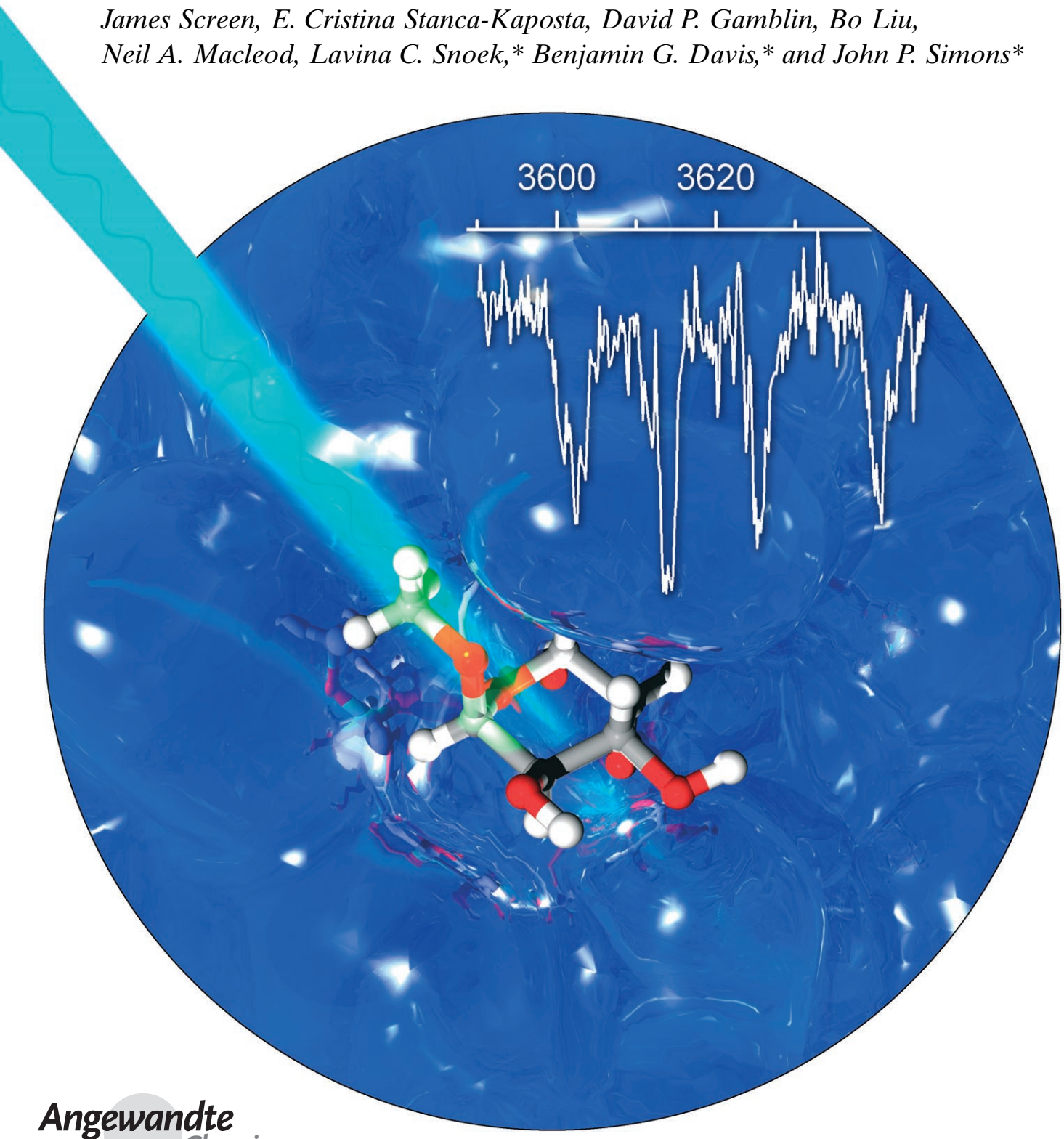
Sugar–Aromatic Complexes

DOI: 10.1002/anie.200605116

IR-Spectral Signatures of Aromatic–Sugar Complexes: Probing Carbohydrate–Protein Interactions**

James Screen, E. Cristina Stanca-Kaposta, David P. Gamblin, Bo Liu, Neil A. Macleod, Lavina C. Snoek, Benjamin G. Davis,* and John P. Simons**



Molecular interactions between carbohydrates and proteins—lectins and enzymes—are central to a broad range of biological processes.[1–3] Binding at carbohydrate-recognition domains can be mediated by hydrogen-bonding, van der Waals, and ionic interactions.[1–8] Despite the high hydrophilicity of most sugars, their interactions with proteins can include binding between aromatic amino acid residues and the apolar faces or "patches" of pyranose rings (Figure 1).[9] The planar aromatic platforms of tryptophan, tyrosine, and less commonly, phenylalanine are often exploited in nature in carbohydrate-binding modules as selective recognition points in stacking or "sugar-tong" structural motifs.[9,10]

The near-IR vibrational spectra of individual carbohydrate conformers, isolated under molecular-beam conditions in the gas phase, are extraordinarily sensitive to the local hydrogen-bond environment of their OH groups. A large shift of a band towards lower wavenumbers, or conversely, the absence of a significant shift, provides a sensitive diagnostic for the incidence or absence of specific OH···X intramolecular hydrogen-bonding interactions.[11] Molecular-beam experiments also provide an ideal "laboratory" for studying carbohydrate interactions within complexes, through a combination of IR spectroscopy and computational modeling.[11]

A key first step along the computational path to understanding such interactions was taken recently by Jiménez-Barbero and co-workers, who explored the nature of specific galactose–lectin interactions through a series of density functional theory (DFT) and ab initio calculations of β-L-fucose and benzene.[4,7] Here, we describe the first results of the experimental approach: a direct determination of the spectroscopic signatures of sugar–arene complexes to identify the presence or absence of specific hydrogen-bonding interactions and any structural changes in the carbohydrate promoted by complex formation. Toluene, chosen as the model aromatic partner, provides a surrogate for the side chains of phenylalanine and tyrosine—amino acids often associated with stacking and sugar-tong binding.[9] Representative carbohydrates chosen for initial investigation included

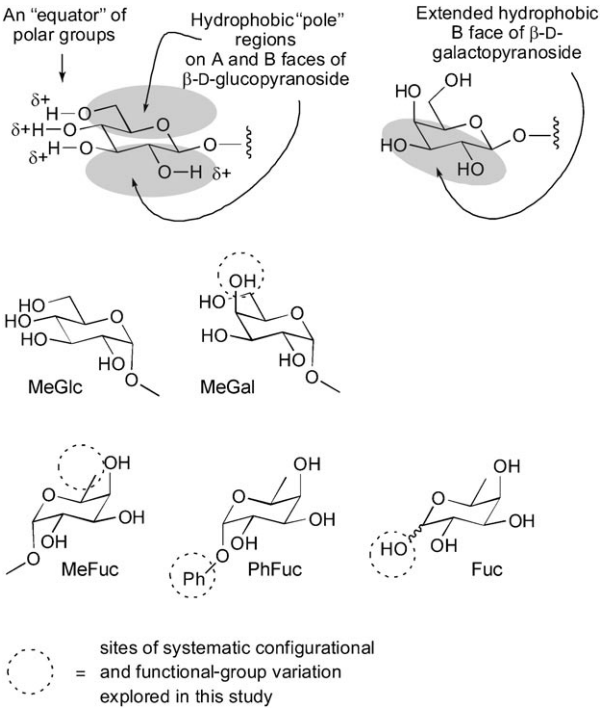


Figure 1. Representative apolar "patches" in carbohydrates and the monosaccharide derivatives chosen for this study.

the α-D-gluco- and -galactopyranosides MeGlc and MeGal, the α-L-fucopyranosides MeFuc and PhFuc, and, in the light of references [4,7], fucose (Fuc) itself (Figure 1). These were used to probe: the effects of varying the configuration at C4 and hence the "shapes" of apolar faces, for example, the B faces of MeGal and MeGlc; the structural influence of an OH group (OH6) more remote from the apolar face, by comparison of the toluene complexes of MeGal and MeFuc; and the influence of the aglycon, by comparison of the binding signatures in the complexes of Fuc, MeFuc, and PhFuc (Figure 2).

The UV spectrum of the MeFuc·Tol complex revealed two structures; their IR spectra (recorded by using the IR ion-dip (IRID) technique[12]) are shown in Figure 2, together with those of PhFuc·Tol, PhFuc,[11] and MeFuc. In the uncomplexed fucosides only the global-minimum conformation is populated; its structure presents a counterclockwise (cc) orientation of the hydrogen-bonded chain (OH4→OH3→OH2→O1). None of the bands in the MeFuc·Tol complex is displaced towards lower wavenumbers, implying bonding through CH···π interactions alone; complexing with the aromatic ring cannot significantly affect the structure of the pyranose unit to which it is attached.

The IRID spectrum of the complex between Fuc and toluene is shown in Figure 3 together with the (computed) spectra of the α and β anomers of uncomplexed Fuc. The Fuc·Tol complex clearly displays a strongly shifted band located at 3530 cm^{-1}, approximately 75 cm^{-1} lower in energy than the next band. This is in sharp contrast with the MeFuc·Tol complex, where any OH1···π interaction is blocked and the binding can be attributed solely to dispersive CH···π

[*] J. Screen, Dr. E. C. Stanca-Kaposta, Dr. B. Liu, Dr. N. A. Macleod, Dr. L. C. Snoek, Prof. J. P. Simons
Department of Chemistry, University of Oxford
Physical and Theoretical Chemistry Laboratory
South Parks Road, Oxford, OX1 3QZ (UK)
Fax: (+44) 186-528-5002
E-mail: Lavina.Snoek@chem.ox.ac.uk
John.Simons@chem.ox.ac.uk
Dr. D. P. Gamblin, Prof. B. G. Davis
Department of Chemistry, University of Oxford
Chemistry Research Laboratory
Mansfield Road, Oxford, OX1 3TA (UK)
Fax: (+44) 186-527-5674
E-mail: Ben.Davis@chem.ox.ac.uk

[**] Financial support was provided by the EPSRC (Grant GR/T26542), the Leverhulme Trust (Grant F/08788G), and the Royal Society (L.C.S., University Research Fellowship). We also acknowledge support from the CLRC Laser Support Facility, the Physical and Theoretical Chemistry Laboratory, and Corpus Christi College, Oxford (L.C.S.).

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

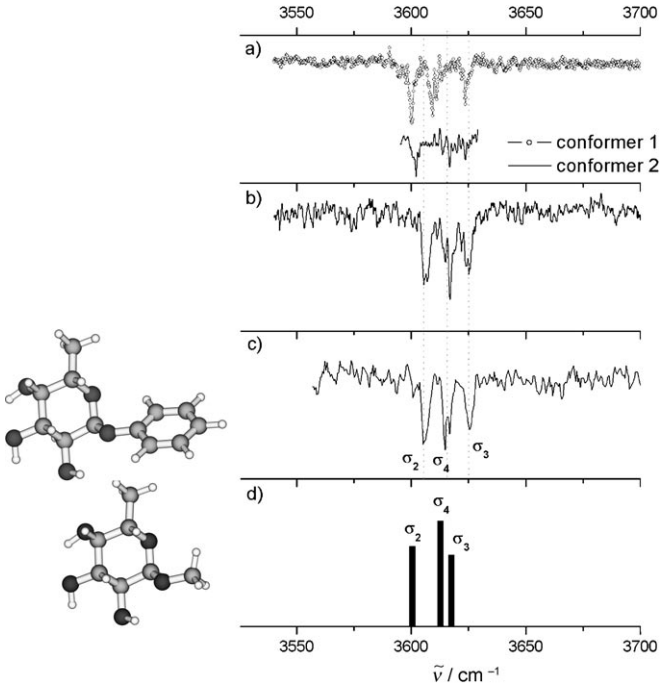


Figure 2. IRID spectra of a) MeFuc·Tol, b) PhFuc·Tol, and c) PhFuc, together with the calculated IR spectrum of d) "free" MeFuc. The dotted lines indicate the positions of the vibrational bands in the (phenyl-tagged) uncomplexed saccharide, PhFuc, that are associated with the three OH stretching modes σ_2, σ_3, and σ_4. The dip in the central band at 3615 cm^{-1} is due to the absorption of atmospheric water which reduces the intensity of the IR laser. The ball-and-stick models next to (c) and (d) correspond to PhFuc and MeFuc, respectively. These and the other structures shown in this Communication were generated by DFT calculations.

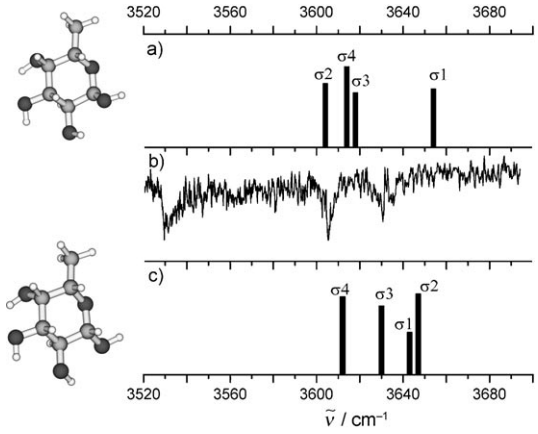


Figure 3. a, c) Computed IR spectra of αFuc (a) and βFuc (c), and b) the experimental IRID spectrum of the complex Fuc·Tol. The computed structures and conformations of αFuc and βFuc are shown on the left.

interactions. The remaining vibrational structure at high wavenumbers corresponds quite well with that calculated for the uncomplexed β anomer, suggesting the occurrence of selective binding, most probably through an OH1–π hydrogen bond. The displaced band at 3530 cm^{-1} can then be assigned to the vibrational mode σ1, and the remaining bands—σ2, σ3, and σ4—to the weak features lying between 3600 cm^{-1} and 3640 cm^{-1}.

This identification of a hydrogen-bonding OH–π interaction in the βFuc·Tol complex contrasts with the computed CH–π interaction in the complex with benzene.[4,7] These computations were initiated from starting geometries based upon crystal structures of related galactoside–protein complexes taken from the protein database. The computations generated stable bound complexes, provided dispersive interactions were taken into account.[7] NMR spectra of methyl β-galactoside in aqueous solutions of benzene or phenol were consistent with the retention of similar structures in the condensed phase, and there was no evidence for specific hydrogen-bonding interactions involving OH groups.[7] In the context of this apparent conflict with our observations it should be noted that restrictions in the calculations of βFuc·benzene imposed a key starting topology and the experimental NMR observations utilized the glycoside MeβGal, in which OH1 is unavailable. Nonetheless, the prevalence of the fucoside rather than free fucose in nature does highlight Fuc as an anomalous carbohydrate for study.

Unlike galactosides and fucosides, in α-methyl glucoside (MeGlc) all of the secondary OH groups (OH2–OH4) are equatorial, potentially allowing access to two parallel hydrophobic faces. In protein complexes in which hydrophobic interactions between glucose and aromatic residues have been identified, both of the hydrophobic (A and B) faces are presented to aromatic protein side chains.[13] The IRID spectrum of the MeGlc·Tol complex is presented in Figure 4 together with the computed spectra of each of the three lowest-lying conformers of isolated αMeGlc. The isolated glucoside presents a weakly hydrogen-bonded counterclockwise chain (OH4→OH3→OH2→O1) in its three lowest-lying conformations. In the complex, one of the vibrational bands is significantly displaced to a lower wavenumber (approximately 3560 cm^{-1}), indicating a hydrogen-bonding OH–X intermolecular interaction, despite the blocking of the anomeric hydroxy group, OH1. The bands observed at 3631 cm^{-1} and 3642 cm^{-1} lie close to those computed for the OH modes σ3 and σ6. The absence of a band around 3600 cm^{-1}, associated in isolated, counterclockwise-oriented α anomers with the vibration σ2,[11] could suggest the possibility of hydrogen bonding via OH2, in which case the feature at 3613 cm^{-1} would correspond to σ4. Alternatively, if the observed H-bonding involved OH4, these assignments would need to be reversed. The band at highest wavenumber (3642 cm^{-1}) is characteristic of a noninteracting hydroxymethyl group which excludes binding via OH6.

Figure 5 compares the IR spectra associated with MeFuc·Tol and MeGal·Tol with the (computed) IR spectra of the lowest-lying conformers of MeGal·Tol and MeGal alone. Like the fucoside (6-deoxy-L-galactoside), none of the vibrational bands of MeGal were significantly shifted towards lower wavenumbers when the sugar was bound to toluene, including σ6, so replacement of the exocyclic methyl (in MeFuc) by a hydroxymethyl group (in MeGal) did not lead to a hydrogen-bonding interaction with toluene through OH6. As with MeFuc, the spectral fingerprint implies a complex bound solely by dispersive CH–π interactions. The pattern displayed

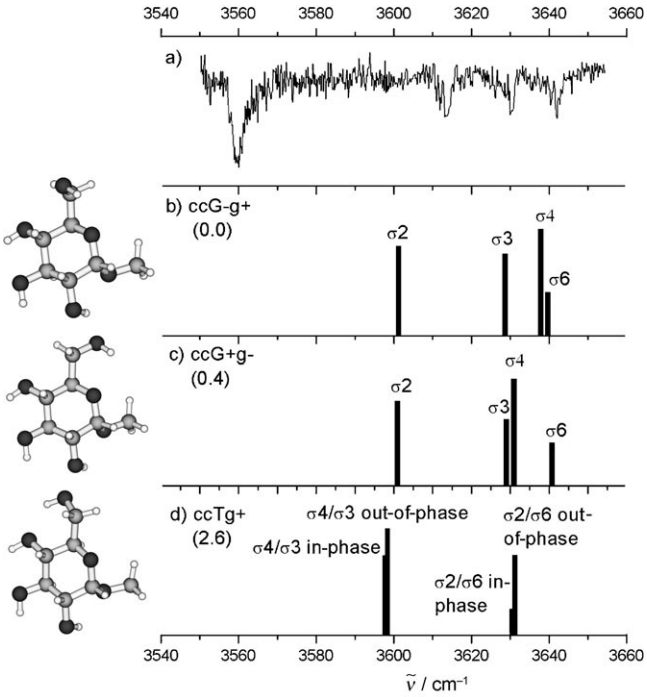


Figure 4. a) IRID spectrum of the complex between MeGlc and toluene recorded experimentally. b–d) Computed IR spectra for the three lowest-lying conformers of the isolated carbohydrate; computed structures of the conformers of MeGlc are shown on the left, and the relative energies (in brackets) are given in kJ mol^{-1}. The designation cc refers to the counterclockwise orientation of the peripheral OH groups; the designations G-g+, G-g-, and Tg+ indicate the orientation of the exocyclic hydroxymethyl and OH6 groups.

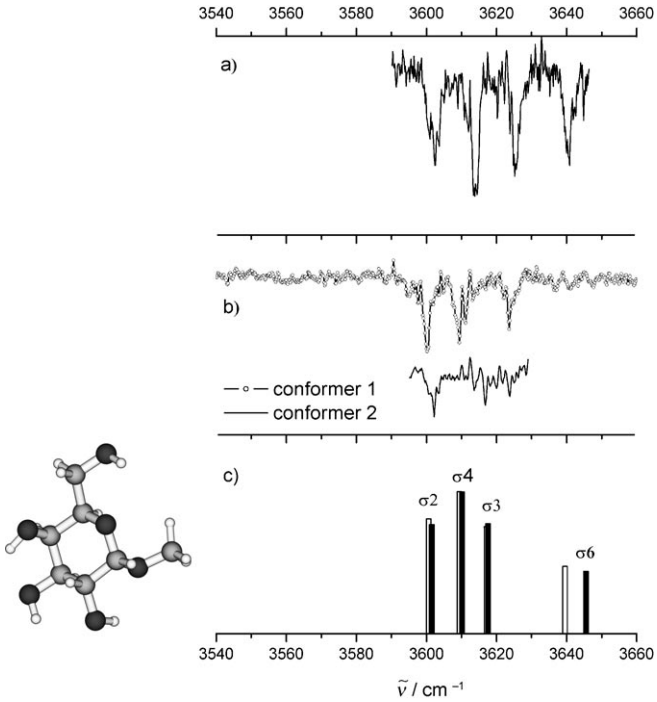


Figure 5. IRID spectra of a) MeGal·Tol and b) MeFuc·Tol; c) the computed IR spectra of MeGal·Tol (unfilled bars) and uncomplexed MeGal (filled bars); the computed conformation of uncomplexed MeGal is shown as a ball-and-stick model.

by the remaining bands reflects retention of the weakly hydrogen-bonded counterclockwise chain, with the band at lowest wavenumber corresponding to the vibration of OH2 (σ2), characteristic of an α anomer.[11] Since none of the bands were significantly perturbed by complex formation, the conformation adopted by the bound monosaccharide is likely to be very similar to that preferred by isolated MeGal.

Initial computational analysis, including single-point MP2 counterpoise corrections, has allowed us to suggest structures for some of these complexes. Analysis of the Tol·MeGal complex reveals a lowest-lying complex (1 kJ mol^{-1} lower in energy than the next) with a clear π–A face interaction (Figure 6). Excitingly, this complex bears a striking similarity to that observed directly in the 3D structure of MeGal bound to the galactose-specific lectin from *Artocarpus hirsute*.[14] This confirms, at least in part, the validity of the approach described here for predicting and understanding biologically relevant hydrophobic sugar–protein interactions.

These investigations were designed to explore and characterize local interactions involved in the molecular recognition of carbohydrates by proteins through binding to aromatic residues. The possibility of forming stable complexes involving dispersive "hydrophobic" CH–π interactions has been established for α-methyl galactoside and its close analogue, α-methyl fucoside. In fucose itself where the anomeric hydroxyl group OH1 is no longer blocked and which is rarely alone as a ligand in nature, the spectra imply a

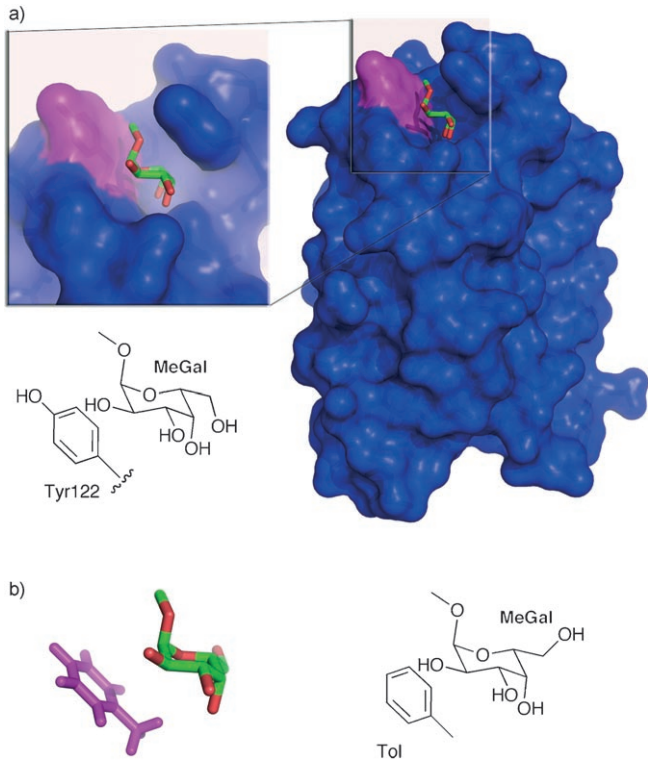


Figure 6. The closely related A-face-binding mode of MeGal in a) the galactose-specific lectin from *Artocarpus hirsute* and b) the DFT-computed structure of MeGal·Tol.

complex bound through an OH1–π interaction. A little surprisingly, a hydrogen-bonding OH–π interaction also operates in the complex with α-methyl glucoside, where OH4 is oriented equatorially rather than axially; in this case the hydrogen-bonding has been associated with OH2 or OH4. Switching the orientation of OH4 from axial to equatorial (MeGal→MeGlc) clearly tips the balance in favor of intermolecular hydrogen bonding. The balance between hydrogen-bonding and dispersive interactions is therefore a delicate one, sensitive to the configuration of the carbohydrate. In the complexes bound exclusively by dispersive CH–π interactions the global-minimum carbohydrate structures are broadly retained. Initial modeling of an arene complex of α-methyl galactoside suggests that this closely mimics a known, observed binding mode in protein–sugar interactions. This and our other results suggest that binding to aromatic apolar platforms does not seem to necessitate large conformational penalties which may in part explain their widespread use by nature, as carbohydrate-binding modules.

Experimental Section

Methyl-α-L-fucopyranoside was obtained in a near-quantitative fashion following a standard Fischer glycosylation from fucose, using a methanolic solution of HCl generated in situ from acetyl chloride.

A detailed description of the molecular-beam experiment has been published previously.[11]

Quantum chemical calculations were performed using the Gaussian 03 package.[15] Geometry optimizations and calculations of harmonic vibrational frequencies 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. Zero-point corrections to the electronic energy were calculated using the harmonic frequencies.

Received: December 19, 2006
Published online: March 27, 2007

Keywords: carbohydrates · density functional calculations · molecular recognition · vibrational spectroscopy

[1] W. I. Weiss, K. Drickamer, *Annu. Rev. Biochem.* **1996**, *65*, 441–473.
[2] R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683–720.
[3] A. Varki, *Glycobiology* **1993**, *3*, 97–130.
[4] J. Jiménez-Barbero, J. L. Asensio, F. J. Cañada, A. Poveda, *Curr. Opin. Struct. Biol.* **1999**, *9*, 549–555.
[5] H. Lis, N. Sharon, *Chem. Rev.* **1998**, *98*, 637–674.
[6] R. Palma, M. E. Himmel, J. W. Brady, *J. Phys. Chem. B* **2000**, *104*, 7228–7234.
[7] M. C. Fernández-Alonso, F. J. Cañada, J. Jiménez-Barbero, G. Cuevas, *J. Am. Chem. Soc.* **2005**, *127*, 7379–7386.
[8] M. I. Chávez, C. Andreu, P. Vidal, N. Aboitiz, F. Freire, P. Groves, J. L. Asensio, G. Asensio, M. Muraki, F. J. Cañada, J. Jiménez-Barbero, *Chem. Eur. J.* **2005**, *11*, 7060–7074.
[9] A. B. Boraston, D. N. Bolam, H. J. Gilbert, G. J. Davies, *Biochem. J.* **2004**, *382*, 769–781.
[10] X. Robert, R. Haser, T. E. Gottschalk, F. Ratajczak, H. Driguez, B. Svensson, N. Aghajari, *Structure* **2003**, *11*, 973–984.
[11] P. Çarçabal, T. Patsias, I. Hünig, B. Liu, C. Kaposta, L. C. Snoek, D. P. Gamblin, B. G. Davis, J. P. Simons, *Phys. Chem. Chem. Phys.* **2006**, *8*, 129–136.
[12] E. G. Robertson, J. P. Simons, *Phys. Chem. Chem. Phys.* **2001**, *3*, 1–18.
[13] A. B. Boraston, D. Nurizzo, V. Notenboom, V. Ducros, D. R. Rose, D. G. Kilburn, G. J. Davies, *J. Mol. Biol.* **2002**, *319*, 1143–1156.
[14] K. N. Rao, C. G. Suresh, U. V. Katre, S. M. Gaikwad, M. I. Khan, *Acta Crystallogr. Sect. D* **2004**, *60*, 1404–1412.
[15] M. J. Frisch, et al., see the Supporting Information.

