IOP Publishing Physica Scripta
Phys. Scr. **78** (2008) 058124 (7pp) doi:10.1088/0031-8949/78/05/058124

Probing the glycosidic linkage: secondary structures in the gas phase

John P Simons[1], E Cristina Stanca-Kaposta[1], Emilio J Cocinero[1], B Liu[1], Benjamin G Davis[2], David P Gamblin[2] and Romano T Kroemer[3]

[1] Chemistry Department, Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford OX1 3QZ, UK
[2] Chemistry Department, Chemical Research Laboratory, 12 Mansfield Road, Oxford OX1 4TA, UK
[3] Sanofi-Aventis, CRVA, 13 quai Jules Guesde, BP14, 94403 Vitry-sur-Seine, France

E-mail: John.Simons@chem.ox.ac.uk

Received 14 July 2008
Accepted for publication 16 July 2008
Published 5 November 2008
Online at stacks.iop.org/PhysScr/78/058124

Abstract
The functional importance of carbohydrates in biological processes, particularly those involving specific molecular recognition, is immense. Characterizing the three-dimensional (3D) structures of carbohydrates and glycoproteins, and their interactions with other molecules, not least the ubiquitous solvent, water, is a key starting point for understanding these processes. The combination of laser-based electronic and vibrational spectroscopy of mass-selected carbohydrate molecules and their hydrated complexes, conducted under molecular beam conditions, with *ab initio* computation is providing a uniquely powerful means of characterizing 3D carbohydrate conformations; the structures of their hydrated complexes, the hydrogen-bonded networks they support (or which support them); and the factors that determine their conformational and structural preferences.

PACS numbers: 87.14.Df, 87.80.Dj

(Some figures in this article are in colour only in the electronic version.)

1. Introduction

The display by Nature of certain recurring carbohydrate structural motifs, conserved across many biological systems, is central to their biological functions. Glycoproteins—proteins containing sugar (oligosaccharide) chains covalently attached to N or O atoms on their peptide side chains during or following their translation from RNA in the ribozyme—are important components of cell membranes, for example. The carbohydrate motifs (glycans) displayed on cell surfaces are key communication markers and their correct recognition is vital to processes ranging from fertilization to immunity, protein stability and the control of protein folding [1–5]. *N*-linked glycans, which are invariably constructed around a conserved core pentasaccharide unit, are often critical to the future prospects of the glycoprotein, see figure 1. The motifs, the so-called 'glycocode', found on maturing glycoproteins are indicative of the protein's age, structure and localization [6]. This 'information' can be interpreted by a variety of lectins, and by the transferase and glycosidase enzymes that assist in the construction and trafficking of both native and defective glycoproteins within the cell.

Despite their familiarity, the underlying reasons for the evolutionary selection of recurrent (and biosynthetically expensive) carbohydrate motifs by Nature are not yet fully understood. Why, for example, is *N*-acetyl glucosamine (GlcNAc) employed at the junction with *N*-linked proteins and not glucose (Glc)? Why does *N*-linked glycosylation occur exclusively at asparagine residues and why are they invariably incorporated in the peptide sequon, Asn-Xxx-Ser/Thr; why is the neighbouring residue Xxx never proline? Certainly these constraints must be related to biological function, perhaps in controlling protein folding during co-translational glycosylation. Perhaps also, through the provision of a powerful and precise structural scaffolding unit by the conserved glycan core, which facilitates the recognition processes ('molecular discussions') taking place at the glycan termini and protects the peptide chain from enzymic attack. The specific glycosidic linkages may have

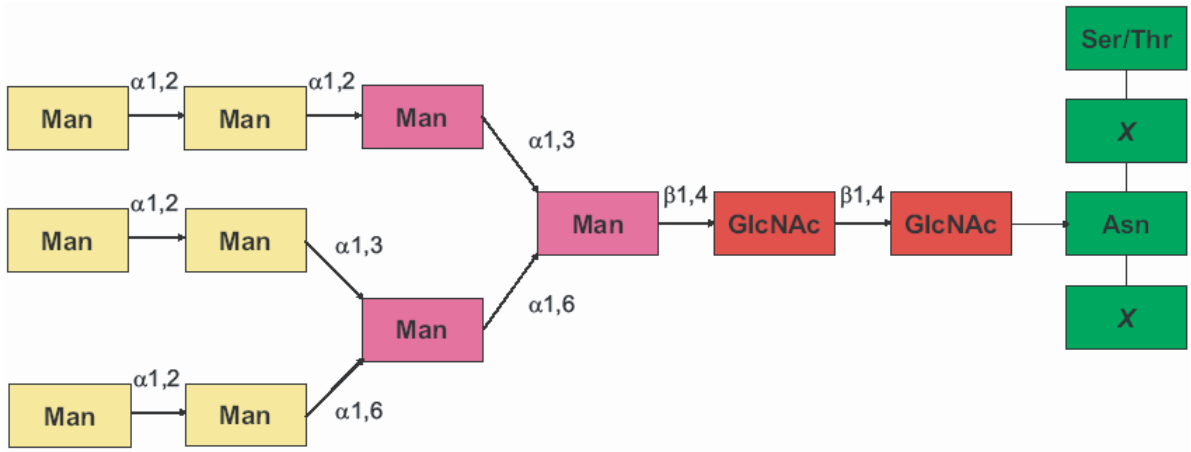

Figure 1. The recurrent core pentasaccharide motif, $(GlcNAc)_2(Man)_3$ (shown in the box), in a mannose-rich *N*-linked glycopeptide. Man and GlcNAc: mannose and *N*-acetyl glucosamine; Asn and Ser/Thr: asparagine and serine/threonine.

been selected because they can be 'stiffened' by direct or solvent-mediated hydrogen-bonded interactions across them, thereby influencing their shape, structural integrity and biological specificity.

Unfortunately, the assignment of oligosaccharide, let alone glycopeptide or glycoprotein structures and conformations, is notoriously demanding because of the flexibility of their glycosidic linkages and the chemical similarity of their component monosaccharide units, all of which contain multiple hydroxyl groups capable of extensive intra- and intermolecular hydrogen-bonded networks. The flexibility hinders crystallographic structural determinations and complicates or limits NMR (solution-phase) structural assignments, which are based upon (slow) time-averaged measurements. Molecular mechanics calculations of (gas-phase) oligosaccharide structures are handicapped by uncertainties in the parameters employed and *ab initio* calculations are restricted by molecular size and complexity constraints (though with the continual increases in computer power, these restrictions are steadily diminishing). Constrained molecular dynamics (MD) calculations of oligosaccharide conformational distributions in solution, used to interpret the results of NMR measurements, are further constrained by the additional complexity introduced by interactions with the solvent. Taken together, these limitations have represented a very significant challenge to experimental observations and to reliable molecular modelling of carbohydrate conformations, whether free or in solution or in a bound environment [6].

A better understanding of the connections between secondary structure and the specificity of molecular interactions is now being sought by meeting this challenge head-on through a combination of molecular synthesis to create a designed set of key oligosaccharide and glycopeptide 'building blocks'; mass-selected, infrared (IR) laser spectroscopy to probe their 'vibrational signatures' in the gas phase—at first free of the environment and then interacting with it in a controlled way within 'custom-made' hydrated and molecular complexes; and quantum chemical calculation, using density functional theory (DFT) and *ab initio* methods to translate these into their conformational and intermolecular structures [7].

Carbohydrates and glycopeptide building blocks are, of course, thermally fragile molecules, but they can be transferred intact into the vapour phase through laser desorption into an expanding rare gas jet, generated by a pulsed nozzle valve. Their collisional cooling during the gas-phase expansion usually traps them in the potential wells associated with their lowest energy conformers. Two cooling regimes can be identified: either the free-energy-driven high-temperature conformational distribution is frozen (fast cooling, i.e. large barriers in comparison with the mean collision energy), or more commonly, the conformational populations relax into the potential-energy-driven low-temperature distribution (slow cooling and low barriers). The high frequency of intermolecular collisions and the decreasing temperature also encourage the formation and stabilization of weakly bound molecular complexes which can be detected through mass-selected, resonantly enhanced two-photon ionization (MS-R2PI). Overlapping R2PI spectra, associated, for example, with different conformers or cluster structures, can be disentangled and identified by ultraviolet (UV) and IR laser 'hole-burning' experiments. Intense UV laser excitation, tuned to any selected band in the R2PI spectrum, depopulates its carrier and leads to a consequent dip in all the associated R2PI signals excited by a second, delayed laser pulse. Similar experiments conducted using tunable IR laser radiation with the delayed UV ionization laser tuned to a band associated with the selected carrier generate the IR ion dip (IRID) spectrum associated with the selected UV band. An example is shown in figure 2, which displays IRID spectra associated with the lowest energy conformational structure of lactose[4]. The conformational assignment was performed through, and also led by, comparisons between the experimental IR vibrational spectrum and those determined through a series of molecular mechanics, DFT and *ab initio* calculations on a wide range of conformers, following procedures that collectively 'homed in' either on a unique or a 'best fit' assignment [8]. Criteria included the overall qualitative match between the patterns of observed and calculated vibrational spectra; the correlation between the relative conformer populations

[4] 'Tagged' with a UV chromophore provided either by an O-linked benzyl group, to facilitate its detection through resonant two-photon ionization.

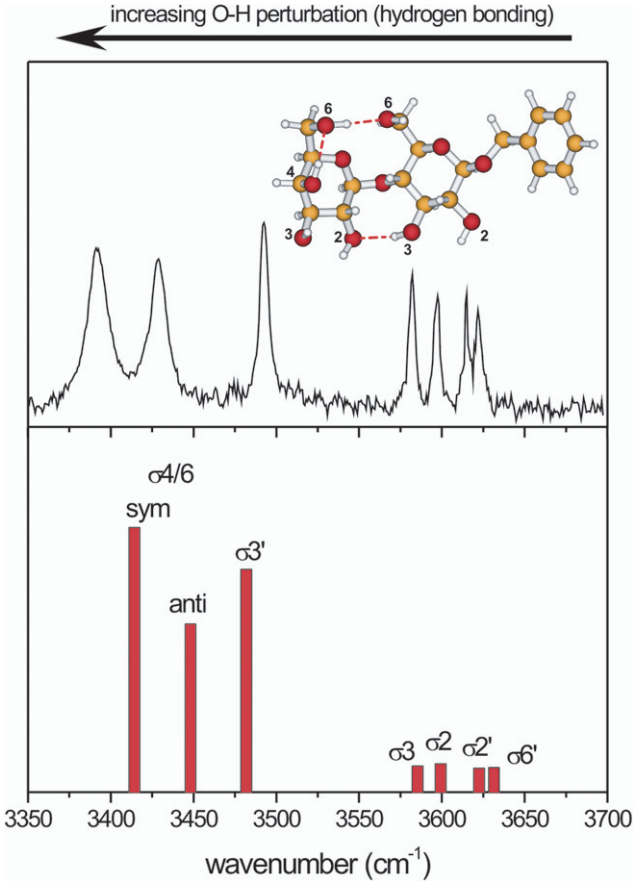

Figure 2. Experimental and computed IRID spectra of (O-benzyl) β-D lactoside. Note the strong displacement to lower wavenumber of the vibrational modes associated with inter-ring H-bonding across the $\beta(1,4)$ glycosidic linkage [8].

(estimated from the relative intensities of their band origins in the R2PI spectrum) and the ordering and magnitude of their calculated relative energies. The most populated conformers/structures generally correspond to those which the *ab initio* calculations have located as those of lowest zero point potential energy.

These strategies can provide a uniquely sensitive probe of the *intrinsic* hydrogen-bonded environments of individual carbohydrate molecules, which create characteristic IR 'spectral signatures' that can be fine-tuned by small (or large) changes in the molecular conformation. They may reflect anomeric and axial versus equatorial differences (or similarities). They also provide characteristic signatures of intermolecular binding motifs (and conformational changes), associated, for example, with 'insertion' or 'addition' structures in hydrated complexes [9]. The spectra of isolated monosaccharide 'building blocks' provide an 'alphabet' of vibrational signatures, which can be used to refine the assignment of larger oligosaccharide structures. When their signatures are not retained the spectral changes can be used to identify modifications in oligosaccharide structure, promoted for example by the retention (or disruption) of secondary structural motifs associated with inter-residue hydrogen-bonding.

The operation of these strategies, combining computational explorations of their conformational landscapes with experimental determinations of their conformer-specific (IR) vibrational spectra, has now revealed the intrinsic gas-phase conformations of a representative series of oligosaccharide building blocks that comprise the conserved core pentasaccharide, $(Man)_3(GlcNAc)_2$ in all *N*-linked glycopeptides (see figure 1). Spectra recorded in the OH (and NH) stretch region (3000–3800 cm^{-1}) and also in the lower wavenumber 'fingerprint' region (1200–2000 cm^{-1}) have led to structural assignments of chitobiose—the $\beta(1,4)$-linked *N*-acetyl glucosamine disaccharide, $(GlcNAc)_2$, which provides the link between the peptide chain and the branched trimannoside unit; the two dimannosides, Man$\alpha(1,3)$Man and Man$\alpha(1,6)$Man—which provide the terminal components of the branching pentasaccharide core; and the complete branched trimannoside, Man$\alpha(1,3)$Man$\alpha(1,6)$Man—which provides the key starting motif of each of the extended oligosaccharide arms that grow out of the core. Finally, the conformational consequences of explicit water binding on this key motif have been explored by probing the structure of its singly hydrated complex.

2. *N*-linked glycopeptides: conformational structures of the conserved core pentasaccharide building blocks

2.1. Chitobiose.

UV hole burning investigations of (*O*-phenyl) β-D GlcNAc$\beta(1,4)$GlcNAc isolated in a cold, free Ar jet expansion have identified two alternative, but very similar conformers. The IRID spectrum and global minimum energy structure of the principal conformer, recorded in the OH/NH stretch and also in the 'fingerprint' regions, are shown in figure 3, together with the computed vibrational spectrum associated with its global minimum conformation. The corresponding conformational structure and spectrum of the monosaccharide unit, (*O*-phenyl) β-D GlcNAc, are also shown for comparison.

The most strongly displaced band in the OH/NH stretch region, in both the mono- and the disaccharide, is associated with the OH3 group. It reflects a strong H-bonded interaction, OH3 $\rightarrow$ O=C, with the neighbouring *N*-acetylamino group. More significantly, in the disaccharide, the NH band associated with the interacting *N*-acetyl amino group is also displaced to lower wavenumber, reflecting its H-bonded interaction with the exocyclic hydroxymethyl group on the *neighbouring* ring, NH $\rightarrow$ OH6′; a second, reverse inter-ring H bond, OH3′ $\rightarrow$ OH6, is also signalled by the displacement of the band $\sigma 3'$ to lower wavenumber, $\sim$3530 cm^{-1}. These two inter-ring bonds combine to create a strongly rigidified scaffold that limits torsional flexibility about the glycosidic linkage. Although it might be anticipated that the H-bonded interactions would be disrupted when the disaccharide is transferred from the gas phase at low temperature into an aqueous solution at 298 K [10], combined NMR and constrained MD investigations have identified a very similar average conformation in solution [11] with the two acetyl amino groups continuing to adopt a *trans* disposition about the glycosidic linkage, sustained perhaps by the presence of explicitly bound water molecules, which take over the task of bridging across the two monosaccharide rings [12].

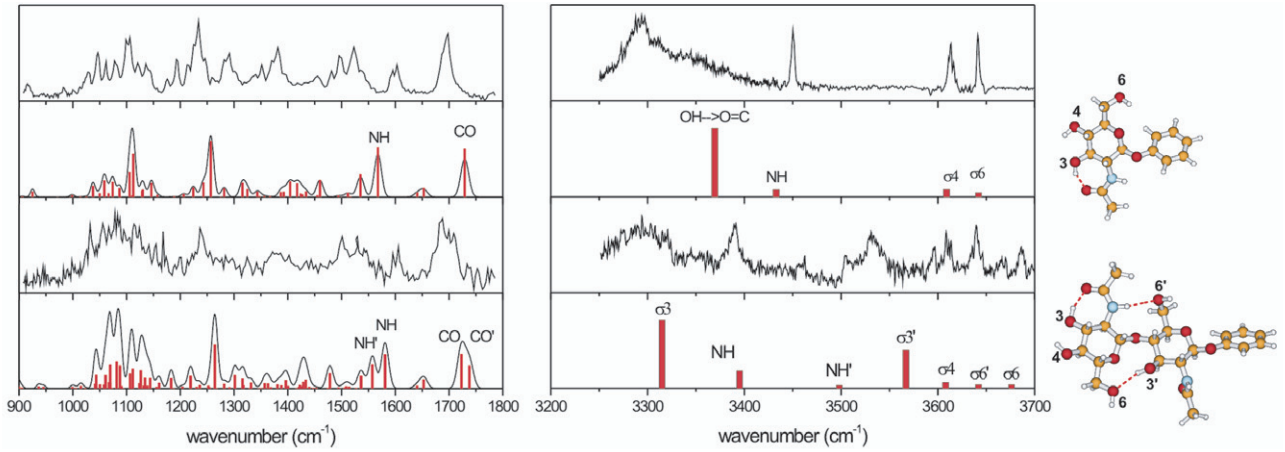

Figure 3. Experimental and computed IRID spectra of the monosaccharide, (*O*-phenyl) β-D GlcNAc (top) and its disaccharide GlcNAc$\beta(1,4)$GlcNAc (bottom). Conformational structures presented on the right-hand side correspond to global minimum energy conformations calculated via DFT (B3LYP/6-31 + G*) and *ab initio* (MP2/3-311 + +G**).

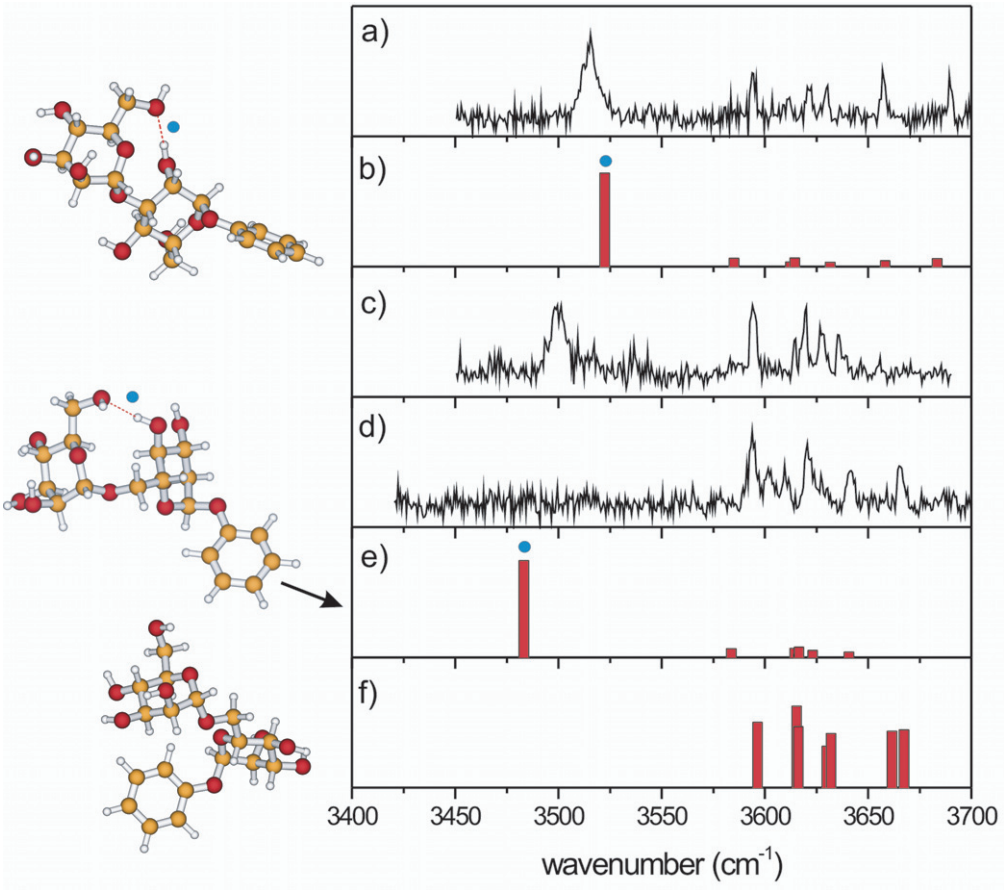

Figure 4. Experimental and computed IRID spectra of (a, b): (*O*-phenyl) α-D Man$\alpha(1,3)$Man and (c–f): (*O*-phenyl) α-D Man$\alpha(1,6)$Man. The blue dots identify the strongly H-bonded OH groups and their corresponding bands in the computed spectra.

2.2. The α (1, 3) and α (1, 6) dimannosides.

When the phenyl tagged unit, (*O*-phenyl) α-D Man$\alpha(1,3)$Man, was isolated in a cold, free jet expansion, the disaccharide relaxed into a single conformation and its IRID spectrum was in good agreement with the vibrational spectrum computed for its global minimum energy structure, see figure 4 [13]. Each of its seven OH bands was separately resolved; the most intense, strongly displaced to low wavenumber and peaking at $\sim$3515 cm^{-1}, was identified as $\sigma 2'$, associated with the inter-ring H bond, OH2′ $\rightarrow$ OH6. Since all the remaining bands were clustered in the range 3600–3700 cm^{-1}, there was no indication of any additional inter-ring H bonding (in contrast with the $\beta(1,4)$-linked disaccharides, lactose and chitobiose), but their relative disposition did provide a further, potentially very useful structural insight.

Comparisons with the corresponding 'vibrational signatures' of each of the (three) conformers of the monosaccharide unit, (*O*-phenyl) α-D mannose populated

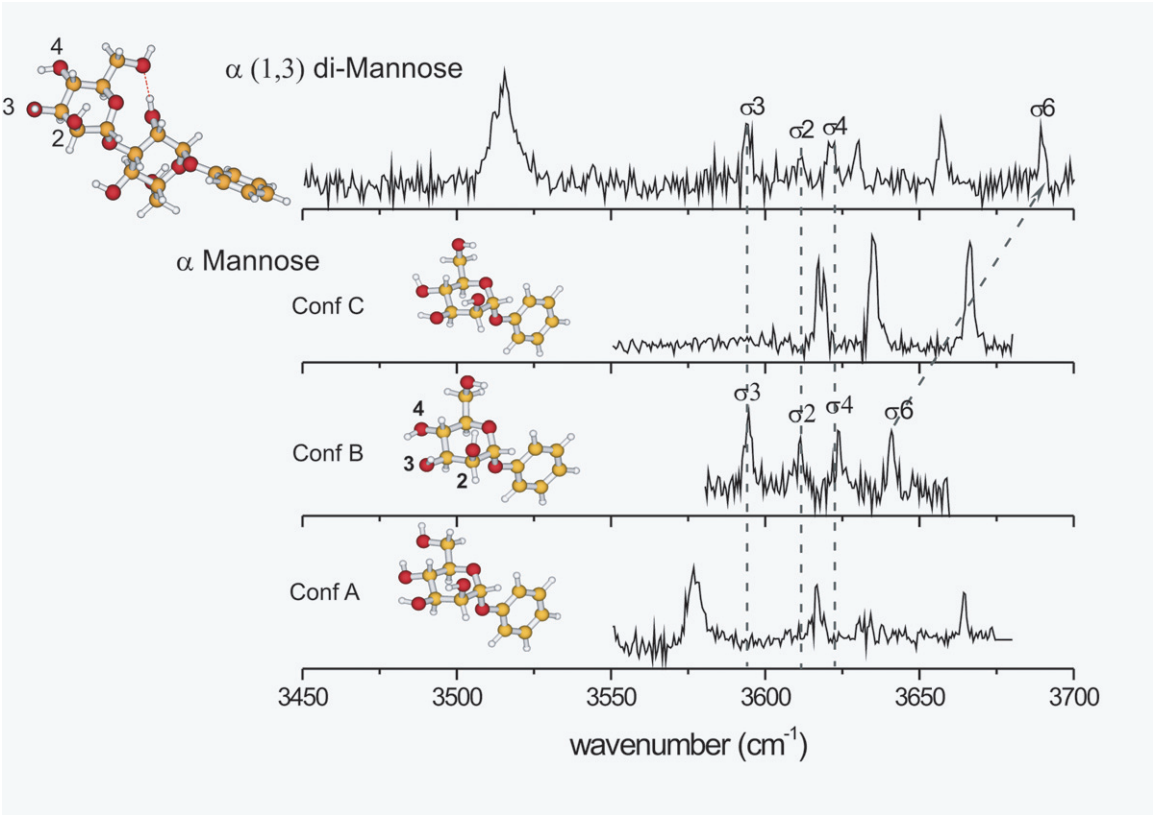

Figure 5. 'Vibrational signatures': comparison between the IRID spectra of the disaccharide (*O*-phenyl) α-D Man$\alpha(1,3)$Man and the three lowest energy conformers of its monosaccharide component (see text).

in a free jet environment, revealed a remarkably close correspondence between the set of bands identified as $\sigma 2$, $\sigma 3$ and $\sigma 4$ in the disaccharide and in one of the monosaccharide conformers; the structural orientation of their corresponding OH groups was virtually identical, see figure 5. The retention of a characteristic vibrational signature carried over from one of the constituent monosaccharide units illustrates the strategy that can be used to refine the assignment of larger oligosaccharide secondary structures, based upon an alphabet of established IR spectral signatures of different conformations of the monosaccharide units. On the other hand, the vibrational band $\sigma 6$, associated with the exocyclic OH6 group and located at $\sim$3690 cm^{-1} in the disaccharide, was located at much lower wavenumber, $\sim$3640 cm^{-1} in the monosaccharide unit. The change reflected its *altered* orientation. In the monosaccharide, it was able to interact with the ring oxygen atom, O5, acting as an H-bond donor, but in the disaccharide it acts as an acceptor and its orientation was reversed by the stronger inter-ring H bond, OH2′ $\rightarrow$ OH6, which moves $\sigma 6$ to higher wavenumbers by 'stiffening' the O–H6 bond (while lowering $\sigma 2'$).

Binding the two mannose units through the $\alpha(1,6)$ glycosidic linkage introduces an extra degree of torsional freedom and the possibility of some additional flexibility, so it was no surprise to discover the population of two distinct conformers of (*O*-phenyl) α-D Man$\alpha(1,6)$Man in the free jet expansion, and possibly some minor ones in addition [13]. One conformer presented an inter-ring H bond (from OH4′ $\rightarrow$ OH6), while the other (principal) conformer did not. Their IRID spectra are shown in figures 4(c) and (d), together with their computed vibrational spectra and associated conformational structures. Although the more strongly populated and more loosely bound conformer did not correspond to the global minimum energy (at 0 K), its looser structure and consequent higher entropy could well reverse the free energy ordering at the elevated temperatures associated with laser ablation prior to collisional relaxation in the jet. Retention of the original relative conformer populations would then reflect their separation by a significant energy barrier.

2.3. The trimannoside Man$\alpha(1,3)$Man$\alpha(1,6)$Man and its singly hydrated complex.

Despite the complexity of its conformational landscape and, *a fortiori*, the vast number of potential water binding sites and conformations, UV hole burning experiments indicate significant population of a single structure only, both for the 'tagged' trimannoside, (*O*-phenyl) α-D Man$\alpha(1,3)$Man$\alpha(1,6)$Man (TriMan), isolated in a free jet gas-phase expansion, and for its singly hydrated complex, TriMan · W1. Their IRID spectra recorded in the OH stretch region are shown in figure 6 together with their computed lowest energy structures and the corresponding vibrational spectra, calculated using DFT. The lowest energy conformer in TriMan is held together by two central, co-operatively linked hydrogen bonds, connecting OH4 on the central mannose ring, labelled **CM** in figure 6(a), to the hydroxymethyl group on the $\alpha(1,3)$ mannosyl arm, labelled **3M**, which is itself connected to the other hydroxymethyl group on the $\alpha(1,6)$ mannosyl arm, labelled **6M**. Comparison between the experimental and computed

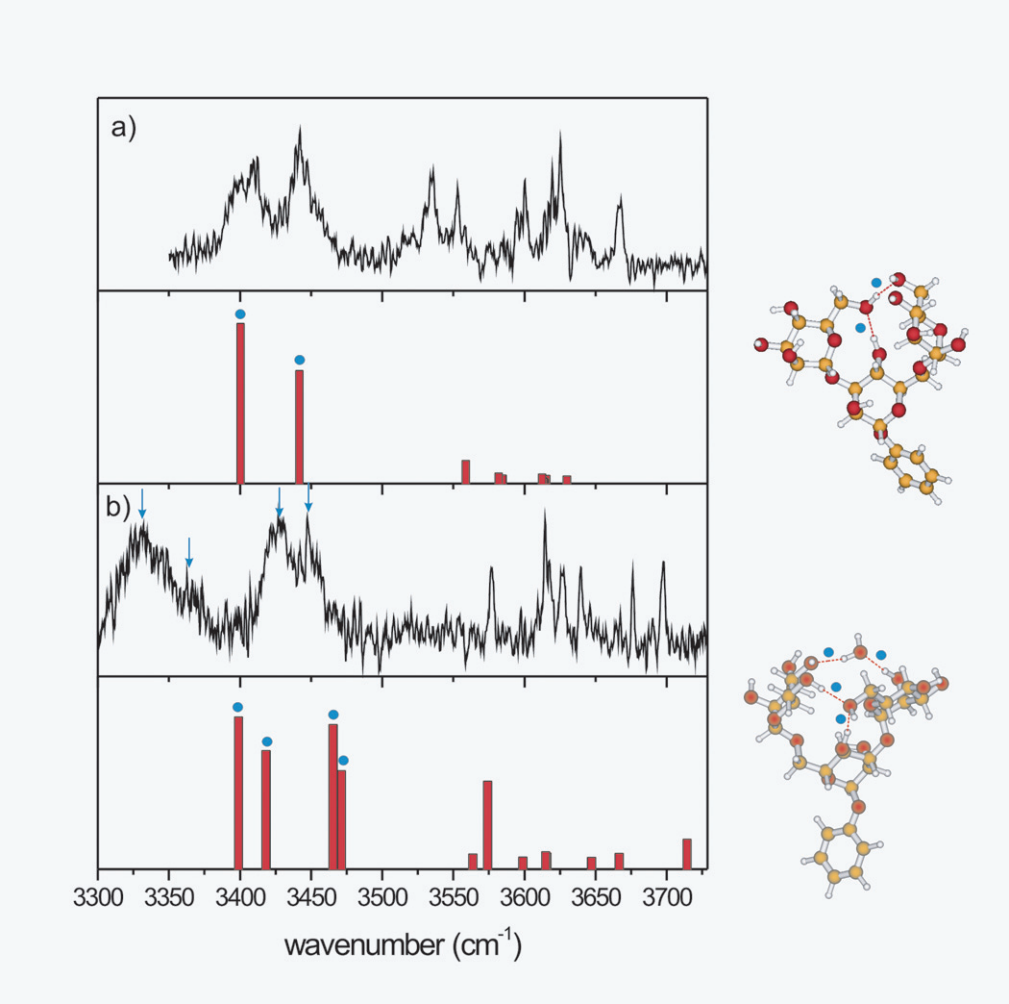

Figure 6. Experimental and computed IR spectra of (a) the trisaccharide (*O*-phenyl) α-D Manα(1, 3)Manα(1, 6)Man and (b) its singly hydrated complex. The blue dots indicate the strongly H-bonded OH groups in the structures and their corresponding vibrational bands in the computed spectra. The arrows identify these bands in the experimental spectra.

spectra identifies the pair of strongly displaced bands as the symmetric ($\sim$3400 cm^{-1}) and antisymmetric ($\sim$3440 cm^{-1}) stretching vibrational modes associated with the central chain, OH4(**CM**) $\rightarrow$ OH6(**3M**) $\rightarrow$ OH6(**6M**). In the absence of water and at low temperature, the trisaccharide adopts a tightly bound 'closed' conformation held together by strong intramolecular hydrogen bonds that link the two arms of the branched trimannoside unit. As with the $\alpha(1, 3)$ dimannoside, apart from the changes associated with hydrogen-bonded interactions of the exocyclic hydroxymethyl group, the $\alpha(1, 3)$ arm continues to retain the conformational structure and bonding displayed in the corresponding conformer of the free monosaccharide, cf figure 5.

The IRID spectrum of the hydrated complex, shown in figure 6(b), presents two *pairs* of strongly displaced bands associated with a structure in which the water molecule links the **3M** and **6M** rings. It has the lowest computed relative energy (at 0 K) and free energy (at 298 K) and its computed IR spectrum is in better accord with experiment than any of the alternative higher energy structures. This includes the cluster of bands recorded at higher wavenumbers, associated with the weakly H-bonded OH vibrational modes and the two pairs of bands at lower wavenumbers, associated with the more strongly H-bonded modes where a slight increase in the 'anharmonicity scaling' coefficient would bring the bands into very close accord with experiment.

3. *N*-linked glycan conformations in aqueous solution: comparisons and speculations

The 'closed' conformational structures of the bare trimannoside and its singly hydrated complex displayed at low temperature in the gas phase contrast with the 'open' structures determined in aqueous solution at 298 K, using NMR techniques coupled with MD calculations [14, 15]. One of the set of computed singly hydrated structures however, the centrally hydrated structure shown in figure 7(a), presented an (entropically favoured) open conformation in remarkably close correspondence to that determined in aqueous solution. The structure was generated by a simple 70° rotation of the **3M** wing in the bare molecule about the dihedral angle ψ[C1(**3M**)–O3(**3M**)–C3(**CM**)–C2(**CM**)] associated with the $\alpha(1, 3)$ glycosidic linkage, and insertion of the water molecule into its interior. An even more striking comparison could be made between the calculated structure of the open singly hydrated trimannoside unit (figure 7(a)) and the average computed root mean square deviation (RMSD) structure of the core pentasaccharide segment,

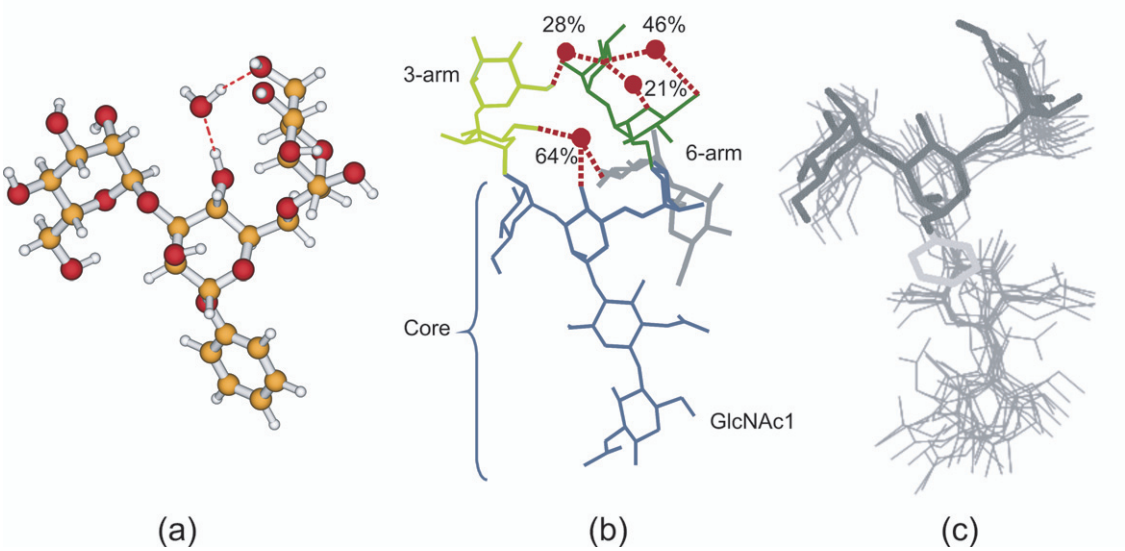

Figure 7. Comparison between (a) the computed open conformational structure of the monohydrate, TriMan·W1, (b) the computed RMSD conformation of the *N*-glycan, $(Man)_9(GlcNAc)_2$, based upon NMR measurements and MD simulations in aqueous solution at ~298 K [16], and (c) its core pentasaccharide segment, $(Man)_3(GlcNAc)_2$. Black rods indicate the trisaccharide framework, light grey rods, the phenyl tag and grey thin wires, the pentasaccharide framework.

$(Man)_3(GlcNAc)_2$, based upon NMR measurements and MD simulations of the 'water-bridged' high-mannose *N*-glycan $(Man)_9(GlcNAc)_2$ in an aqueous environment at ~298 K (figure 7(b)) [16]. This is highlighted in figure 7(c), which compares the 'frozen' singly hydrated trimannoside structure shown in figure 7(a) with the structure of the core pentasaccharide unit in $(Man)_9(GlcNAc)_2$, where the most persistent explicit hydration site incorporates a 'keystone' water molecule H-bonded to OH4(**CM**). The similarity between the open structure and the relative orientations of the **3M** and **6M** mannosyl 'wings' in the trimannoside and in the corresponding unit in the core pentasaccharide is remarkable. Perhaps the conformation of the key branched (trimannoside) segment is driven, in aqueous solution, by the creation of a water 'pocket' in the branch 'fork'; indeed it would not be difficult for the central water molecule linked to OH4(**CM**) to slip into the water pocket. The repeated ocurrence of this motif in many fully solvated glycan structures suggests a more general interplay between their favoured motif and hydration, with water pockets that conserve bisecting water molecules acting as beneficial structural elements.

Acknowledgments

We acknowledge support from the EPSRC, the Leverhulme Trust (grant no. F/08788G), the Spanish Ministry of Education and Science (EJC), the STFC Laser Loan Pool, the Oxford Supercomputing Centre and the Physical and Theoretical Chemistry Laboratory.

References

[1] Lowe J B 2001 *Cell* **104** 809
[2] Talbot P, Shur B D and Myles D G 2003 *Biol. Reprod.* **68** 1
[3] Rudd P M, Joao H C, Coghill E, Fiten P, Saunders M R, Opdenakker G and Dwek R A 1994 *Biochemistry* **33** 17
[4] Parodi A J 2000 *Annu. Rev. Biochem.* **69** 69
[5] Wormald M R, Petrescu A J, Pao Y L, Glithero A, Elliott T and Dwek R A 2002 *Chem. Rev.* **102** 371
[6] Helenius A and Abei M 2001 *Science* **291** 2364
[7] Simons J P, Çarçabal P, Davis B G, Gamblin D P, Hünig I, Jockusch R A, Kroemer R T, Marzluff E M and Snoek L C 2005 *Int. Rev. Phys. Chem.* **24** 489
[8] Jockusch R A *et al* 2004 *J. Am. Chem. Soc.* **126** 5709
[9] Hünig I, Painter A J, Jockusch R A, Çarçabal P, Marzluff E M, Snoek L C, Gamblin D P, Davis B G and Simons J P 2005 *Phys. Chem. Chem. Phys.* **7** 2474
[10] Kirschner K N and Woods R J 2001 *Proc. Natl. Acad. Sci. USA* **98** 10541
[11] Yu F and Prestegard J H 2006 *Biophys. J.* **91** 1952
[12] Almond A and Sheehan J K 2003 *Glycobiology* **13** 255
[13] Çarçabal P, Hünig I, Gamblin D P, Liu B, Jockusch R A, Kroemer R T, Snoek L C, Fairbanks A J, Davis B G and Simons J P 2006 *J. Am. Chem. Soc.* **128** 1976
[14] Sayers E W and Prestegard J H 2000 *Biophys. J.* **79** 3313
[15] Tian F, Al-Hashimi H M, Craighead J L and Prestegard J H 2001 *J. Am. Chem. Soc.* **123** 485
[16] Woods R J, Pathiaseril A, Wormald M R, Edge C J and Dwek R A 1998 *Eur. J. Biochem.* **258** 372