



Carbohydrate–aromatic interactions: A computational and IR spectroscopic investigation of the complex, methyl α -L-fucopyranoside · toluene, isolated in the gas phase

Zheng Su^a, Emilio J. Cocinero^a, E. Cristina Stanca-Kaposta^{a,1}, Benjamin G. Davis^b, John P. Simons^{a,*}

^a Department of Chemistry, Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford OX1 3QZ, UK

^b Department of Chemistry, Chemistry Research Laboratory, Mansfield Road, Oxford OX1 3TA, UK

ARTICLE INFO

Article history:

Received 22 January 2009

In final form 13 February 2009

Available online 20 February 2009

ABSTRACT

A carbohydrate–aromatic complex, methyl α -L-fucopyranoside · toluene, which provides a model for probing the physical basis of carbohydrate–protein ‘stacking’ interactions, has been created in a molecular beam and probed through IR ion dip spectroscopy in the CH and OH regions. The results are interpreted in the light of DFT calculations using the MO5-2X functional. They indicate the creation of stacked structures with the aromatic molecule bonded either to the upper or to the lower face of the pyranoside ring, through $\text{CH}_{3,4}$ - π (upper) or CH_1 - π (lower) interactions leading to binding energies $\leq 18 \text{ kJ mol}^{-1}$.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Molecular recognition of carbohydrates by proteins often involves their selective binding at sites adjacent to aromatic amino acid residues, tryptophan, tyrosine or phenylalanine, where they can adopt a stacking motif with the aromatic side group providing a platform for the bound ligand [1,2]. This has been associated with ‘apolar CH- π bonding’ [2] which can supplement hydrogen bonded interactions involving the polar OH groups on the carbohydrate and neighbouring side groups in the local protein environment, or bonding that may be mediated through electrostatic interactions with nearby ions, such as Ca^{2+} [3–9]. Protein–carbohydrate interactions are fundamentally important in a wide array of biological organisms and many processes such as infection, fertilization, inflammation, and cellular recognition, depend upon them [10,11].

CH- π interactions in protein or oligopeptide–carbohydrate complexes have been probed experimentally through X-ray crystallographic investigations and also through NMR spectroscopy in aqueous solution. ‘Hydrophilic’ OH- π hydrogen bonded and ‘hydrophobic’ CH- π interactions have been inferred from the crystallographic structural data, and the perturbation of (C)-H hydrogen nuclei by aromatic ring currents, determined through proton NMR measurements in solution, has been used to identify specific CH- π interactions [12], though their nature is less easy to define. [13–17] IR vibrational spectroscopy of O-H modes has been used in smaller model carbohydrate–aromatic systems isolated in the

gas phase and free of neighbouring water molecules, to provide a more direct experimental probe since their frequencies and intensities are extraordinarily sensitive to their local inter (and also intra) molecular hydrogen bonded environments [18–20]. The incidence of OH- π hydrogen bonding in a series of carbohydrate–toluene complexes stabilized in a cold molecular beam [19,20], was signaled by a red-shift, broadening and intensification of one (or more) of the O-H bands. On the other hand, the *absence* of any significant displacement of the OH modes in the aromatic complex was taken as an indirect indication of carbohydrate binding through ‘CH- π ’ interactions, provided the OH- π interaction was stronger than any pre-existing OH-O interaction with a neighbouring OH group; (the operation of CH- π interactions need not necessarily preclude a minor contribution from OH- π bonding as well).

To date, similar spectroscopic measurements of shifts in the IR frequencies of the C-H modes between free and bound carbohydrates, which might be used to ‘report’ directly on the incidence of CH- π bonding in their aromatic complexes, have not been undertaken. The observation of specific displacements might, with the aid of predictive quantum chemical calculations, help in their structural assignment since the magnitude and direction of the spectral shifts that can be associated with CH- π bonding have been intensively discussed in recent years, particularly the information they may convey about the nature of the interaction [13–17,21,22].

Selection of the appropriate level of theory that should be used has to take into account the role that dispersion plays [23] in the binding of carbohydrate–aromatic complexes and also the computational expense. The latter can be greatly reduced by using density functional theory (DFT); the former can be achieved within the

* Corresponding author.

E-mail address: john.simons@chem.ox.ac.uk (J.P. Simons).

¹ Present address: Institut für Experimentalphysik, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany.

framework of DFT theory by using for example, the dispersion corrected version, D-DFT [23–25], which adds a weighted empirical pair-wise potential of the form $\sum_{ij} C_{ij}^d/R_{ij}^6$, or by calculations which employ a functional that takes proper account of electron exchange and correlation [26]. The former approach was used very recently by Hillier's group [25] following our earlier experimental, molecular beam investigations, to calculate the vibrational spectra (OH and CH stretch modes) and binding energies of a number of carbohydrate–aromatic complexes, two of which could be compared with our IR spectra recorded in the OH region [18,19]. Their D-DFT calculations, employing a BLYP functional, were in broad agreement with the experimental spectra of the fucose · toluene and α -methyl glucopyranoside · toluene complexes, predicting OH red-shifts that were comparable with those that we had observed and had associated with OH– π hydrogen bonded structures. They also predicted (surprisingly) large CH blue shifts associated with (calculated) complex structures that were bound through CH– π interactions, though spectroscopic measurements in the CH stretch region were not then available.

Two complexes which had previously been associated with CH– π , rather than OH– π bound structures [18,19] because of the absence of any significant change in the OH spectrum of the bound carbohydrate, included methyl α -D-galactopyranoside · toluene (MeGal · Tol) and its 6-deoxy analogue, methyl α -L-fucopyranoside · toluene (MeFuc · Tol), see Fig. 1; (the enantiomeric choices reflect Nature's preferences). Remarkably, preliminary DFT calculations of MeGal · Tol employing the B3LYP functional, which has a good 'track record' in predicting the frequencies of OH (and NH) bands within hydrogen bonded carbohydrate structures [20], predicted a CH– π bound structure as the lowest in energy [18] despite the functional being unable to provide a proper description of complexes bound predominantly by dispersion interactions.

Encouraged by this result, the earlier experiments have now been extended using MeFuc · Tol as a benchmark system and toluene- d_8 , to obtain the IR spectrum of the bound carbohydrate in the CH stretch region without spectral interference from its aromatic partner. At the same time further DFT calculations have been conducted for both the isolated and the complexed carbohydrate, using Zhao and Truhlar's MO5-2X functional [26] and the predictions have been compared with experiment, both for the OH spectra reported earlier [18,19] and for the newly recorded CH spectra.

2. Experimental

2.1. Molecular beam spectroscopy

The essential details of the experimental procedure were similar to those reported previously [19]. Powdered samples of the carbohydrate were ground with graphite powder in a ratio

~80% sample: 20% graphite (by volume), deposited as a thin uniform surface layer on a graphite substrate, and placed in the vacuum chamber close to the exit of a pulsed, cylindrical nozzle expansion valve (0.8 mm diameter). Molecules were desorbed from the surface using the fundamental of a pulsed and focused Nd:YAG laser (~4 mJ/pulse) and entrained and cooled in an expanding argon jet (~4 bar backing pressure) before passing into the detection chamber through a 2 mm diameter skimmer. The aromatic complexes were formed by seeding the carrier gas with fully deuterated toluene (99.6 atom% D, Sigma-Aldrich) vapor prior to the expansion in ~0.25% in Ar. Mass-selected R2PI spectra were recorded using a frequency-doubled pulsed Nd:YAG-pumped dye laser (Continuum Powerlite II/Sirah PS-G, 1–2 mJ/pulse UV) operating at 10 Hz. Conformer-specific CH vibrational spectra were recorded at the parent ion mass and selected R2PI absorption bands of the complex, through IR ion dip (IRID) double resonance spectroscopy [27] using radiation in the range 2800–3150 cm^{-1} , line-width ~2–3 cm^{-1} , 5 mJ/pulse, generated by the idler output of an OPO/OPA laser system (LaserVision), pumped by a pulsed Nd:YAG laser (Surelite II).

2.2. Computation

The structural search began with an unrestricted survey of the many possible MeFuc · Tol- d_8 structures, using the Monte Carlo multiple minimization procedure as implemented in the MacroModel software (MacroModel v.8.5, Schrödinger, LLC21). [28] Subsequent B3LYP/6-31+G* or MO5-2X/6-31+G** geometry optimization, led ultimately to the identification of eight structures with energies lying <10 kJ mol^{-1} above the minimum – calculated using the MO5-2X functional and the Gaussian03 suite of programs [29]. Their relative energies and binding energies were reduced a little when corrections were included for zero point energies (using the harmonic frequencies computed at the MO5-2X/6-31+G** level) and for BSSE but the rank order remained the same. The predicted harmonic wavenumbers of their CH and OH stretch modes were scaled by the recommended factors [30] of 0.94 (MO5-2X) or 0.96 (B3LYP) to correct for anharmonicity.

3. Results and discussion

3.1. Phenyl α -L-fucopyranoside

In order to establish the experimental feasibility of measuring the IRID spectrum of a carbohydrate in the CH region and its correspondence with the spectrum predicted by DFT calculations, an initial trial investigation of bare phenyl α -L-fucopyranoside (PhFuc) was undertaken; (the phenyl tag was needed to provide the UV chromophore that enables detection through resonant two-photon

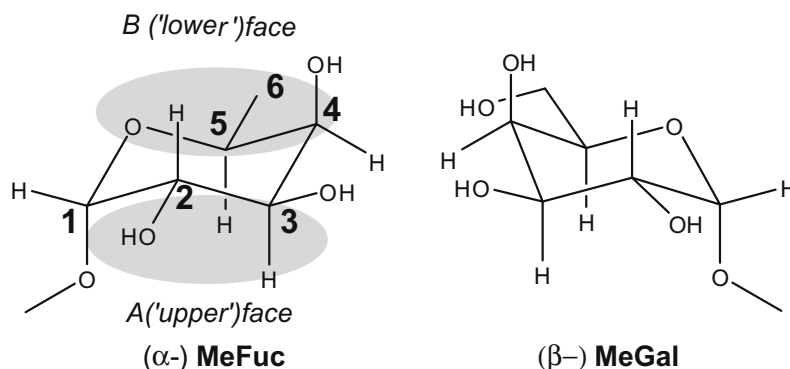


Fig. 1. Schematic structures of methyl α -L-fucopyranoside (MeFuc) and methyl α -D-galactopyranoside (MeGal).

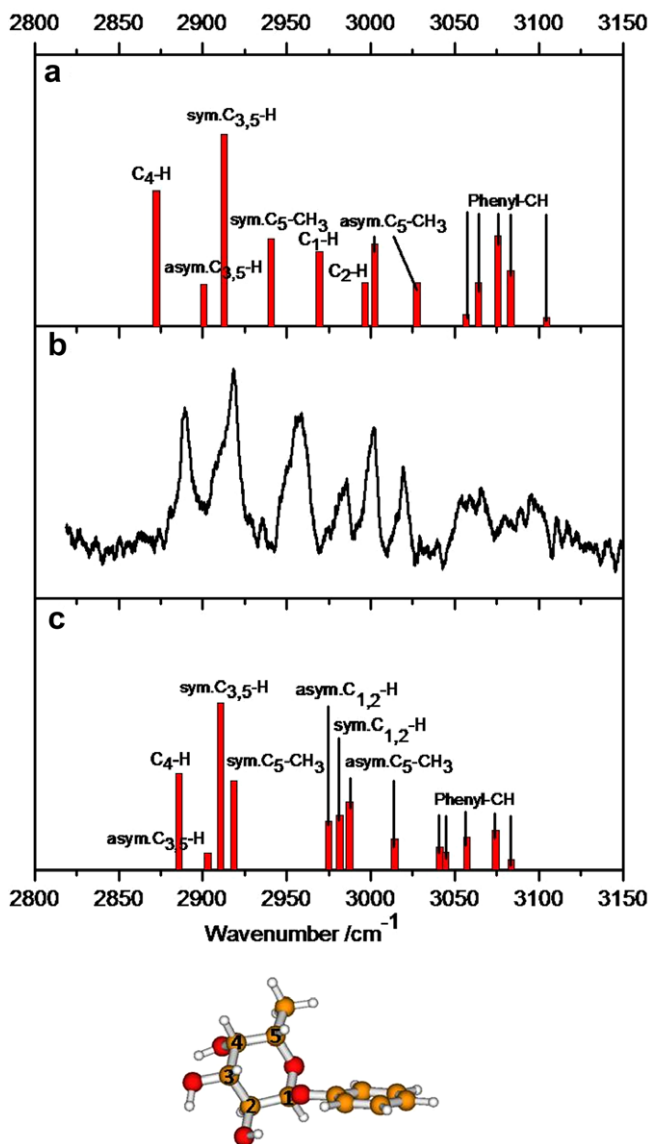


Fig. 2. Experimental and computed IRID CH spectra of phenyl α -L-fucopyranoside: (a) calculated using the B3LYP functional and a 6-31+G⁺⁺ basis set, (b) experimental, (c) calculated using the MO5-2X functional and a 6-31+G⁺⁺ basis set. Its computed minimum energy structure is shown at the bottom of the figure.

ionization, an essential component of the mass-selected IRID experiment). Its success can be gauged from the spectra shown in Fig. 2, where the experimental CH spectrum is compared with the predictions of DFT calculations using the B3LYP and MO5-2X functionals. Each one is in qualitative correspondence with experiment though the B3LYP functional appears to provide the closer agreement, principally because of the differences in the predicted positions of the CH₃ symmetric stretch mode, labeled as C₅-CH₃ and the CH mode labeled C₁-H. Since dispersion is not expected to be a major factor influencing the structure of the isolated molecule, this result favours the use of the B3LYP functional (for the isolated molecule). On the other hand this is unlikely to remain valid when the carbohydrate, methyl α -L-fucopyranoside is bound to an aromatic molecule/UV chromophore² through CH- π interactions and in this case the MO5-2X functional will be the preferred option.

² The IRID scheme, which requires a UV chromophore, cannot be used to provide an experimental vibrational spectrum of the uncomplexed carbohydrate; in the complex of course, the necessary chromophore is provided by the aromatic partner.

Table 1

Relative energies of the eight lowest energy structures of MeFuc · Tol-d₈ in kJ mol⁻¹, calculated using the MO5-2X functional and a 6-31+G⁺⁺ basis set, including those incorporating harmonic zero point energy corrections and counterpoise corrections for BSSE. The ' π , b, ...' notation identifies different orientations of the aromatic ring, which differ only by rotation in the aromatic plane.

Structure	ΔE	$\Delta E + \text{ZPE}$	$\Delta E + \text{ZPE} + \text{BSSE}$
cc-CH ₁ - π a	0.00	0.00	0.00
cc-CH ₁ - π b	2.70	1.49	1.17
cc-CH ₁ - π c	2.96	1.73	1.51
cc-CH _{3,4} - π	3.96	2.41	1.85
cc-CH ₁ - π d	3.07	3.62	3.27
cc-CH ₂ - π	5.42	4.63	4.00
cc-CH ₁ - π e	7.66	5.98	5.30
cc-OH ₂ - π	9.22	7.61	7.60

Table 2

Binding energies, D_e , of the eight lowest energy structures of MeFuc · Tol-d₈ in kJ mol⁻¹, calculated using the MO5-2X functional and a 6-31+G⁺⁺ basis set, together with those including ZPE (D_0) and BSSE corrections.

Structure	D_e	D_0	$D_0 + \text{BSSE}$
cc-CH ₁ - π a	-25.44	-22.06	-18.27
cc-CH ₁ - π b	-22.74	-20.57	-17.10
cc-CH ₁ - π c	-22.48	-20.33	-16.76
cc-CH _{3,4} - π	-21.48	-19.65	-16.42
cc-CH ₁ - π d	-22.37	-18.45	-14.99
cc-CH ₂ - π	-20.02	-17.43	-14.26
cc-CH ₁ - π e	-17.78	-16.09	-12.96
cc-OH ₂ - π	-16.23	-14.46	-10.67

3.2. Methyl α -L-fucopyranoside · toluene

The lowest lying structures (<10 kJ mol⁻¹ above the minimum) of MeFuc · Tol-d₈ calculated using the MO5-2X functional are listed, together with their relative energies and binding energies, in Tables 1 and 2. They are described by the term 'cc', which identifies a counterclockwise orientation of the hydrogen-bonded peripheral OH groups on the carbohydrate, OH₄ → OH₃ → OH₂ → O₁, and 'CH₁- π ' or 'OH₁- π ' which identifies the character of the intermolecular interaction with the aromatic partner. Seven of the eight calculated structures are bound through CH- π interactions, leading to binding energies in the range \sim (13–18) kJ mol⁻¹. The only one involving an OH- π hydrogen bond has the highest relative energy, 7.6 kJ mol⁻¹ and the rest involve binding either at the CH₁ site, which is located on the lower face of the pyranoside ring (see Fig. 1) or the CH_{3,4} site which binds the fucose to an apolar patch on the upper face of the ring. Both types of interaction have been identified in X-ray crystallographic investigations of complexes between proteins and methyl α -L-fucopyranoside (MeFuc) or its analogue methyl α -D-galactopyranoside (MeGal) – for example, the CH_{3,4,5}- π bonding of MeFuc to a tyrosine residue in Lectin 1 (from *Ulex europaeus*) [31] and CH₁- π bonding of MeGal, again at a tyrosine site in the galactose specific lectin from *Autocarpus hirsute* [32].

Recording the experimental IR spectra of MeFuc · Tol-d₈ complexes proved experimentally demanding since the complexes did not form readily and required very high ratios of toluene:MeFuc partial pressures (presumably because the complexes are very weakly bound). In consequence the signal:noise ratios in the recorded CH (and OH) spectra were very much smaller than those obtained with PhFuc. Fig. 3 compares the computed OH vibrational spectra of the uncomplexed carbohydrate (Fig. 3a), calculated using the MO5-2X functional rather than B3LYP for the sake of consistency in making comparisons, with the experimental

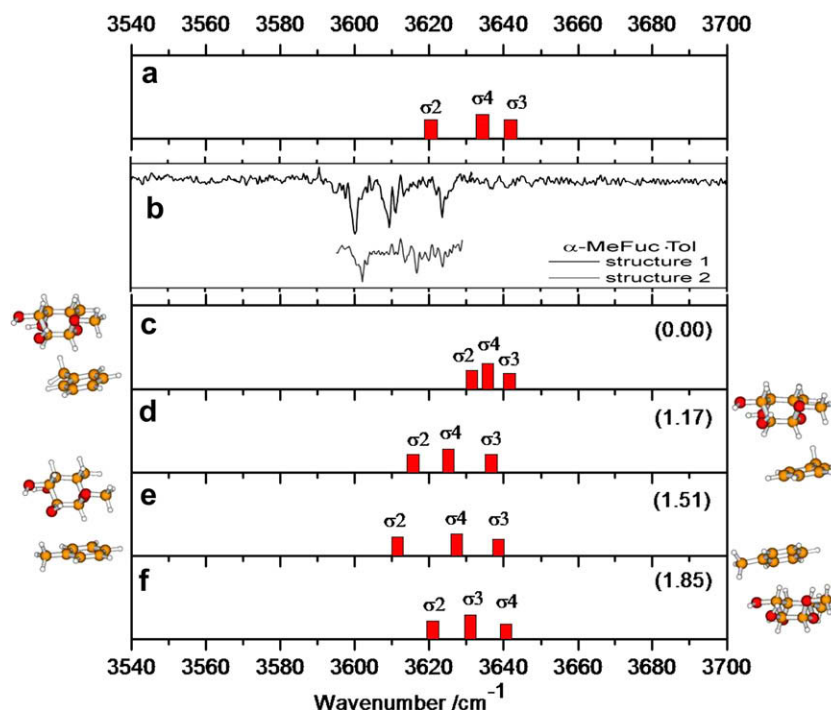


Fig. 3. (a) Computed OH vibrational spectrum of methyl α -L-fucopyranoside; (b) IRID OH spectra (shown as 'dip' spectra) of the two populated conformers of MeFuc · Tol-d₈ and (c)–(f) the corresponding vibrational spectra and conformations of its lowest lying computed structures, calculated using the MO5-2X functional and a 6-31+G** basis set.

IRID OH spectrum of the complex, which populates two alternative structures, or structural families, (Fig. 3b) and the OH vibrational spectra associated with the lowest energy structures of MeFuc · Tol-d₈, CH₁- π (Fig. 3c–e) and CH_{3,4}- π (Fig. 3f), again calculated using the MO5-2X functional. They are all very similar although the CH- π interactions do lead to very minor changes in the spectral patterns. Apart perhaps, from the minimum energy structure, each one of them is in good correspondence with each of the experimental spectra (though this is *not* true for the OH spectra associated with any of the calculated higher energy structures). It is tempting to associate the two experimental spectra with CH₁- π and CH_{3,4}- π bonded structures.

Similar comparisons conducted in the CH spectral region are presented in Fig. 4, which presents the computed (MO5-2X) vibrational spectra of uncomplexed MeFuc and the four lowest energy MeFuc · Tol-d₈ structures, and the experimental IRID spectrum associated with the more strongly populated complex ('structure 1' in Fig. 3b). Disappointingly, all the spectra are qualitatively similar and this, when coupled with their complexity and the poor resolution and intensity of the experimental spectrum, prevents assignment of the complex to a particular structure or the vibrational assignment of its experimental spectrum. Comparisons between the computed spectra of each of the complex structures and the bare carbohydrate do however, reveal some interesting results. Each of the spectra shown in Fig. 4c–e is associated with a structure bound through a C₁H- π interaction; when they are compared with the spectrum of bare MeFuc, 'blue shifts' can be identified in the respective C₁-H bands \sim 35 cm⁻¹ (c), \sim 64 cm⁻¹ (d) and \sim 57 cm⁻¹ (e) but none of the other bands is significantly displaced. Similarly, the C₃-H and C₄-H bands in spectrum (f), which is associated with a C_{3,4}- π bound complex, are blue-shifted by \sim 30 cm⁻¹ and \sim 66 cm⁻¹, respectively, but not the other bands which are virtually unaffected. These large, selective displacements towards higher wavenumber are comparable in magnitude (and direction) with those calculated for similar CH- π bonded carbohydrate-tolu-

ene structures, using the dispersion corrected D-DFT method [25]. Remarkably, blue shifts (about one third as large) are also predicted by calculations using the B3LYP functional although it is not capable of dealing with dispersive interactions.

4. Concluding remarks

Comparisons between the IR spectra recorded in the OH region, of a carbohydrate-aromatic complex, methyl α -L-fucopyranoside-toluene isolated in the gas phase, and the results of DFT calculations using the MO5-2X functional, reinforce the earlier qualitative proposal [18,19] of a complex bound through CH- π interactions alone. These can create stacked structures with the aromatic molecule bonded to the upper or the lower face of the pyranoside ring, through CH_{3,4}- π (upper) or CH₁- π (lower) interactions. Although it has also been possible to record an IR spectrum of the complex in the CH stretch region, its density and quality do not allow more detailed assignments to be made but the theoretical prediction of a strong blue shift in the CH bands associated with the CH- π interactions parallels the predictions made recently, on the basis of D-DFT calculations for other carbohydrate-toluene complexes [25], and encourages a further experimental pursuit based upon a simpler benchmark system such as the xylose-toluene complex, methyl α -D-xylopyranoside-toluene-d₈. In xylose, which provides an analogue of glucose rather than galactose since all its OH groups adopt an equatorial orientation, the hydroxymethyl group is replaced by an H atom. In addition, we have already discovered that the two anomers of the glucopyranoside behave differently: the toluene complex of the α -anomer is supported by OH- π bonding but the other anomer appears to be bound solely through CH- π interactions [19]. If the methoxy group were fully deuterated, methyl-d₃ α / β -D-xylopyranoside-toluene-d₈, the number of CH bands would be almost halved, falling from 11 (in MeFuc · Tol-d₈) to 6 (in MeXyl · Tol-d₈) to allow a considerable spectral simplification.

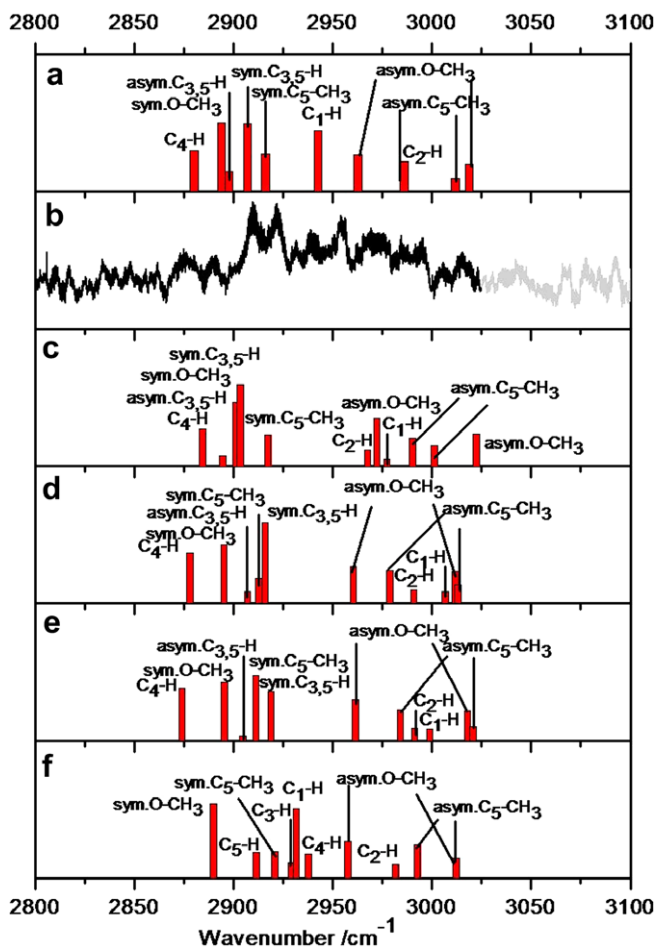


Fig. 4. (a) Computed CH vibrational spectrum of methyl α -L-fucopyranoside; (b) IRID CH spectra of conformer 1 of MeFuc · Tol- d_8 ; (c)–(f) corresponding vibrational spectra of its lowest lying computed structures, calculated using the MO5-2X functional and a 6-31+G** basis set. (The weak features lying above ~ 3025 cm^{-1} lie outside the range of aliphatic C–H modes: if real, they could be due to aromatic C–H bands associated with a small population of undeuterated toluene clusters. The difference between the mass ratios of MeFuc · Tol- d_8^+ and (Tol) $_3^+$ is only 2 amu. Its aliphatic C–H₃ bands could also account for some of the ‘crowding’ at lower wavenumbers).

Acknowledgements

We have appreciated helpful discussions with Professors Tim Zwier and Ian Hillier and with Dr. Timothy Vaden. Early B3LYP cal-

culations were performed by Mr. James Screen. Financial support has been provided by EPSRC (BGD, JPS), the Leverhulme Trust (JPS, Grant F/08788G), and NSERC (ZS). Support has also been provided by the Oxford Supercomputing Centre, the STFC Laser Loan Pool and the Physical and Theoretical Chemistry Laboratory.

References

- [1] X. Robert, R. Haser, T.E. Gottschalk, F. Ratajczak, H. Driguez, B. Svensson, N. Aghajari, *Structure* 11 (2003) 973.
- [2] A.B. Boraston, D.N. Bolam, H.J. Gilbert, G.J. Davies, *Biochem. J.* 382 (2004) 769.
- [3] W.I. Weis, K. Drickamer, *Annu. Rev. Biochem.* 65 (1996) 441.
- [4] R.A. Dwek, *Chem. Rev.* 96 (1996) 683.
- [5] A. Varki, *Glycobiology* 3 (1993) 97.
- [6] H. Lis, N. Sharon, *Chem. Rev.* 98 (1998) 637.
- [7] J. Jiménez-Barbero, J.L. Asensio, F.J. Cañada, A. Poveda, *Curr. Opin. Struct. Biol.* 9 (1999) 549.
- [8] R. Palma, M.E. Himmel, J.W. Brady, *J. Phys. Chem. B* 104 (2000) 7228.
- [9] M.C. Fernández-Alonso, F.J. Cañada, J. Jiménez-Barbero, G. Cuevas, *J. Am. Chem. Soc.* 127 (2005) 7379.
- [10] H.-J. Gabius, H.-C. Siebert, S. André, J. Jiménez-Barbero, H. Rüdiger, *Chem. Bio. Chem.* 5 (2004) 740.
- [11] C.R. Bertozzi, L.L. Kiessling, *Science* 291 (2001) 2357.
- [12] J. Jiménez-Barbero, J.L. Asensio, G. Cuevas, A. Canales, M.C. Fernandez-Alonso, F.J. Cañada, *Biotransform.* 24 (2006) 10.
- [13] P. Hobza, Z. Havlas, *Chem. Rev.* 100 (2000) 4253.
- [14] K. Hermansson, *J. Phys. Chem. A* 106 (2002) 4695.
- [15] M. Nishio, *Cryst. Eng. Commun.* 6 (2004) 130.
- [16] W. Zierkiewicz, P. Jurečka, P. Hobza, *Chem. Phys. Chem.* 6 (2005) 609.
- [17] S. Tsuzuki, A. Fujii, *Phys. Chem. Chem. Phys.* 10 (2008) 2584.
- [18] J. Screen et al., *Angew. Chem. Int. Ed.* 46 (2007) 3644.
- [19] E.C. Stanca-Kaposta, D.P. Gamblin, J. Screen, B. Liu, L.C. Snoek, B.G. Davis, J.P. Simons, *Phys. Chem. Chem. Phys.* 9 (2007) 4444.
- [20] J.P. Simons et al., *Int. Rev. Phys. Chem.* 24 (2005) 489.
- [21] C. Ramos, P.R. Winter, J.A. Stearns, T.S. Zwier, *J. Phys. Chem. A* 107 (2003) 10280.
- [22] A.D. Buckingham, J.E. Del Bene, S.A.C. McDowell, *Chem. Phys. Lett.* 463 (2008) 1.
- [23] J. Čerňý, P. Hobza, *Phys. Chem. Chem. Phys.* 9 (2007) 5291.
- [24] S. Grimme, *J. Comput. Chem.* 25 (2004) 1463.
- [25] R.K. Raju, A. Ramraj, M.A. Vincent, I.H. Hillier, N.A. Burton, *Phys. Chem. Chem. Phys.* 10 (2008) 6500.
- [26] Y. Zhao, D.G. Truhlar, *J. Chem. Theory Comput.* 3 (2007) 289.
- [27] T.S. Zwier, *J. Phys. Chem. A* 105 (2001) 8827.
- [28] F. Mohamadi et al., *J. Comput. Chem.* 11 (1990) 440.
- [29] M.J. Frisch et al., *Gaussian, Inc.*, Wallingford CT, 2004.
- [30] J.P. Merrick, D. Moran, L. Radom, *J. Phys. Chem. A* 111 (2007) 11683.
- [31] G.F. Audette, D.J.H. Olson, A.R.S. Ross, J.W. Quail, L.T.J. Delbaere, *Can. J. Chem.* 80 (2002) 1010. pdb ref. 1jxn.
- [32] K.N. Rao, C.G. Suresh, U.V. Katre, S.M. Gaikwad, M.I. Khan, *Acta Crystallogr. Sect. D* 60 (2004) 1404. pdb ref. 1tp8.