MeGal changes the configuration at C4 and the OH4 group is now directed axially, enhancing the interaction OH4 (axial) → OH3 (equ). MeFuc (6-deoxy-L-galactose) similarly contains an axial OH4. Like MeGlc, the lowest energy conformers of isolated MeGal and MeFuc (and also the corresponding phenyl pyranosides[17]) present counterclockwise oriented OH chains, OH4 → OH3 → OH2, further promoted in the α-anomers, by a strong OH2 → O1 interaction.

The IRID spectra of the α-MeFuc · toluene and α-MeGal · toluene complexes are shown in Fig. 5 together with the computed IR spectra of the isolated carbohydrates in their lowest energy conformations (calculated to lie at energies ~7 kJ mol^{-1} and ~4 kJ mol^{-1} below those of their nearest neighbours). The α-MeFuc complex displays two distinct, though very similar structural isomers, identified through hole-burning experiments, but only one is observed for α-MeGal. In every case however, the bound, and the corresponding free carbohydrate spectra are almost identical and the binding must be associated with 'structurally benign' interactions that conserve the conformation of the carbohydrate, namely CH → π interactions—in sharp contrast to the α-MeGlc · toluene complex where the binding is provided by an OH → π interaction. Switching the orientation at C4 from equatorial to axial creates a more 'focused', denser apolar patch on the B face of the carbohydrate, associated with the three neighbouring C–H bonds, CH3, CH4 and CH5, which are located away from the axially oriented OCH_3 at C1. It also has the effect of amplifying binding in the OH chain rather more strongly due to the presence of an enhanced OH4 (ax) → OH3 (eq) bond, further supported in the α-anomer, by the OH2 (eq) → O1 (ax) bond. These effects would each favour binding at the B face mediated by CH3,4,5 → π interactions and decrease the likelihood of a specific OH → π interaction. Excitingly, the structures displayed in the protein-bound complexes of α-MeFuc with Lectin 1 (from *Ulex europaeus*),[26] and α-MeGal (with the *Winged Bean Acidic* Lectin[25] shown in Fig. 4), show a similar contact with this extended B face. In each case, the counterclockwise orientation OH4 → OH3 → OH2 → O1 is preserved but the hydroxymethyl group in α-MeGal adopts the Tg+ conformation, associated in the free carbohydrate with its second lowest energy conformer.

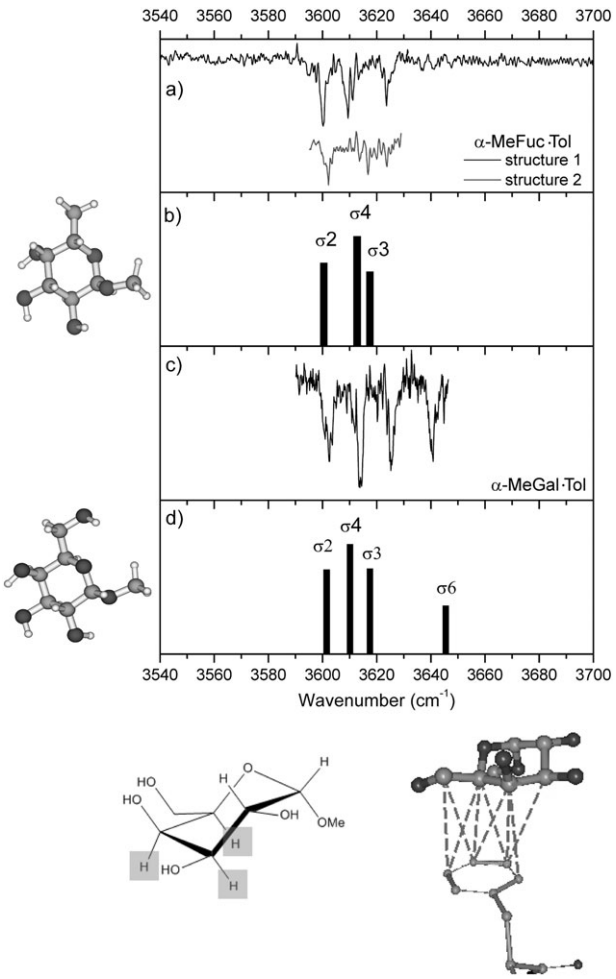

Fig. 5 IRID spectra of (a) α-MeFuc · toluene and (c) α-MeGal · toluene complexes. (b) and (d) show the predicted stick spectra for the energetically lowest lying conformers of uncomplexed α-MeFuc and α-MeGal together with the frequency assignments and their corresponding structures. Also shown: a suggested 'three point landing surface' for α-MeGal bound to toluene *via* a CH3,4,5–π interaction together with the local structure at a phenylalanine-containing carbohydrate binding site in *Winged Bean Acidic* lectin.[25] The dashed lines indicate the hydrophobic interactions between the sugar (CH3,4,5,6) and the aromatic residue.

Figs. 6 and 7 display the experimental IRID spectra of the β-MeFuc · toluene and β-MeGal · toluene complexes together with the computed and observed spectra of the lowest lying conformers of the uncomplexed methyl (and phenyl) pyranosides. Binding to toluene slightly perturbs the IR spectrum of β-MeFuc (for which only a single complex structure was observed) and the more weakly populated of the β-MeGal complexes, Fig. 7(b), where a second more strongly perturbed structure was also detected; unfortunately, because of UV spectral overlap it was not possible to separate its IRID spectrum fully from the first.

Consider first, the clearer experimental data for β-MeFuc · toluene, Fig. 6(a). If the band at highest wavenumber, 3641 cm^{-1} (the wavenumber associated with $\sigma 2$ in the β-anomer) is associated with the OH2 group, the two bands at 3595 cm^{-1} and 3620 cm^{-1} are most likely to correspond to OH4 and OH3, each displaced to lower wavenumber by ~15 cm^{-1} with respect to the corresponding bands in the uncomplexed anomer. While this could reflect a weak hydrogen-bonded interaction involving OH4 or OH3, *cf.* the α-MeGlc · toluene complex, where the perturbed band is much more strongly displaced, it might instead reflect some distortion of the conformation of the bound pyranose ring. In the β-anomer of MeFuc (and MeGal) the re-orientation at C1 extends the apolar patch on the B-face still further to include CH1 as well CH3, 4 and 5, and it also directs the OCH_3 group away from the B-face. In this situation binding to the B-face mediated by CH–π interactions would seem the more favoured option.

The behaviour of β-MeGal is more complex. The IRID spectrum of its more weakly populated complex (Fig. 7(b)) is

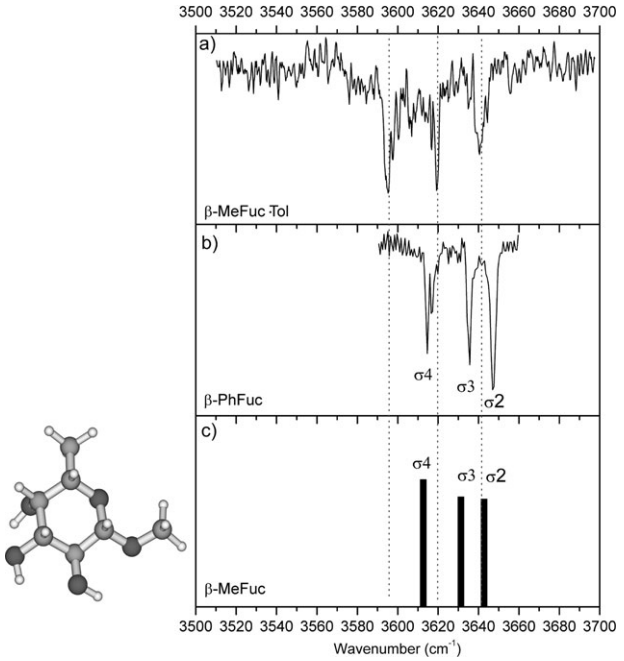

Fig. 6 IRID spectra of (a) the β-MeFuc · toluene complex, (b) uncomplexed β-PhFuc and (c) the calculated stick spectrum of β-MeFuc. Peak assignments are indicated for the experimental and predicted bands of the free molecules. The calculated structure of β-MeFuc shows a counter-clockwise orientation of the co-operative H-bonds. The dotted lines indicate the position of the peaks in the complex relative to the bare molecules.

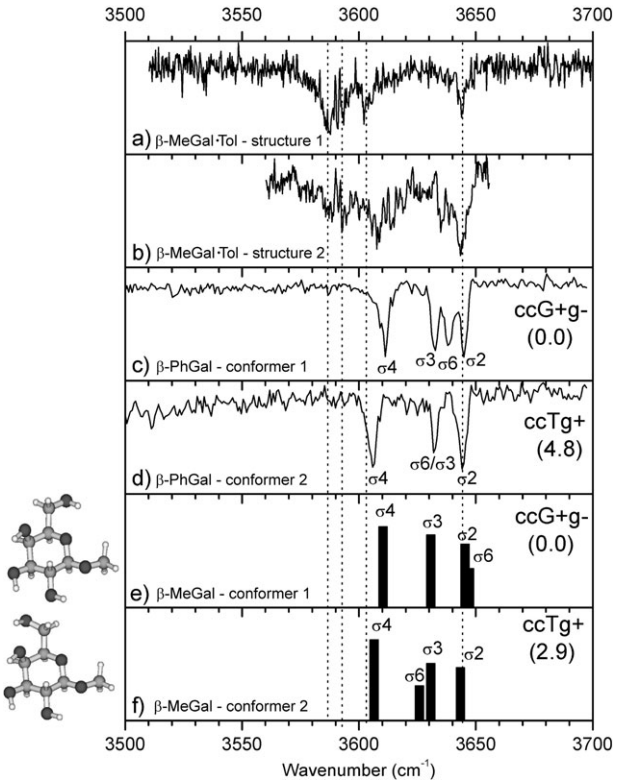

Fig. 7 (a) and (b) IRID spectra associated with two β-MeGal · toluene structures. (c) and (d) experimental IRID spectra and assignments based on previous calculation[17] for the energetically two lowest lying conformers of β-PhGal. (e) and (f)-predicted stick spectra for the lowest conformers of the uncomplexed β-MeGal. The dotted lines indicate the band positions in the more strongly populated complex. Calculated structures for the uncomplexed β-MeGal are shown in the bottom-left corner. Relative energies in kJ mol^{-1} are indicated in the brackets.

very similar to those of the two lowest lying conformers of the free carbohydrates, β-MeGal and β-PhGal (Fig. 7(c,d)) and there is no evidence of any contribution from OH–π bonding. On the other hand, the IRID spectrum associated with its more strongly populated complex, Fig. 7(a), is quite strongly perturbed. Three of its four bands are displaced by some 30 cm^{-1} towards lower wavenumber and a contribution from OH–π interaction(s) can no longer be discounted reflecting perhaps, an alternative orientation of the aromatic and carbohydrate partners.

3.3 Fuc

In sharp contrast to the complexes of MeFuc, the IRID spectrum of the Fuc · toluene complex displays by far the clearest signal of binding through a specific OH → π bonded interaction, with one of its four bands displaced to 3530 cm^{-1}, see Fig. 8. The obvious candidate for this is OH1, since the interaction OH1 → π would be unavailable (blocked by a methyl) in MeFuc. If the OH → π interaction lies at the terminus of the extended co-operative, counter-clockwise sequence, OH4 → OH3 → OH2 → OH1 → π this assignment would also be consistent with the large shift, ∼110–120 cm^{-1} below the corresponding band in the (computed) IR spectra of the α- or β-anomers of uncomplexed fucose. In β-MeGlc the sequence length corresponds to $n = 3$; in the analogous Fuc · toluene complex it would be $n = 4$. Co-operativity should promote corresponding but much smaller shifts in the remaining bands in the experimental spectrum. When comparisons are made with the remaining (computed) bands in the β-anomer, but not the α-anomer, the spectra do indeed reveal small shifts towards lower wavenumber.

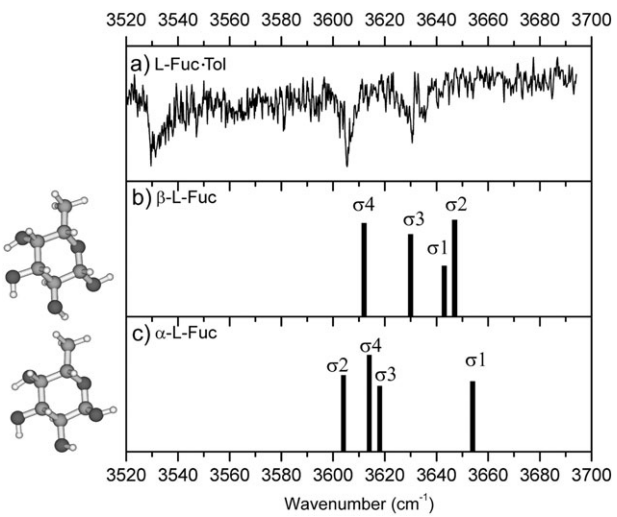

Fig. 8 IRID spectrum of L-Fuc · toluene (a) and predicted vibrational spectra, assignments and structures of uncomplexed β-L-Fuc (b) and α-L-Fuc (c).

4. Concluding remarks

The experimental programme set out to establish the generality of intermolecular binding in carbohydrate-aromatic complexes; to identify and characterize some of the interactions which promote it, particularly the operation of hydrophobic dispersion interactions *vs.* specific hydrogen bonded interactions; and to explore their sensitivity to structural change. The first aim has been achieved—at least for the 'basis set' of carbohydrates explored to date. Each one is seen to bind to toluene at low temperatures—there are without doubt, potential energy minima which support bound complexes. Since the carbohydrate-aromatic complexes were formed at low temperatures in the absence of water, entropic contributions associated with the reorganization of surrounding water molecules, the essence of the proposed hydrophobic interaction[11] cannot contribute—which is not to say that they might not do so in aqueous media at more elevated temperatures.

In most of the cases explored, dispersion forces involving CH–π interactions, which promote little if any distortion of the bound carbohydrate, predominate though complexes bound through specific OH–π H-bonded interactions have also been identified.

Qualitative attempts to correlate the balance of their relative contributions with the conformational structures of the bound carbohydrates, or the choice of possible orientations with respect to the aromatic platform, have only been partially successful. The factors which may favour localized H-bonding over the more broadly based dispersion interaction and which determine the global minimum energy structures of the isolated complexes are delicately balanced. In a 'real life' situation, where a carbohydrate is bound at an aromatic protein molecular recognition site, other interactions with the protein backbone or neighbouring residues (determined by the environment and protein architecture) will also be present. Given the delicacy of the local π-electron interactions it should not be difficult for other neighbouring interactions to modulate the relative carbohydrate-aromatic dispositions and orientations at individual binding sites. Indeed a survey of the deposited crystallographic structures of carbohydrate (*e.g.* Gal or Glc)– protein complexes listed in the Brookhaven PDB[14] reveals a broad spread of sugar-aromatic interactions.[13] These suggest, at least in part, that the primary role of the aromatic platform is the provision of an attractive but malleable contribution to the net intermolecular potential energy that can be modulated by the local protein environment to provide the remarkable selectivity displayed by carbohydrate recognition proteins. Attempts to calculate the interactions accurately through *ab initio* computation[7,12,19,27] will need to be conducted at high levels of theory if they are to provide quantitative agreement with experimental spectroscopic or structural data and DFT methods will certainly fail without the inclusion of dispersion interactions.[28,29]

There are many other experimental approaches that may now be explored. The nature of hydrophobic interactions—or water-mediated binding—can be investigated by creating hydrated aromatic complexes in a molecular beam, passing them through a plume of carbohydrate vapour generated from an oven or by laser ablation, and recording the vibrational spectra of the resulting cluster(s). CH–π interactions can be explored directly, by recording the vibrational spectra of bound and free carbohydrates in the C–H region; deuteration will shift the vibrational spectrum of the aromatic partner 'out of sight'.[30] Assignment of specific OH–π interactions can be facilitated by recording the IRID spectra of ^{18}O labeled carbohydrates. Moving beyond the simplest model systems, the role of the phenolic OH group in tyrosine, which is more common in protein binding sites than phenylalanine, can be explored by substituting *p*-cresol for toluene. Does the phenolic OH provide a hydrogen-bonding 'anchor' to help bind the carbohydrate to the aromatic platform? 3-methyl indole provides a truncated model for tryptophan. The influence of back-bone interactions may be explored by substituting a suitably protected phenylalanine or tyrosine for toluene. Chiral discrimination can be explored by comparing the complexes of their D and L enantiomers with D and L sugars. The carbohydrate partners need not be restricted to *mono*saccharides. There is much to be done and much to learn.

Acknowledgements

We have appreciated helpful discussions and contributions made by Drs N. A. Macleod, T. van Mourik and Dr J. Frey. Financial support has been provided by the EPSRC, the Leverhulme Trust (Grant F/08788G) and the Royal Society (LCS, University Research Fellowship). Support from the CLRC Laser Support Facility and Corpus Christi College (LCS) is also acknowledged.

References

1 W. I. Weis and K. Drickamer, *Annu. Rev. Biochem.*, 1996, **65**, 441.
2 R. A. Dwek, *Chem. Rev.*, 1996, **96**, 683.
3 A. Varki, *Glycobiology*, 1993, **3**, 97.
4 A. B. Boraston, D. Nurizzo, V. Notenboom, V. Ducros, D. R. Rose, D. G. Kilburn and G. J. Davies, *J. Mol. Biol.*, 2002, **319**, 1143.
5 A. B. Boraston, D. N. Bolam, H. J. Gilbert and G. J. Davies, *Biochem. J.*, 2004, **382**, 769.
6 H. Lis and N. Sharon, *Chem. Rev.*, 1998, **98**, 637.
7 J. Jiménez-Barbero, J. L. Asensio, F. J. Cañada and A. Poveda, *Curr. Opin. Struct. Biol.*, 1999, **9**, 549.
8 R. Palma, M. E. Himmel and J. W. Brady, *J. Phys. Chem. B*, 2000, **104**, 7228.
9 M. C. Fernández-Alonso, F. J. Cañada, J. Jiménez-Barbero and G. Cuevas, *J. Am. Chem. Soc.*, 2005, **127**, 7379.
10 X. Robert, R. Haser, T. E. Gottschalk, F. Ratajczak, H. Driguez, B. Svensson and N. Aghajari, *Structure*, 2003, **11**, 973.
11 R. U. Lemieux, *Acc. Chem. Res.*, 1996, **29**, 373.
12 V. Spiwok, P. Lipopová, T. Skálová, E. Vondráčková, J. Dohnálek, J. Hašek and B. Králová, *J. Comput.-Aided Mol. Des.*, 2006, **19**, 887.
13 M. S. Sujatha, Y. U. Sasidhar and P. V. Balaji, *Biochemistry*, 2005, **44**, 8554.
14 The protein data bank: url: www.rcsb.org.
15 G. Terraneo, D. Potenza, A. Canales, J. Jiménez-Barbero, K. K. Baldridge and A. Bernardi, *J. Am. Chem. Soc.*, 2007, **129**, 2890.
16 J. Screen, E. C. Stanca-Kaposta, D. P. Gamblin, N. A. Macleod, L. C. Snoek, B. G. Davis and J. P. Simons, *Angew. Chem., Int. Ed.*, 2007, **46**, DOI: 10.1002/anie.200605116.
17 J. P. Simons, R. A. Jockusch, P. Çarçabal, I. Hünig, R. T. Kroemer, N. A. Macleod and L. C. Snoek, *Int. Rev. Phys. Chem.*, 2005, **24**, 489.

18 P. Çarçabal, T. Patsias, I. Hünig, B. Liu, C. Kaposta, L. C. Snoek, D. P. Gamblin, B. G. Davis and J. P. Simons, *Phys. Chem. Chem. Phys.*, 2006, **8**, 129–136.
19 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. Montgomery, J. A. T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004.
20 G. Chang, W. C. Guida and W. C. Still, *J. Am. Chem. Soc.*, 1989, **111**, 4379.
21 F. O. Talbot and J. P. Simons, *Phys. Chem. Chem. Phys.*, 2002, **4**, 3562.
22 C. Divne, J. Ståhlberg, T. T. Teeri and T. A. Jones, *J. Mol. Biol.*, 1998, **275**, 309: pdb 7CEL.
23 P. Çarçabal, R. A. Jockusch, I. Hünig, L. C. Snoek, R. T. Kroemer, B. G. Davis, D. P. Gamblin, I. Compagnon, J. Oomens and J. P. Simons, *J. Am. Chem. Soc.*, 2005, **127**, 11414.
24 R. N. Pribble, F. C. Hagemeister and T. S. Zwier, *J. Chem. Phys.*, 1997, **106**, 2145.
25 N. Manoj, V. R. Srinivas, A. Surolia, M. Vijayan and K. Suguna, *J. Mol. Biol.*, 2000, **302**, 1129: pdb 1F9K.
26 G. F. Audette, D. J. H. Olson, A. R. Ross, J. W. Quail and L. T. Delbaere, *Can. J. Chem.*, 2002, **80**, 1010: pdb 1JXN.
27 M. I. Chávez, C. Andreu, P. Vidal, N. Aboitiz, F. Freire, P. Groves, J. L. Asensio, M. Muraki, F. J. Cañada and J. Jiménez-Barbero, *Chem.–Eur. J.*, 2005, **11**, 7060.
28 M. Elstner, P. Hobza, T. Frauenheim, S. Suhai and E. Kaxiras, *J. Chem. Phys.*, 2001, **114**, 5149.
29 T. van Mourik, P. G. Karamertzanis and S. L. Price, *J. Phys. Chem. A*, 2006, **110**, 8.
30 S. Morita, A. Fujii, N. Mikami and S. Tsuzuki, *J. Phys. Chem. A*, 2006, **110**, 10583.