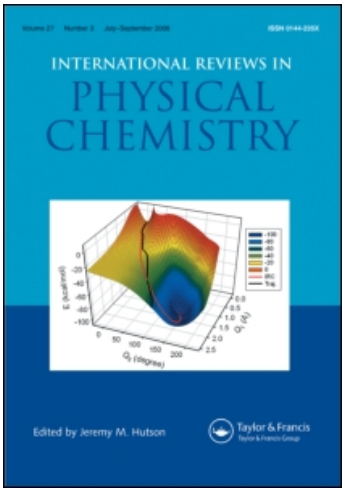


International Reviews in Physical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713724383

Sugars in the gas phase. Spectroscopy, conformation, hydration, co-operativity and selectivity

John P. Simons[a]; Rebecca A. Jockusch[a]; Pierre ÇarÇabal[a]; Isabel Hünig[a]; Romano T. Kroemer[b]; Neil A. Macleod[a]; Lavina C. Snoek[a]
[a] Chemistry Department, Physical and Theoretical Chemistry Laboratory, Oxford, OX1 3QZ, UK [b] Sanofi-Aventis, Centre de Recherche de Paris, BP14, 94403 Vitry-sur-Seine, France

To cite this Article Simons, John P. , Jockusch, Rebecca A. , ÇarÇabal, Pierre , Hünig, Isabel , Kroemer, Romano T. , Macleod, Neil A. and Snoek, Lavina C.(2005) 'Sugars in the gas phase. Spectroscopy, conformation, hydration, co-operativity and selectivity', International Reviews in Physical Chemistry, 24: 3, 489 — 531
To link to this Article: DOI: 10.1080/01442350500415107
URL: http://dx.doi.org/10.1080/01442350500415107

International Reviews in Physical Chemistry,
Vol. 24, Nos. 3–4, July–December 2005, 489–531

Sugars in the gas phase. Spectroscopy, conformation, hydration, co-operativity and selectivity

JOHN P. SIMONS*†, REBECCA A. JOCKUSCH†, PIERRE ÇARÇABAL†, ISABEL HÜNIG†, ROMANO T. KROEMER‡, NEIL A. MACLEOD† and LAVINA C. SNOEK†

†Chemistry Department, Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford, OX1 3QZ, UK
‡Sanofi-Aventis, Centre de Recherche de Paris, 13 quai Jules Guesde, BP14, 94403 Vitry-sur-Seine, France

(*Received 6 July 2005; in final form 1 September 2005*)

The functional importance of carbohydrates in biological processes, particularly those involving specific molecular recognition, is immense. Characterizing the three-dimensional structures of carbohydrates and glycoconjugates and their interactions with other molecules, particularly the ubiquitous solvent, water, are key starting points on the road towards the understanding of these processes. The review introduces a new strategy, combining electronic and vibrational spectroscopy of mass-selected carbohydrate molecules and their hydrated (and also protonated) complexes, conducted under molecular beam conditions, with *ab initio* computation. Its early successes have revealed a uniquely powerful means of characterizing carbohydrate conformations and hydrated structures, the hydrogen-bonded networks they support (or which support them) and the specificity of their interactions with other molecules. The new information, obtained in the gas phase, complements that provided by more 'traditional' condensed phase methods such as NMR, X-ray diffraction, molecular mechanics and molecular dynamics calculations. The review concludes with a vision of the challenges and opportunities offered by applications of molecular beam spectroscopy and their relevance in a biological context.

Contents

*Corresponding author. Email: john.simons@chem.ox.ac.uk

International Reviews in Physical Chemistry
ISSN 0144–235X print/ISSN 1366–591X online
http://www.tandf.co.uk/journals
DOI: 10.1080/01442350500415107

1. Preamble

There are many ways of exploring biomolecular structure and the forces which control it: in condensed phase environments, in situations approximating the 'natural' environment, or more fundamentally, in the gas phase isolated from the environment. The differing strategies, environments and degrees of molecular size and complexity tend to define individual scientific communities addressing related issues but evolving along complementary paths. The 'gas phase' communities include:

- **Optical spectroscopists** exploring the structures, conformations and non-bonded interactions in small, mainly neutral, biomolecules and their hydrated and molecular complexes: e.g. neurotransmitters; amino acids, oligopeptides and protein fragments; nucleobases and base pairs.
- **Mass spectrometrists** exploring equilibria, interactions, and the primary and higher order structures of larger polymeric biomolecular ions and their molecular complexes: e.g. protonated oligopeptides and nucleic acids; their hydrated, intermolecular and metal ion complexes; proteins, carbohydrates and glycoconjugates.
- **Computational chemists and molecular modellers**, employing *ab initio* or density functional theory, and molecular mechanics or molecular dynamics methods to

Figure 1. Bridging the physical and life sciences.

provide the menu of structural possibilities from which the experimentalists – and Nature – make their choice.

Over the last decade, driven (subconsciously perhaps) by Thomas Huxley's *dictum*, 'Vital forces are molecular forces', the optical spectroscopic community has joined the mass spectrometric and computational chemistry communities to help construct a new approach road towards the bridge that links the physical and life sciences, figure 1.

An exciting new milestone on that road, which also helps to connect the gas phase to the condensed phase, has recently been provided by the first optical spectroscopic investigation of a protein, cytochrome c, transferred into the gas phase through electrospray ionization (ESI), stored in a trapping mass spectrometer and detected in a series of individual, *charge-selected* states through their mid-infrared (IR) photodissociation spectra [1]. Uniquely, gas-phase spectra provide 'vibrational signatures' of the isolated protein structures free from environmental disturbance. The clear resolution of the amide I and amide II vibrational bands lying close to, but slightly shifted from, those of cytochrome c in aqueous solution (where charge selection is *not* possible), suggests the gas-phase structure is an intrinsic property of the protein which is maintained in solution. The growth of an additional band with increasing protein charge, accompanied by small displacements of the amide I and II bands, however, also points to the likelihood of structural changes dependent upon pH in a solution environment.

Conversion of a protein's IR spectral data into its conformational structure requires reliable calculation of its normal modes, a dream which lies beyond the scope of current computational methods, because of the large molecular size. This constraint does not apply to 'small molecule' ligands, however; their structures *can* be determined in the gas phase, isolated from any molecular interactions, and then compared with their structures at protein receptor binding sites, determined through X-ray diffraction or NMR spectroscopy.

2. Sweetness and light: Sugars in the gas phase

An important step along this alternative route, which also provides the *leitmotif* for the current review, was provided by the first gas-phase infrared structural investigation of a carbohydrate disaccharide unit. When lactose (galactosyl β(1–4) glucoside), tagged with a benzyl chromophore (Galβ(1–4)Glc(1–)OBn) was evaporated from an oven, entrained in an expanding argon jet and detected through laser excitation of its ultraviolet (UV) and IR spectra in the low-temperature ($\sim$5 K) gas-phase environment, only one structure was significantly populated [2]. The strong displacement to lower wavenumbers of three of its seven (individually resolved) mid-infrared O–H vibrational bands reflected strong hydrogen bonding across the glycosidic linkage (see figure 2). The inter-ring H-bonding supports a rigid *cis*-conformational structure

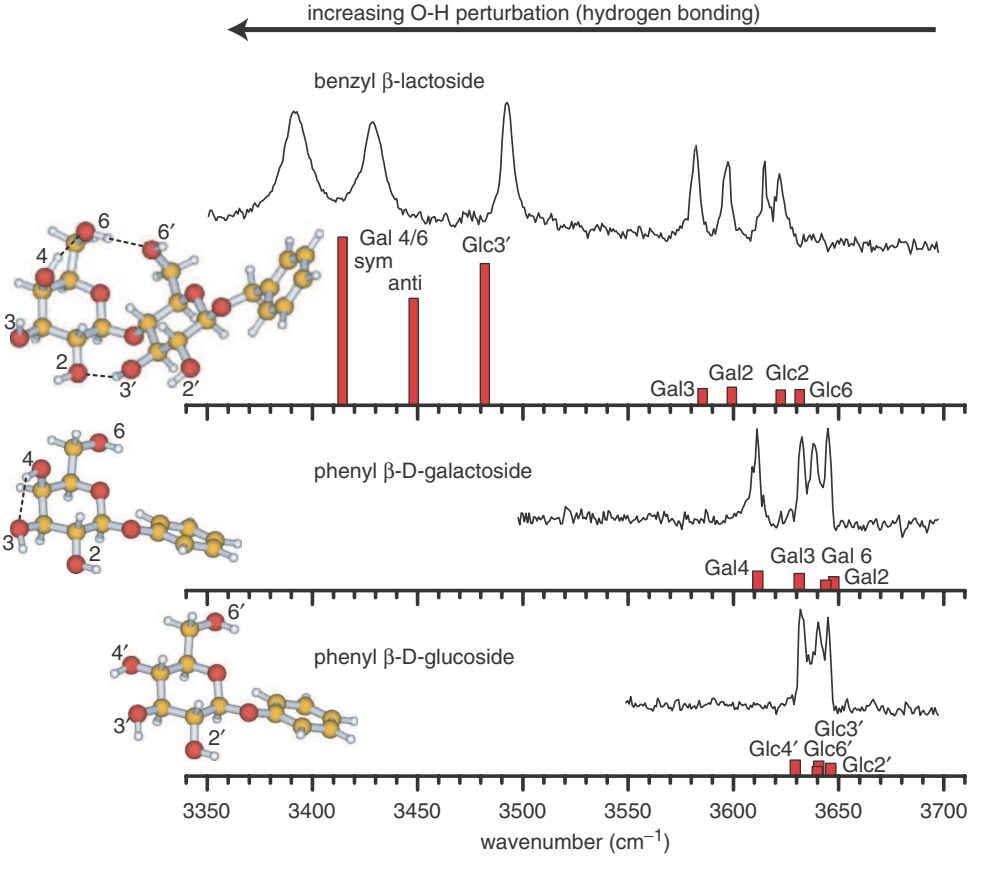


Figure 2. Experimental and computed near-infrared ion-dip spectra of the disaccharide benzyl-β-D-lactoside, and of its component monosaccharides, (phenyl)-β-D-galactoside and -β-D-glucoside, recorded in the 'O–H vibrational signature' spectral region [2]. Note the sensitivity of the individual hydroxyl stretch modes to their differing hydrogen-bonded environments. The benzyl or phenyl 'tags' do not significantly disturb the sugars to which they are attached; they are required to provide the UV chromophore utilized in the 'ion-dip' detection scheme (see section 3).

(identified with the aid of supporting *ab initio* calculations) favoured at low temperature by their enthalpic contribution to the free energy. At elevated temperatures (~300 K), however, a combined NMR and molecular dynamics investigation of lactose conducted in solution [3], indicates a preference for the population of a much freer, *trans*-conformational structure which does not facilitate such strong inter-ring hydrogen bonding, and is favoured therefore by an enhanced entropic factor.

These studies reveal striking parallels, and differences, between gas-phase, condensed-phase and 'protein-bound' carbohydrate structures. The conformational flexibility of carbohydrates, although thought to be an important factor in their molecular recognition by proteins (lectins) or antibodies [4] cannot be the only relevant factor. Examination of the (few) available protein-bound structures where lactose (Galβ(1–4)Glc) or its close neighbour, cellobiose (Glcβ(1–4)Glc), is incorporated as a *tightly* bound ligand demonstrates a preference for a rigidly hydrogen-bonded structure similar to the one populated in the low-temperature gas-phase environment rather than the flexible structure favoured in the free ligand in solution: its tight binding provides an enthalpic contribution which compensates for the entropic cost.

The β(1–4) glycosidic linkage incorporated in lactose and cellobiose is in fact a very common *motif* in many natural polysaccharides and importantly in many of the *glycan* (oligosaccharide) side-chains incorporated into glycoconjugate structures (glycoproteins and glycolipids). This is a prevalent protein modification; about 90% of all eukaryotic protein species are thought to be *N*-glycosylated [5], and *N*-linked glycans are involved in a multitude of recognition processes within and between cells. All incorporate glycan side-chains containing the same core structure, $(Man)_3(GlcNAc)_2$ – a β(1–4) linked *N*-acetyl glucosamine disaccharide unit $(GlcNAc)_2$ attached to the peptide through a β(1–) linkage to an asparagine (Asn) residue, and to the oligosaccharide chain through a second β(1–4) linkage to a branched α(1–3), α(1–6) mannose trisaccharide unit, see figure 3 [6]. The predominant glycan monosaccharide units, e.g. GlcNAc, Man, Glc, Gal,

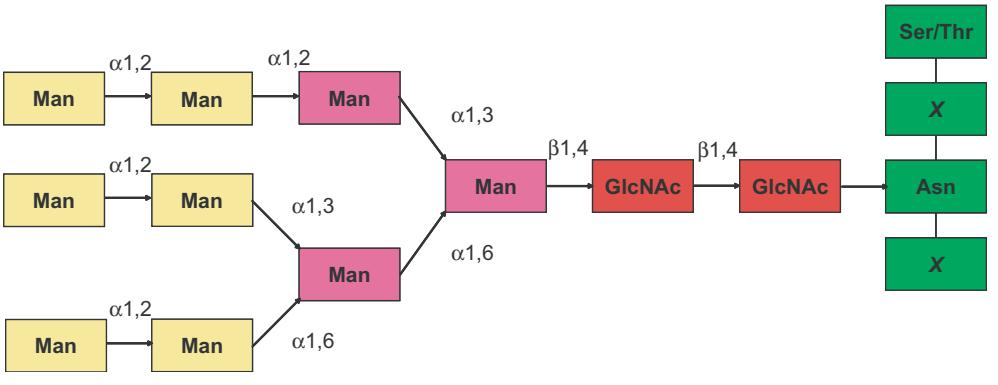


Figure 3. The common glycan core in *N*-linked glycopeptides includes a branched mannose trisaccharide unit, shown in violet, and a β(1–4) linked *N*-acetyl glucosamine (GlcNAc) disaccharide unit (chitobiose), shown in red. This unit connects the glycan to the peptide through an asparagine residue followed, almost always by a serine or threonine, two residues away. Further carbohydrate units, shown in yellow, comprising mannose (or other sugars) grow out of the common core and their termini serve as 'multivalent recognition antennae'.

fucose (Fuc), are all hexoses, though xylose (Xyl), a pentose sugar, is also a common and important molecular recognition unit.

The natural occurrence of common, conserved carbohydrate structural *motifs* must be related to their biological function and the overall glycan structure, potentially influencing that of the linked protein and 'animating' the 'molecular discussions' taking place at the glycan termini. The underlying reasons for their evolutionary selection remain to be established but elegant studies using primarily NMR and computer modelling techniques [7–9] do suggest their fundamental role as precise scaffolding units, providing support for the multivalent spatial presentation of key 'recognition antennae' growing from the conserved core, see figure 3. The presence of multiple hydroxyl groups in the constituent monosaccharide rings allows the growth of glycosidic connections creating both linear and branched oligosaccharide and glycan structures. They can also support extensive intra- and intermolecular linked hydrogen-bonded networks. The glycosidic linkages may be highly flexible or, as we have seen, they may be stiffened by direct or environmentally mediated hydrogen-bonded interactions across them. Collectively, these factors can operate in a complex way to influence the structure, shape and biological specificity of individual carbohydrates and glycoconjugates [4, 6, 7, 10]. Unfortunately, the assignment of carbohydrate structures and conformations in the condensed phase (or the gas phase) is notoriously demanding [7].

The application of crystallographic methods to determine oligosaccharide and glycan structures is often compromised by the flexibility of (some of) the glycosidic linkages, making the growth of regular crystals a challenging task [7]. Also, structures that have been determined by X-ray diffraction measurements may be influenced by their solid state environment though the close packing in crystals, and the presence of impurities (mainly water) might somehow mimic the molecular crowding of living organisms. NMR measurements of glycosidic conformations in solution, using nuclear Overhauser (NOESY) [7] or, more recently, residual dipolar coupling methods [9], combined with molecular dynamics calculations to convert the data into carbohydrate conformational structures, are based necessarily upon measurements averaged over millisecond time scales. Interpretation of the data can be complicated by the population of multiple conformers [7] and by their inter-conversion at reciprocal rates comparable with or much faster than the time scale of the NMR experiment. NOESY based assignments can also be biased by the $1/\langle r^6 \rangle$ dependence of ^{1}H or ^{13}C spin coupling constants, which emphasizes conformations which allow the closest approach of the coupled nuclei. Vibrational Raman optical activity (ROA) spectroscopy offers a promising alternative strategy [11, 12] but, as with all spectroscopic methods, quantitative structural interpretation of ROA spectra requires the availability of reliable computational support. Accurate modelling of oligosaccharide vibrational force fields, Raman optical activities and equilibrium structures in the gas phase, let alone the condensed phase, soon becomes computationally exorbitant. In the gas phase, where comparisons with experiment are possible, *ab initio* calculations of the higher frequency bond-stretching motions agree well with observation but for the coupled low-frequency torsion and bending motions involved in conformational change the accord is less satisfactory [2, 13] and unlikely to allow unequivocal structural assignments.

The scope of *ab initio* or density functional theoretical (DFT) calculations [8, 14] is limited by the constraints of computational cost imposed by molecular size, complexity

and scaling laws, though the limits are continually being extended by the sustained increase in computer power. A promising new development in DFT theory [15] which enables calculations to be conducted on large biomolecules (comprising 100 atoms or more) through the introduction of linear (as opposed to cubic) scaling codes running on parallel computers, also offers hope for the future. The accuracy of structural modelling through molecular mechanics or molecular dynamics programs necessarily depends upon the quality of the empirical force field parameters they employ and, in the absence of reliable experimental comparisons, are open to critical review [8, 16]. Calculations of oligosaccharide conformations in the condensed phase, e.g. in aqueous solution and other interactive environments, face additional complexities, not least the accurate description of explicit hydrogen-bonded interactions [8, 17, 18] versus the influence of the bulk properties of the medium. Taken together, these limitations present a very significant challenge to the experimental assignment and reliable modelling of carbohydrate structures and conformations.

So how should a physical chemist rise to the challenge? Well, physical chemists being physical chemists, they naturally favour a quantitative and reductionist approach – though wisely, one that is also guided by their interactions with their glyco-scientific friends. The authors of this review (and their glyco-scientific friends), encouraged by the unexpected simplicity of the combined computational and IR spectroscopic investigations of lactose and its monosaccharide components, glucose and galactose in the low-temperature gas-phase environment of a free jet expansion (see figure 2), have been applying the same strategy to allow the structural determination of:

(a) a representative set of the basic monosaccharide units and key di-, tri- and oligosaccharide 'glycan building blocks', initially as isolated molecules, followed by
(b) similar determinations for a set of custom-synthesized key 'glyocopeptide building blocks' including both peptide and glycopeptide fragments, leading to
(c) an exploration of the nature and consequences of internal non-bonded and H-bonded interactions and, equally importantly, external interactions with other molecules, particularly water and their relationship to specific molecular recognition processes.

Through these studies, an acute IR spectral sensitivity to the carbohydrates' conformational and anomeric structures has been revealed and the fine structural control exerted by extended co-operative hydrogen bonded networks is emerging as a recurrent theme. This review provides an unashamedly focused overview of current progress and a taste of other bio-related applications of carbohydrate molecular beam spectroscopy; it concludes with a vision of its future prospects and relevance in a biological context.

3. Experimental and computational strategies

Laser excitation of flexible carbohydrate molecules entrained in a pulsed supersonic jet expansion at low temperatures provides a very powerful strategy for the isolation and detection of their molecular conformers and also their intermolecular

complexes [19–23]. In combination with computation it enables conformational and structural assignments to be made both for the isolated carbohydrate molecules and their molecular complexes, typically hydrates with a defined number of water molecules, in contrast to NMR studies conducted in solution or X-ray diffraction studies conducted in the crystalline state.

Carbohydrates and glycoconjugates are thermally fragile molecules but they can be transferred into the vapour phase intact, using a carefully controlled oven or more generally through laser desorption,[1] into the expanding rare gas jet generated by a pulsed nozzle valve. The gas phase expansion promotes collisional cooling down into potential wells associated with individual conformers where, if the barriers are high enough, they will be trapped. During the cooling process, two regimes can be identified: in one, it is the free energy-driven high-temperature conformational distribution that is frozen (fast cooling, i.e. large barriers in comparison with the mean collision energy); in the other, 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.

The intermingled ultraviolet R2PI spectra of all the species entrained by the supersonic gas pulse, recorded downstream through mass-selected resonantly enhanced two-photon ionization (MS-R2PI) spectroscopy,[2] may be associated with several different conformers, complexes or isomers. Fortunately, these can usually be disentangled through ultraviolet 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 ion-dip experiments conducted using tunable infrared laser radiation with the delayed UV ionization laser tuned to a band associated with the selected carrier, generates the IR spectrum associated with the selected band. Examples of both are shown in figure 4, which presents the 'intermingled' R2PI spectrum of phenyl β-D-glucopyranoside and its (three) component UV 'hole-burn' spectra, and the near- and mid-IR 'ion-dip' spectra of its three separately resolved conformers [24, 25].

Spectral, conformational and structural assignments are based upon a combination of several factors, which collectively 'home in' either on a unique assignment, or a limited set of alternatives that best fit observation. These include: the overall qualitative match between the patterns of observed and calculated vibrational spectra; the correlation between the relative conformer populations (estimated from the relative intensities of their band origins in the R2PI spectrum) and the ordering and magnitude of their calculated relative energies; and finally, the near universal rule that the most populated conformers/structures correspond to those which the *ab initio* calculations have located as those of lowest zero-point potential energy.

[1]Oven sources are much more stable and 'user-friendly' than desorption/ablation sources but problems generated by thermal decomposition have limited their use to the smaller systems (monosaccharides and their hydrates, and some disaccharides); for larger systems, desorption (which also uses much smaller samples) becomes the method of choice.

[2]Which requires a UV chromophore, usually provided by attaching a 'structurally benign' phenyl or benzyl tag to the O1 atom.

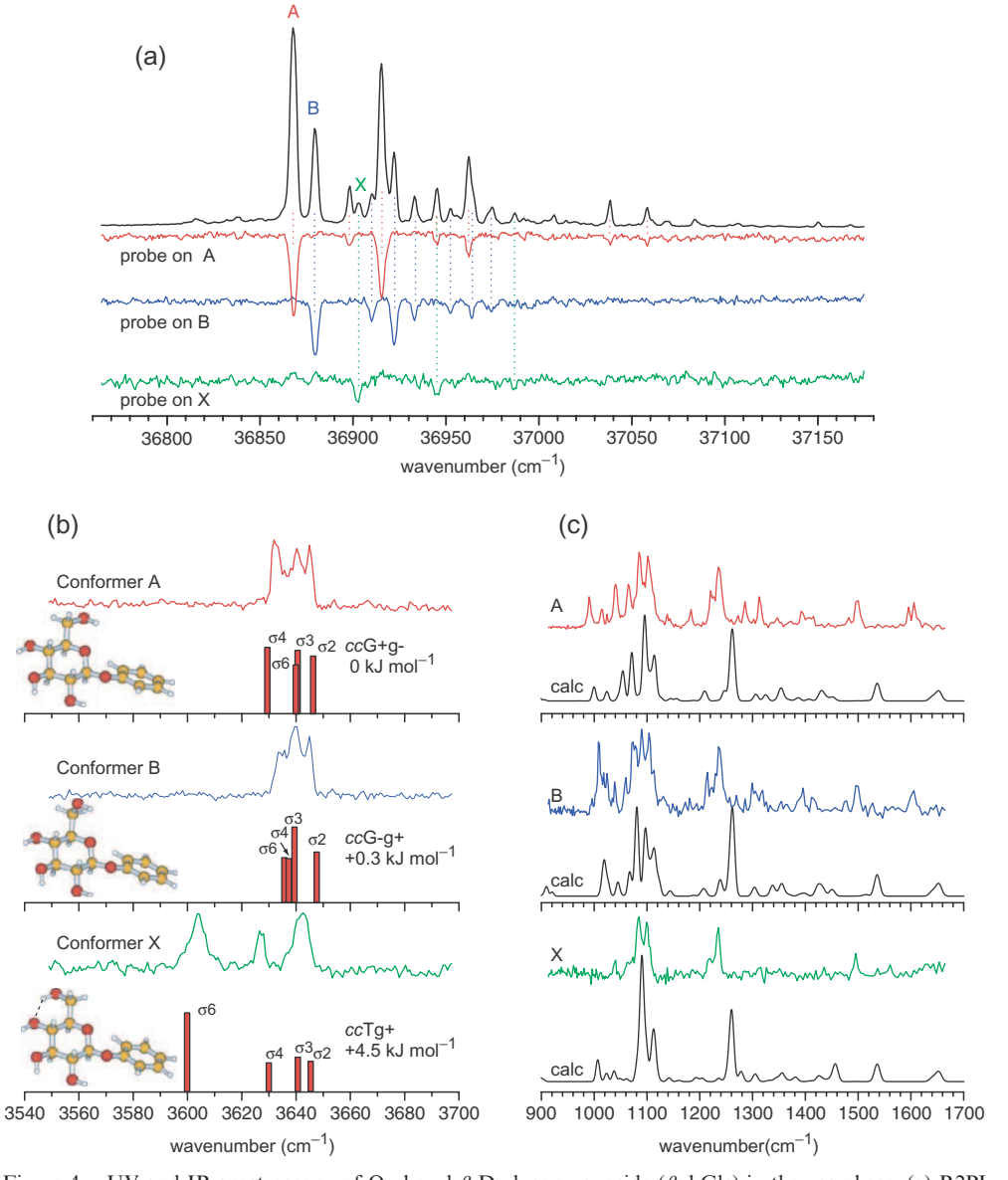


Figure 4. UV and IR spectroscopy of O-phenyl β-D-glucopyranoside (βphGlc) in the gas phase. (a) R2PI and UV 'hole-burn' spectra: A, B and X are the band origins of three different conformers. (b) and (c) Experimental and computed near- and mid-IR ion-dip spectra monitored via the band origins, A, B and X, together with the corresponding computed conformational structures and relative energies [24, 25]. Calculated mid-IR spectra are presented as convolutions of the computed spectra with a Gaussian line shape function.

The level of theory employed is a compromise between the known reliability of the method and the computational cost. The latter is strongly dependent on the size of the system and the number of possible structural permutations. In most cases, since many structures will be possible, a typical strategy begins with the generation of a large set of

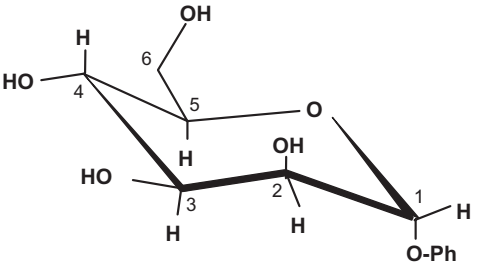
feasible conformers or cluster structures using a Monte Carlo search procedure implemented from a molecular modelling suite, and continues with the optimization of each one using density functional theory (DFT) and standard basis sets of moderate size, e.g. 6–31+G(d). Stationary points, characterized by calculation of harmonic frequencies using analytical second derivatives, yield the vibrational frequencies used to determine zero-point energy (ZPE) corrections and also frequencies and intensities for comparison with experimental IR spectra. Depending upon the level of theory used and the relevant IR spectral region, appropriate scaling factors can be introduced to compare computed frequencies with those actually observed. DFT/B3LYP calculations do not fully take dispersion interactions into account and the relative energies they provide can be unreliable; better estimates are usually obtained through single-point Møller–Plesset perturbation theory (MP2) calculations using a bigger basis set, typically 6–311++G(d,p). For the larger, more complicated molecules, a second round of conformational searching may be performed using either the results from experiments or the MP2 level relative energies as a guide. For example, in the model disaccharides, subsequent searches using a backbone (i.e. glycosidic torsion angles) fixed at low-energy geometries while allowing the OH groups to rotate results in closely related conformers that may have been missed in the initial conformation search. New candidate structures thus identified are then subjected to the same sequence of DFT optimization and frequency calculations and MP2 single-point energy evaluations. This enables a sufficient sample of the possible conformations and structures of the molecule under investigation to be made to generate a comprehensive set of candidate spectra and relative populations. These can then be compared with the experimental spectroscopic data to allow conformer-specific structural assignments of isolated sugars or their molecular complexes and, in the process, to provide a rigorous test of the computational modelling. To ensure this, the computed conformational landscapes of the phenyl (or benzyl) 'UV tagged' molecules need to be compared with those of the 'free' carbohydrates to which they are attached, to confirm (or not) the negligible influence of the tag on the structure of the isolated sugar. In all the examples to be discussed the UV tag has proved to be 'structurally benign'.

4. The conformational landscapes of some key monosaccharides: glucose, galactose, mannose, fucose and xylose

4.1. *Notation*

The flexibility of carbohydrate and oligosaccharide structures can generate a very large number of different conformations, though thankfully not too many of them are populated in practice. The notation [26] that is employed to describe them is introduced using a hexose, phenyl α-D-mannopyranoside, as a representative example. The carbon atoms of the sugar ring are numbered 1–6, starting from the anomeric carbon, C1, and finishing with the exocyclic hydroxymethyl group (scheme 1).

In the α-anomer as shown, the oxygen atom, O1, is oriented in an axial configuration; in the β-anomer it would occupy the equatorial position. The pyranose ring is shown in the $^4C_1$, *chair* conformation with the C4 and C1 atoms, respectively, 'up' and 'down'. Other possible ring conformations include the inverted chair ($^1C_4$), half-chair, boat or



Scheme 1. Phenyl α-D-mannopyranoside.

skew conformations. In the common hexose sugars, the $^4C_1$ chair conformation is, generally but not exclusively, the most favoured.

Monosaccharides differ from each other in their chirality at different carbon centres. For example, the OH2, OH3 and OH4 ring hydroxyl groups of glucose (in the $^4C_1$ chair configuration) are all equatorial, but in mannose, OH2 is axial and in galactose OH4 is axial. The hydroxyl groups can also be rotated and may be oriented in *trans*, (*t*) or *gauche* (*g*) conformations, dependent upon the H_n–O_n–C_n–C_{n+1}, $n = 2$–4, dihedral angles. *Gauche* conformations can be subdivided into *g*+ or *g*–, defined by dihedral angles lying between 0° and 120° or 240° and 360°. In the same way, the exocyclic hydroxymethyl group may adopt a T, G+ or G– conformation, dependent upon the heavy atom dihedral angle O_6–C_6–C_5–O_5, while the orientation of its terminal OH6 group, specified by the dihedral angle H_6–O_6–C_6–C_5, can assume a t, g+ or g– conformation. In phenyl α-D-mannopyranoside the minimum energy structure adopts the conformation $g+g+g$–G–g+, with the three ring OH groups linked in a 'clockwise' sense (when viewed in the orientation shown in scheme 2 with the ring oxygen located at the top right), to create a co-operatively H-bonded chain, OH2→OH3→OH4 [25]; because of the 'heaviness' of the detailed notation, however, the labelling is simplified here to cG–g+, where c denotes 'clockwise'.

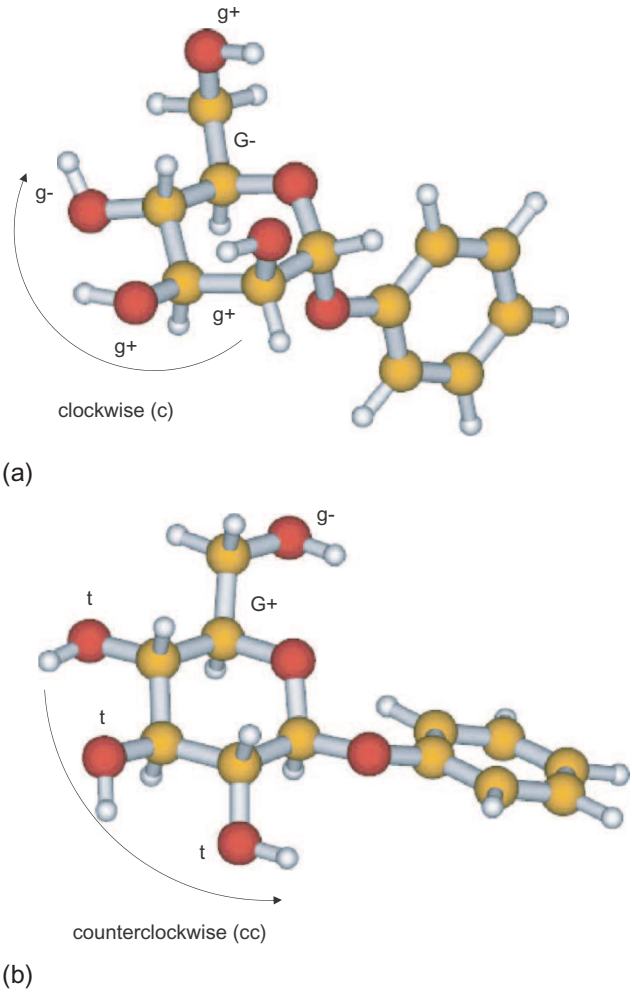
In contrast, the minimum energy conformation adopted by phenyl β-D-glucopyranoside [24, 25] is *ttt*G + g–, or more simply, ccG + g–, with the OH groups oriented in an all-*trans*, 'counter-clockwise' sense, OH4→OH3→OH2.

The OH stretching vibrations are labelled σn, where *n* refers to the number of the carbon atom to which the OH group is attached, numbered as indicated for phenyl α-D-mannopyranoside in scheme 1.

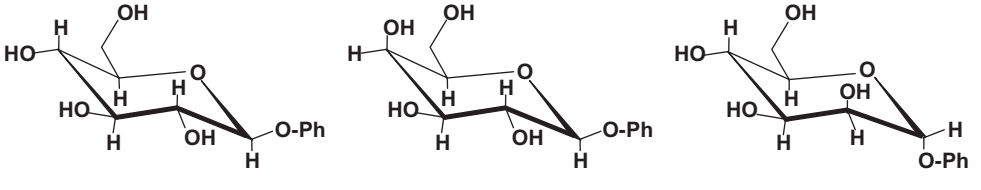
4.2. *Glucose, galactose and mannose*

The molecular structures of the three 'tagged' sugars, phenyl β-D-glucopyranoside (βphGlc), β-D-galactopyranoside (βphGal) and α-D-mannopyranoside (αphMan), are displayed in scheme 3.

The computed and near-IR ion-dip spectra of their global minimum energy conformers, which prove in each case to be the most strongly populated conformers in the free jet expansion [24, 25, 27], are shown in figure 5.



Scheme 2. (a) Phenyl α-D-mannopyranoside (cG–g+); (b) Phenyl β-D-mannopyranoside (ccG+g–).



Scheme 3. Monosaccharide molecular structures.

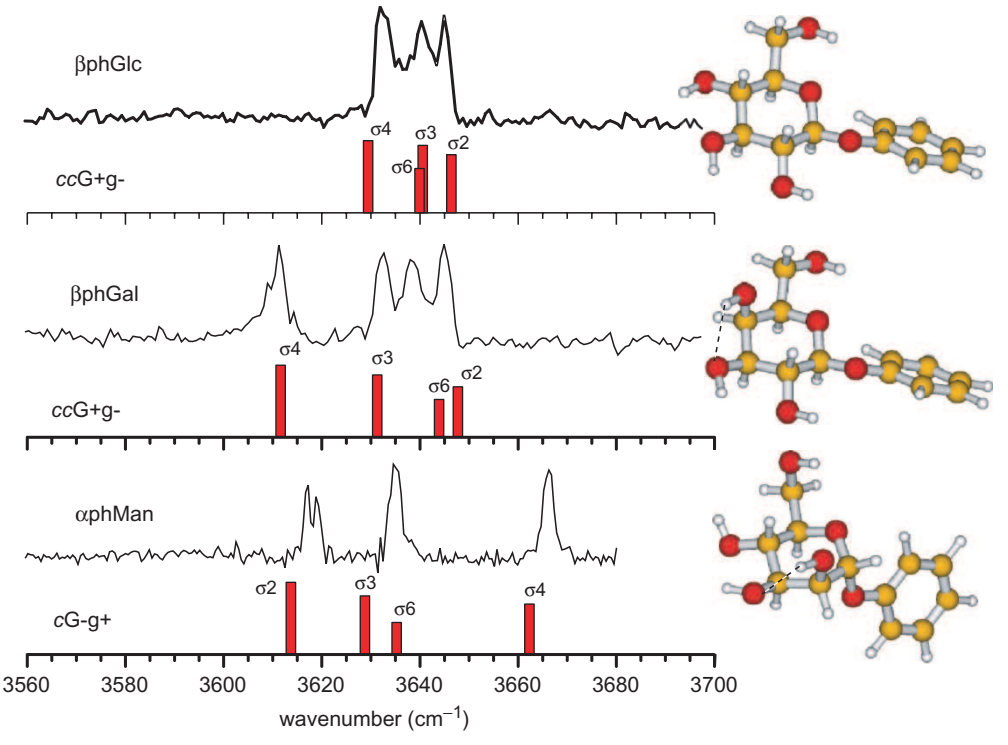


Figure 5. Computed and experimental near-IR spectra of the most stable conformers of phenyl 'tagged' glucose, galactose and mannose, (βphGlc, βphGal and αphMan) [24, 25, 27].

The spectra generated by the four O–H stretch modes, $\sigma 2$, $\sigma 3$, $\sigma 4$ and $\sigma 6$, are quite distinct and they reflect an extraordinary sensitivity to subtle changes in their hydrogen-bonded environment, whether promoted by changes in sugar structure or sugar conformation, a quality that is shared by *all* the sugars and oligosaccharides that have been investigated to date. The switch in the orientation of the OH4 group from equatorial in glucose to axial in galactose, for example, reduces the OH4$\cdots$O3 distance[3] from 2.41 Å (βphGlc) to 2.22 Å (βphGal). This is reflected in the shift of the vibrational band $\sigma 4$ towards lower wavenumber, displaced from the main cluster of bands, centred around 3640 cm^{-1}, by $\sim$25 cm^{-1}. These are themselves displaced by $\sim$30–40 cm^{-1} from the position of an unperturbed OH vibrational band (located in *trans*-ethanol, for example at 3673 cm^{-1}). More subtle changes are reflected in the relative displacements of the bands within the cluster. The least displaced OH band in the spectra of the minimum energy conformers of βphGlc and βphGal, $\sigma 2$, correlates with the longest OH$\cdots$O distance, OH2$\cdots$O1 $\sim$ 2.55 Å.

The most radical change in conformation is promoted by the change in configuration to an axial OH orientation at C2 in mannose, coupled with the change to an α-anomeric configuration at C1 in αphMan. The interaction between OH2 and O1 is no longer

[3]Determined through DFT calculation.

possible and the OH2 group flips around to focus its 'attention' instead on its nearest neighbour, OH3. Their separation is very much less, OH2···O3 ~ 2.23 Å, and the band $\sigma 2$ is now the most strongly shifted (its apparent splitting is adventitious – the dip is associated with an absorption band of atmospheric water vapour). More significantly, reflecting their co-operative interaction, the three OH groups all flick from the counter-clockwise 'cc' conformation adopted by βphGlc and βphGal, to the clockwise 'c' conformation in αphMan; this will prove to be particularly significant in later discussions of their hydrated structures, see section 6. The changed conformation directs the OH4 group away from OH3 and in αphMan, in contrast to βphGal, $\sigma 4$ is the least shifted band in the spectrum since the OH4···OH6 distance is now very long, ~3.19 Å.

Similar kinds of correlation can be discerned through comparisons of the experimental and computed IR spectra of different conformers of the same monosaccharide. A combination of UV hole-burn and IR ion-dip spectroscopy has revealed the population of several distinct conformers in each of the three examples [24, 25, 27]. Figure 6 displays for illustration the IR spectra and the conformational structures of one of them, βphGal, while table 1 lists the calculated relative energies of the most stable conformers of all three. The ring structures of βphGlc, βphGal and αphMan all adopt the $^4C_1$ chair conformation but their conformational landscapes are quite distinct. The most stable conformer of βphGal is analogous to that of βphGlc with the hydroxymethyl group in a *gauche*, G+ orientation. Its stronger OH4 ... OH3 interaction (a result of the change in configuration about C4) is easily visible in the corresponding IR spectrum. In its second lowest lying conformation the hydroxymethyl group of βphGal is rotated through ~130° from a G + g– to a Tg+ conformation but the counter-clockwise, 'cc', orientation of the peripheral OH groups is retained and both the G + g– and Tg+ conformers generate a similar set of H-bonded interactions. The most weakly populated conformer (~5% of the total ion signal, relative energy, 6.6 kJ mol^{-1}) adopts a clockwise orientation, however, cG–g+, which is the orientation favoured in the *minimum* energy conformation of αphMan. This allows all four OH groups to interact co-operatively, behaviour that is reflected in a general IR spectral shift towards lower wavenumbers. In addition, unlike mannose, the orientation of the OH4 group (*axial* in βphGal) coupled with the G–g+ orientation of the hydroxymethyl group, brings OH4 and OH6 into close proximity. The separation OH4...6 falls to ~1.92 Å to create a strongly H-bonded interaction which is signalled by the large displacement and enhanced intensity of the vibrational band, $\sigma 4$.

The close correlation between the near-IR spectra and the local molecular geometries demonstrates the power of such measurements for conformational and structural assignment; by measuring OH stretching frequencies, the experiments probe the relative H-bonding strengths, a major determinant of conformational stability. Mid-IR ion-dip spectra (some of which are shown later in this review) are dominated by C–O(–C) stretching and C–O torsional modes, and they probe coordinates along which conformational change can occur. While rich in information, it is difficult to relate mid-IR signatures to specific conformations since interpretations of this kind rely mainly on changes in transition intensity, which are both harder to measure and to calculate.

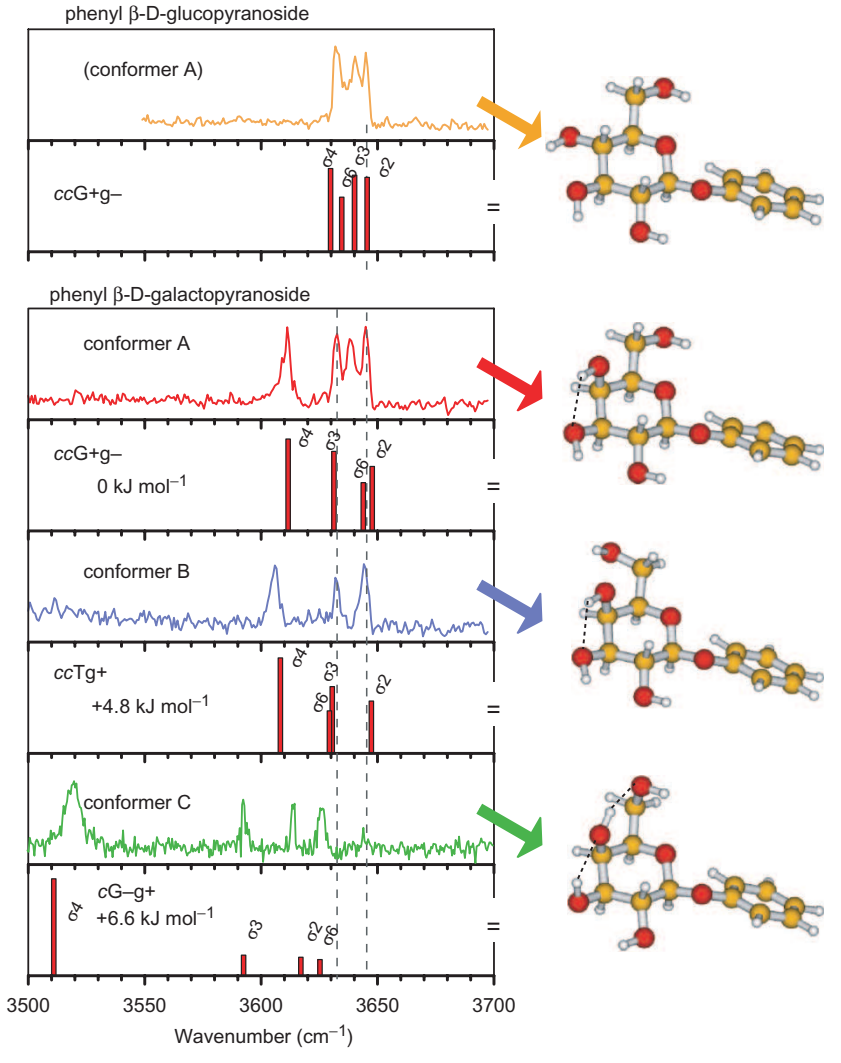


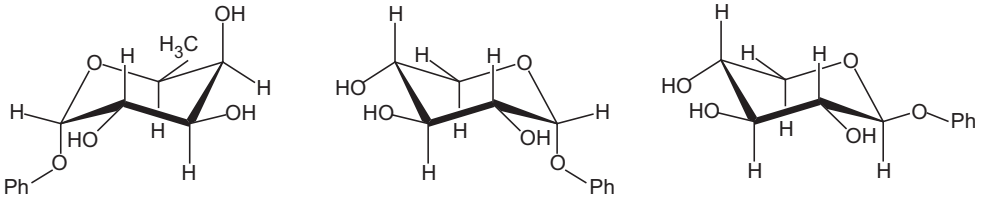
Figure 6. Computed and experimental near-IR spectra of the three assigned conformers of phenyl β-D-galactopyranoside (βphGal) [25, 27].

4.3. *Fucose and xylose*

In L-fucose, a close relative of galactose, the OH4 group is axially oriented but the exocyclic hydroxymethyl is replaced by a methyl group: L-fucose is the mirror image of 6-deoxy-galactose. Xylose is a pentose sugar analogue of glucose in which the hydroxymethyl group is replaced by an H atom. The molecular structures of the 'tagged' sugars, phenyl α-L-fucopyranoside, αphFuc, selected because of their importance in many molecular recognition processes (it is, for example, the immunodominant monosaccharide of the ABO blood group antigen determinants [29]) and phenyl α- and β-D-xylopyranoside, αphXyl, βphXyl (like α-L-fucose also of immunological importance) are displayed in scheme 4.

Table 1. Relative total energies (ΔE_0 in kJ mol^{-1}) of monosaccharide conformers, calculated at the MP2/6–311++G**//B3LYP/6–31+G* level of theory. ZPE corrections have been included from the B3LYP/6–31+G* frequency calculation. The numbers in parentheses indicate the energy ordering of the lowest lying conformers starting from the most stable. Conformers populated and detected in the jet are highlighted in **bold type**; their relative populations follow an ordering similar to their relative energies, though in βphGal a conformation, ccG–g–, is apparently 'missed' [24, 25, 27, 50].

	αphMan	βphGal	βphGlc
cG–g+	**0 (1)**	**6.6 (4)**	9.7
ccG–g+	**3.2 (2)**	15.3 (5)	**0.3 (2)**
cTt	**3.7 (3)**	20.0	13.4
cTg–	5.2 (4)	21.1	16.3
ccG + g–	5.7 (5)	**0 (1)**	**0 (1)**
ccTg+	7.4 (6)	**4.8 (2)**	**4.5 (3)**
ccG–g–	8.7	6.0 (3)	10.5



Scheme 4. Monosaccharide molecular structures.

The computed and measured near-IR ion-dip spectra of their minimum energy conformers, the only ones that are significantly populated in the free jet expansion [28, 30], are displayed in figure 7. All three are shown in a chair conformation but in the L-sugar, αphFuc, the structure is inverted, $^1C_4$ rather than $^4C_1$. The general features of their IR spectra are in good agreement with the proposed conformational assignments, each one of which presents a counter-clockwise chain of H-bonded OH groups. The notable difference between the IR 'signature' of the α- and β-xylosides reflects their different anomeric configurations. The (equatorial) OH2···O1 separation in βphXyl is 2.54 Å but in the α-anomer it falls to 2.25 Å, promoting a stronger H-bond and a shift of the associated band, σ2, by ~40 cm^{-1} towards lower wavenumber. The same (equatorial) OH2···O1 interaction is also favoured in αphFuc and σ2 is located at virtually the same position; the other two bands, σ3 and σ4, lie close to the corresponding bands in the cc conformers of the related monosaccharide, βphGal (cf. figure 6). The slightly increased spacing of the three bands in αphFuc, cf. their computed spacings, could be caused through weak coupling of the modes σ2 and σ3.

5. Probing the glycosidic linkage: lactose and glycan 'building blocks'

Having established the conformational landscapes of some of the key monosaccharide units involved in glycan structures the logical next step is to explore the structural

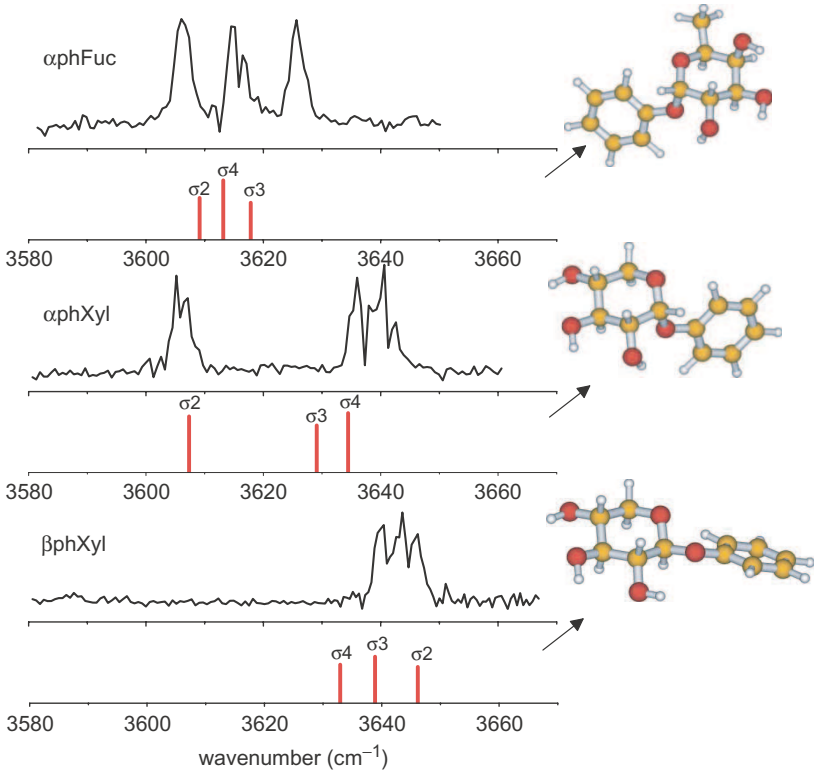
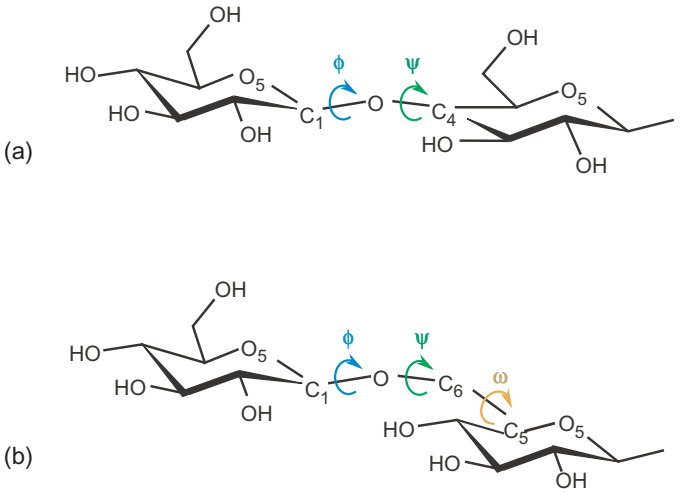


Figure 7. Computed and experimental near-IR spectra of phenyl α-L-fucopyranoside (αphFuc) [28] and α-D- and β-D-xylopyranoside (α,βphXyl) [29]. The apparent splitting of the central band in αphFuc is actually due to an absorption feature associated with atmospheric water vapour but (uniquely) each band in the IR spectrum of the xyloside is a doublet. The splitting is resolved in the α-anomer (~4 cm^{-1}) but is somewhat smaller in the β-anomer, where it generates an asymmetric line-shape. The splitting has been attributed tentatively to a tunnelling process involving ring-bending or OH motions [28].

consequences of their assembly into larger glycosidic units. In oligosaccharide chains, the heavy atom conformations of their component sugar ring structural units are generally assumed to remain fixed[4] but the same assumption cannot be made about the glycosidic linkages between them (or about the orientation of their exocyclic hydroxymethyl, peripheral OH or other substituent groups). The primary determinants of the overall 3D conformation(s) of oligosaccharide and glycan chains include the linkage points, e.g. 1–2, 1–3, 1–4, 1–6; their contributions to linear or branched chain structures; their anomeric structures, e.g. α(1–4), β(1–4); and the torsion angles about the glycosidic linkages – the primary focus of structural measurements and computational modelling.

[4]In conformational searches and *ab initio* optimizations the ring atoms are free to relax into alternative configurations, but to date the most stable predicted structures of all the D-sugars adopt the $^4C_1$ ring conformation and it is this conformation that is observed experimentally.



Scheme 5. Glycosidic torsion angles. (a) (1–4) linkage; (b) (1–6) linkage.

5.1. *Notation*

The angular structures about glycosidic linkages of the type $1 - n$, $n \neq 6$ (i.e. not involving an exocyclic CH_2OH group) are defined by the two dihedral angles, φ and ψ, in close analogy with the peptide link in amino acid chains. The two angles are often defined in terms of the heavy atom structures, $\varphi(O_5–C_1–O–C'_n)$ and $\psi(C_1–O–C'_n–C'_{n-1})$, see below, or alternatively, in terms of the sequences $\varphi_H(H_1–C_1–O–C_n')$ and $\psi_H(C_1–O–C_n'–H_n')$.

Linkage of a monosaccharide unit through the CH_2OH group provides an extra degree of flexibility, defined by the angle $\omega(O–C_6'–C_5'–C_4')$.

5.2. *Lactose*

The near-IR spectrum of the disaccharide lactose, tagged with a benzyl group to provide a UV chromophore, was presented earlier, in figure 2. Since its component units comprise galactose (Gal) and glucose (Glc), it presented an obvious first choice target for structural investigation [2]. The β(1–4) glycosidic linkage, which connects the two units, is a very common structural element, found in the core chitobiose unit of *N*-linked glycoproteins, GlcNAc β(1–4) GlcNAc, see figure 3, and in many carbohydrate structures such as those creating plant cell walls and insect and crustacean exoskeletons. Its ubiquity is an indication of its major biological importance.

Despite the potentially large range of conformations that might have been expected *a priori* in the vapour (different rotamers of the hydroxyl and hydroxymethyl groups in conjunction with three possible sets of glycosidic linkage torsion angles), only one is significantly populated in the free jet expansion [2]. As has been seen, under the same experimental conditions its component monosaccharides, Gal and Glc, each populate three conformers, associated with distinct orientations of the hydroxymethyl group

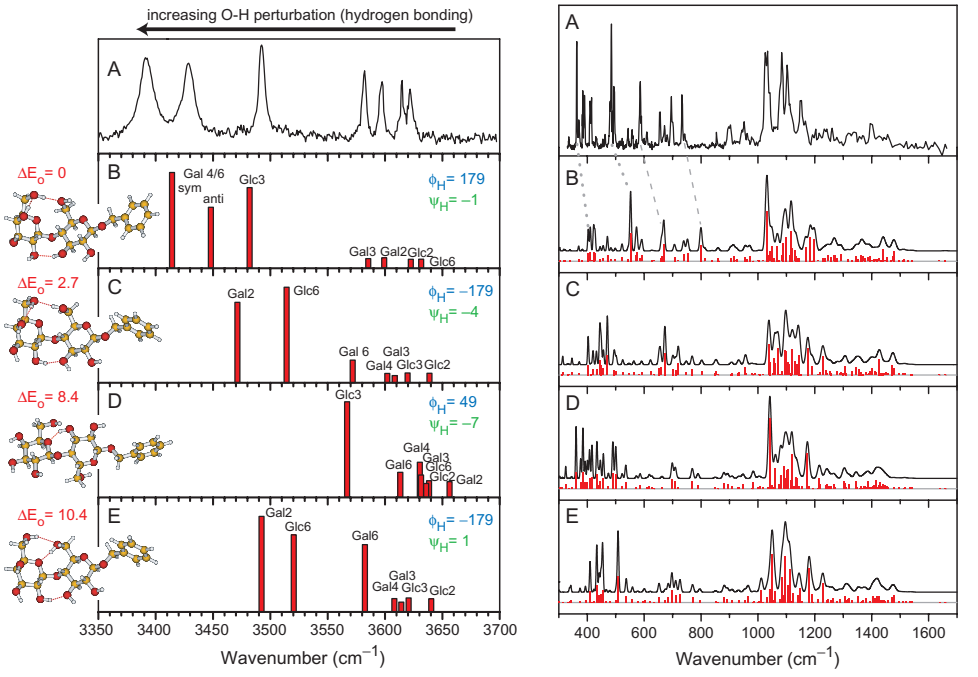


Figure 8. Experimental (A) and computed (B–E) near- and mid-IR spectra of benzyl β-D-lactoside [2]. The intense bands in the lower wavenumber region of the mid-IR spectrum are predominantly associated with coupled C–O torsional modes and are less well described by the calculations than the more localized modes at higher energy.

and/or their peripheral OH groups. When the near-IR spectrum of the lactoside is compared with the computed spectra associated with each of its low-lying conformations (four of the lowest are shown in figure 8), its assignment to the minimum energy structure is compelling [2], a conclusion that is reinforced by further comparisons based upon spectra recorded in the mid-IR 'fingerprint' region, see figure 8.

In the minimum-energy structure the seven OH groups form an uninterrupted hydrogen bonded chain, OH2′→OH3′→OH2→OH3→OH4→OH6→OH6′, which circles around the molecular framework, locking the orientation of the OH groups and, most significantly, generating a rigid scaffold across the glycosidic bond through the linked pair of inter-ring hydrogen bonds, OH3′→OH2 and OH6→OH6′. The strongly displaced vibrational modes associated with the last two members of the chain are strongly coupled and split into a symmetric and an antisymmetric pair, involving the sequence Gal:OH4–Gal:OH6–Glc:OH6′. The minimum-energy structure is achieved by rotation of the two hydroxymethyl groups from the G+g– conformation favoured in the free monosaccharides to the G–g+ conformation favoured in the disaccharide (and also favoured, as will be seen in section 6, in their mono-hydrate structures), and by rotation of all of the peripheral OH groups in the component Gal and Glc units from a counter-clockwise to a clockwise orientation. The co-operative hydrogen-bonded network which incorporates the inter-ring hydrogen bonding must provide a considerable enthalpic gain to offset the cost of the conformational change

(the cG-g+ conformers of the two *monosaccharide* units are located, respectively, ~7 kJ mol^{-1} and ~10 kJ mol^{-1} above their minimum energy conformation, ccG + g– [2]) as well as the entropic cost of the rigidified structure. The quartet of bands lying in the higher wavenumber range is associated with the four remaining hydroxyl groups, Gal:OH3, Gal:OH4, Glc:OH2′, and Glc:OH6′.

In solution at ~300 K, where NMR studies of lactose indicate the dominance of a far more flexible conformation, akin to structure D in figure 8, with the hydroxymethyl groups of the two rings adopting a *trans* orientation about the glycosidic linkage [3, 9, 31, 32], inter-ring hydrogen bonding is limited. The increased flexibility increases the entropy and at ~300 K structure D is calculated to have the lowest free energy, rather than the rigidified structure B [2].

5.3. *Mannose disaccharides*

The clear resolution of each one of the OH vibrational modes of the lactoside and the unambiguous correspondence between its experimental vibrational spectrum and the computed spectrum of its minimum energy conformer, both in the near-IR and in the 'fingerprint' mid-IR regions, has encouraged a similar 'attack' on the core structural units in *N*-linked glycans. Obvious first choices include the initial disaccharide unit, chitobiose (GlcNAc β(1–4) GlcNAc) and the α(1–3) and α(1–6) disaccharide components of the branched mannose trisaccharide unit that follows it, see figure 9.

The experimental and computed near-IR spectrum of the α(1–3) mannose disaccharide ('tagged' with a phenyl group and recorded under free jet expansion conditions) is presented in figure 10.

As with the lactoside, only a single conformer is populated in the free jet expansion but otherwise the contrast with βBnLac is dramatic. Apart from the one strongly shifted and intensified vibrational band, indicating a *single* strong band, assigned to the inter-ring Man b OH2→ Man a OH6 hydrogen bond σ2b, the rest of its IR spectrum does not shift towards lower wavenumbers. Unlike the β(1–4) linked lactoside, the peripheral groups of the α(1–3) linked mannose disaccharide are not oriented in a consistently co-operative fashion to circle the molecule as a whole but are disposed in a much more 'higgledy piggledy' manner. The single inter-ring H-bond rotates the hydroxymethyl acceptor on the Man a unit into a G + t orientation, directing it away from the neighbouring OH groups on the ring. The remaining weakly interacting set of OH groups {OH2a, OH3a, OH4a} adopt a 'counter-clockwise' orientation, which closely resembles that of one of the populated conformers, ccG–g+, of the monosaccharide, αphMan. Not surprisingly, there is a close correlation between the bands σ2a, σ3a and σ4a in the IR spectrum of the disaccharide and the corresponding bands, σ2, σ3 and σ4, of the monosaccharide conformer (see figure 11). The OH band at the high wavenumber end of the spectrum corresponds to σ6a, the 'free' OH6 vibration of the Man a unit.

In the α(1–6) mannose disaccharide, the glycosidic linkage is formed through one of the hydroxymethyl groups, introducing additional flexibility into the linkage structure. Its R2PI spectrum suggests the population of two conformers: the spectrum is shown in figure 12, together with the near-IR spectrum associated with each band. They are quite distinct and their association with two distinct conformers is confirmed. The absence of

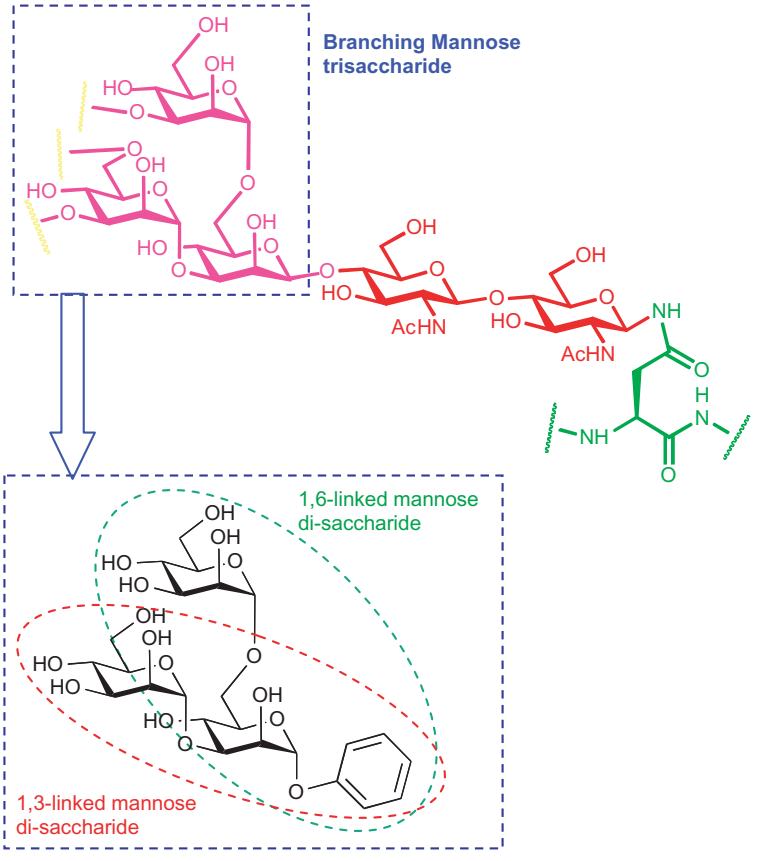


Figure 9. The conserved core structural scaffold in *N*-linked glycopeptides (top) and the two component disaccharides examined (bottom).

strong inter-ring hydrogen bonding in the more strongly populated conformer 1, in red in figure 12, is clear – *none* of its seven OH bands are shifted significantly towards wavenumbers below those of the OH bands in the three observed conformers of its component monosaccharide unit, αphMan. Furthermore, most of the bands appear to correlate with features in one or other of the isolated conformers, ccG–g+ or cG–g+ of αphMan. In contrast to the $\beta(1–4)$ linked lactoside, and the $\alpha(1–3)$ linked mannose disaccharide, the structure must be quite flexible. The near-IR spectrum of the minor conformer 2, in blue in figure 12, which displays an intense, strongly shifted band at $\sim$3500 cm^{-1}, does suggest an inter-ring hydrogen bond. A full conformational assignment of the $\alpha(1–6)$ disaccharide will follow the completion of current *ab initio* calculations.

Section 2 included the broad observation, 'The natural occurrence of common, conserved carbohydrate structural *motifs* must be related to their biological function. The underlying reasons for their evolutionary selection remain to be established but elegant studies [7–9] suggest their fundamental role as precise scaffolding units, providing support for the multivalent spatial presentation of key "recognition

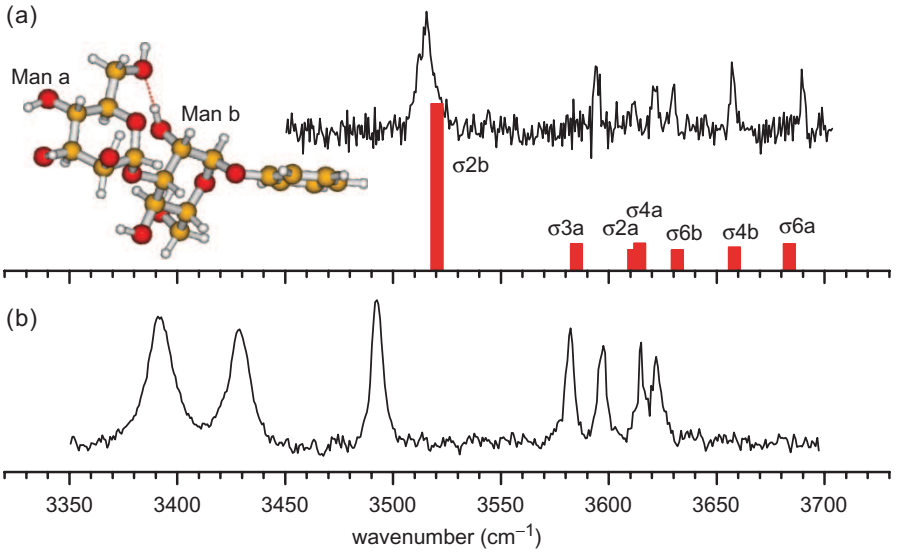


Figure 10. (a) Experimental and computed near-IR spectrum of the minimum energy conformer of O-phenyl α(1–3)-linked mannose disaccharide, recorded in a free jet expansion. (b) The corresponding spectrum of the benzyl tagged lactose disaccharide, βBnLac.

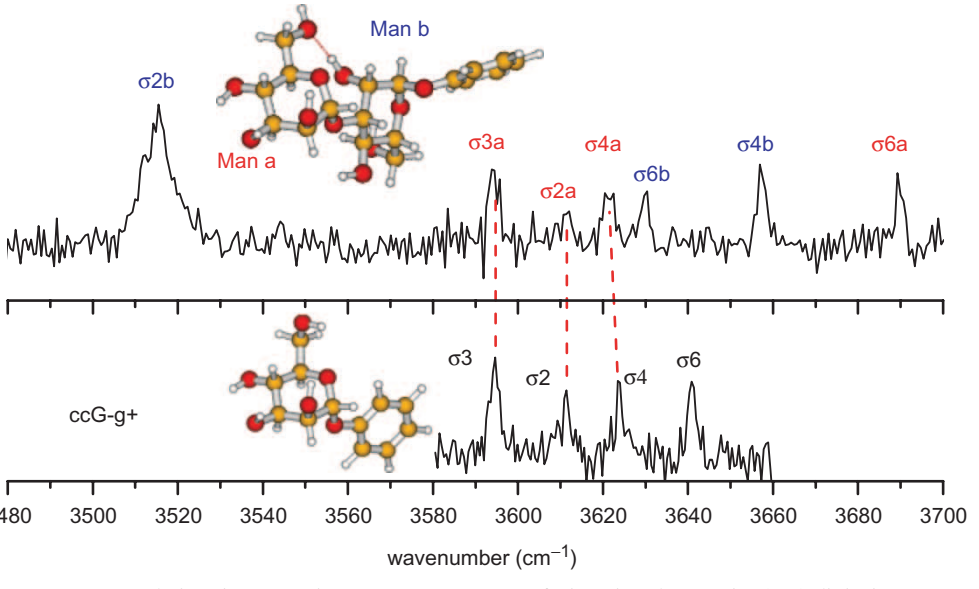


Figure 11. Correlation between the near-IR spectrum of the phenyl tagged α(1–3) linked mannose disaccharide and the IR spectrum the counter-clockwise conformer of αphMan.

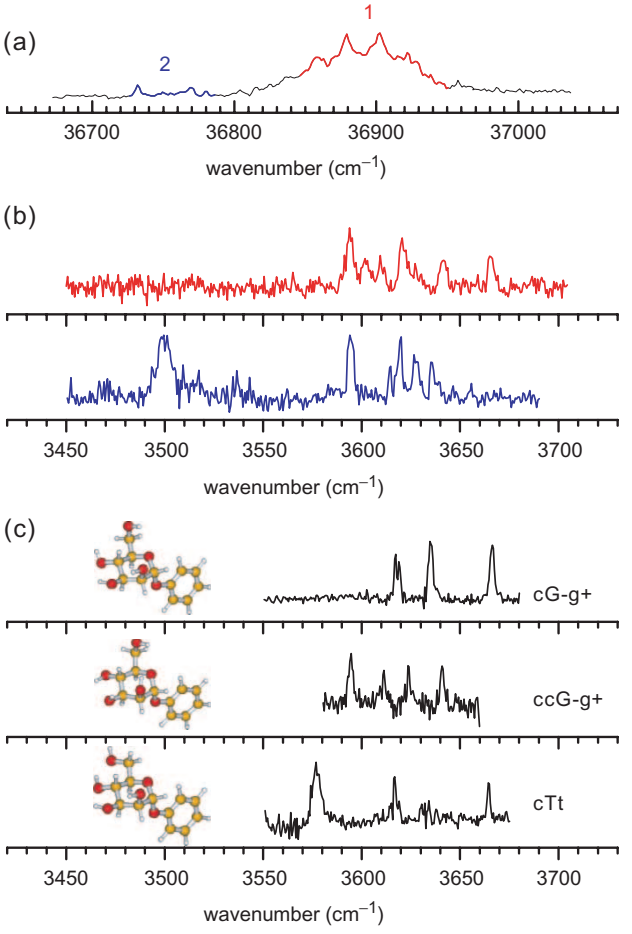


Figure 12. (a) R2PI and (b) near-IR ion-dip spectra of the phenyl tagged α(1–6) mannose disaccharide. (c) Near-IR ion-dip spectra of the three populated conformers of its component monosaccharide units.

antennae'' growing from the conserved core'. The pioneering structural investigations of lactose and the mannose disaccharides are beginning to provide additional support for this assertion. The leading β(1–4) linked chitobiose disaccharide unit $(GlcNAc)_2$, see figure 3, is likely to adopt the same rigid conformation as lactose (enhanced perhaps by the substitution of an acetamide group in the 2-position) but the branched mannose trisaccharide unit that follows it may well be intrinsically more flexible, as the mannose disaccharides appear to be. NMR experiments combined with molecular dynamics simulations on the $Man_9GlcNAc_2$ oligosaccharide unit [7] and the branched trimannoside [33] do indicate flexibility in these mannose linkages and it has been suggested that solvent-mediated hydrogen bonding between the mannosyl rings helps to modify the structure [7]. Taken together these results suggest *N*-linked glycopeptide structures in which the glycan antennae grow out of a rigid chitobiose spur through a more flexible branched mannose trisaccharide structure, which helps the antennae to adjust the

spatial presentation of their termini and optimize the specific, multivalent recognition processes in which they are involved.

6. Adding water to sugar: hydrogen-bonding, co-operativity and selectivity

Hydrogen bonding is one of the key components in the machinery of carbohydrate molecular recognition and, as has now been demonstrated for mono- and disaccharides, in the maintenance of their preferred conformational structures. *Intra*-molecular hydrogen bonds within and between individual and linked monosaccharide units can work together in a co-operative way to stabilize favoured conformations. The generation of co-operative hydrogen-bonded networks involving neighbouring hydroxyl groups may also play a major role in molecular recognition [34–37]. Many studies have identified water not simply as a spectator solvent but as an active participant in carbohydrate-protein molecular recognition processes [38–49] and the influence of this ubiquitous biological solvent on carbohydrate conformational landscapes needs to be understood if its interaction with biological molecules is to be properly modelled. Following this aspiration, the spectroscopic and computational strategies that have proved so successful for some key, isolated carbohydrates are being extended to include the structural investigation of their singly (and multiply) hydrated complexes [25, 28, 30, 50]. As will be seen, the results are revealing how water can come into play to 'reshape' to its own advantage the host carbohydrate conformation and they are providing a much clearer picture of the way in which this is achieved. Hydrogen bonding and co-operativity are the driving forces: the near-infrared measurements provide a 'spectroscopic indicator' of their manifestation [25].

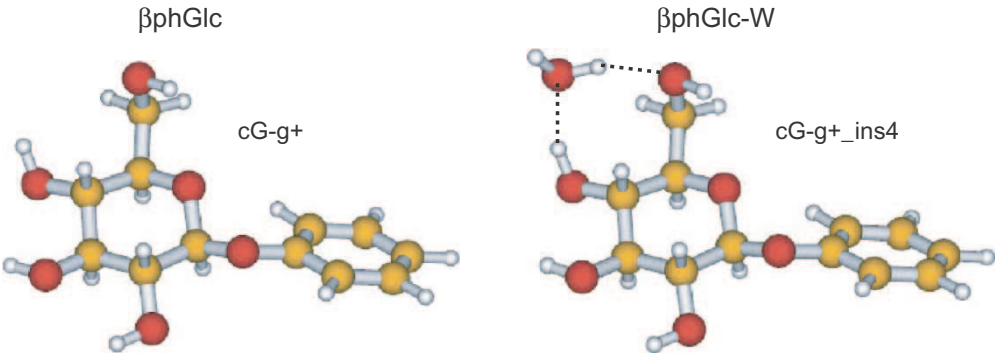
Comparisons of protein-carbohydrate crystal structures with the favoured water-binding sites on carbohydrates have provided strong evidence for proteins recognizing conformational structures preferred by the *hydrated* carbohydrate [25]. Structural water molecules present in the crystal are frequently located at the same carbohydrate positions as those identified in the gas-phase structures. Where the water molecule is displaced, it often does so to favour interaction with the protein at the *same* selected site on the carbohydrate.

6.1. *Notation*

Once again, the simplified notation 'c', and 'cc' will be used in what follows and the conventional notation describing the conformation of the hydroxymethyl group is retained. If a water molecule is inserted between two hydroxyl groups on the sugar, bound as an H-bond acceptor from the group OH*n*, then it is labelled as 'ins *n*'. Thus insertion of a water molecule bridging across the groups OH4 and OH6 in a glucose conformation, βphGlc (cG–g+), would be termed 'cG–g+_ins 4' (scheme 6).

6.2. *Mono-hydrated complexes: glucose, galactose and mannose*

When phenyl-tagged β-D-gluco-pyranoside, βphGlc, is expanded into a moist, expanding Ar jet, new resonant two-photon ionization signals can be detected corresponding to singly and multiply hydrated complexes [25, 50]. The R2PI and UV



Scheme 6. Molecular structures of βphGlc and its monohydrate.

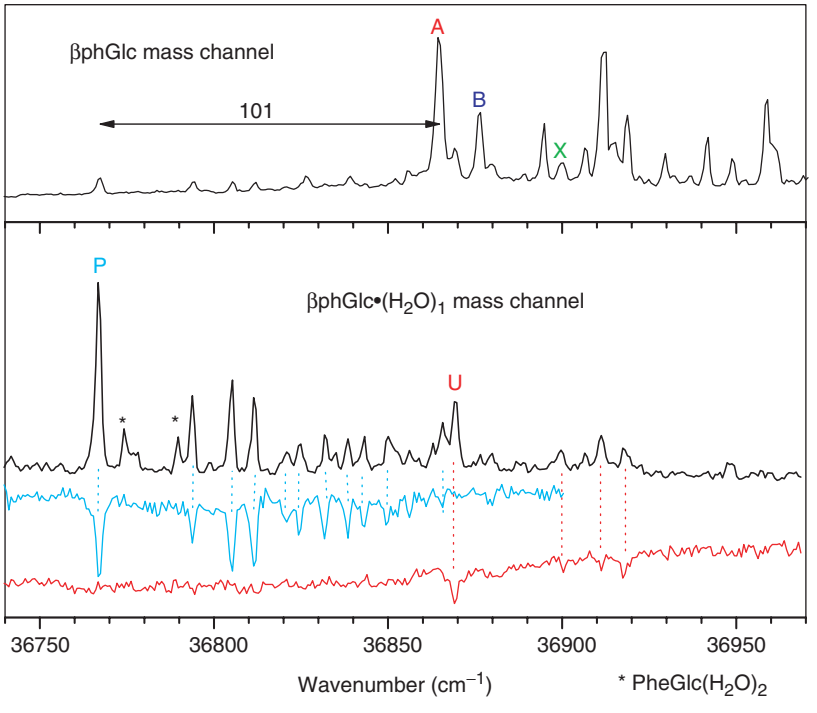


Figure 13. R2PI and UV 'hole-burn' spectra of the singly hydrated complexes of βphGlc [50]. Peaks marked with an asterisk are associated with a doubly hydrated complex. The R2PI spectrum recorded in the $\beta phGlc^+$ ion channel, primarily corresponding to the 'bare' carbohydrate, is shown in the top frame for comparison.

'hole-burn' spectra recorded in the $\beta phGlc–W1^+$ and $\beta phGlc^+$ ion channels are shown in figure 13. Two distinct mono-hydrate structures, labelled as 'P' and 'U', can be identified (and also a few weaker features associated with a doubly hydrated complex, which partially fragment following ionization).

The near and mid-IR ion-dip spectra associated with the two mono-hydrates are shown in figure 14, where they can be compared with those calculated *ab initio* for the

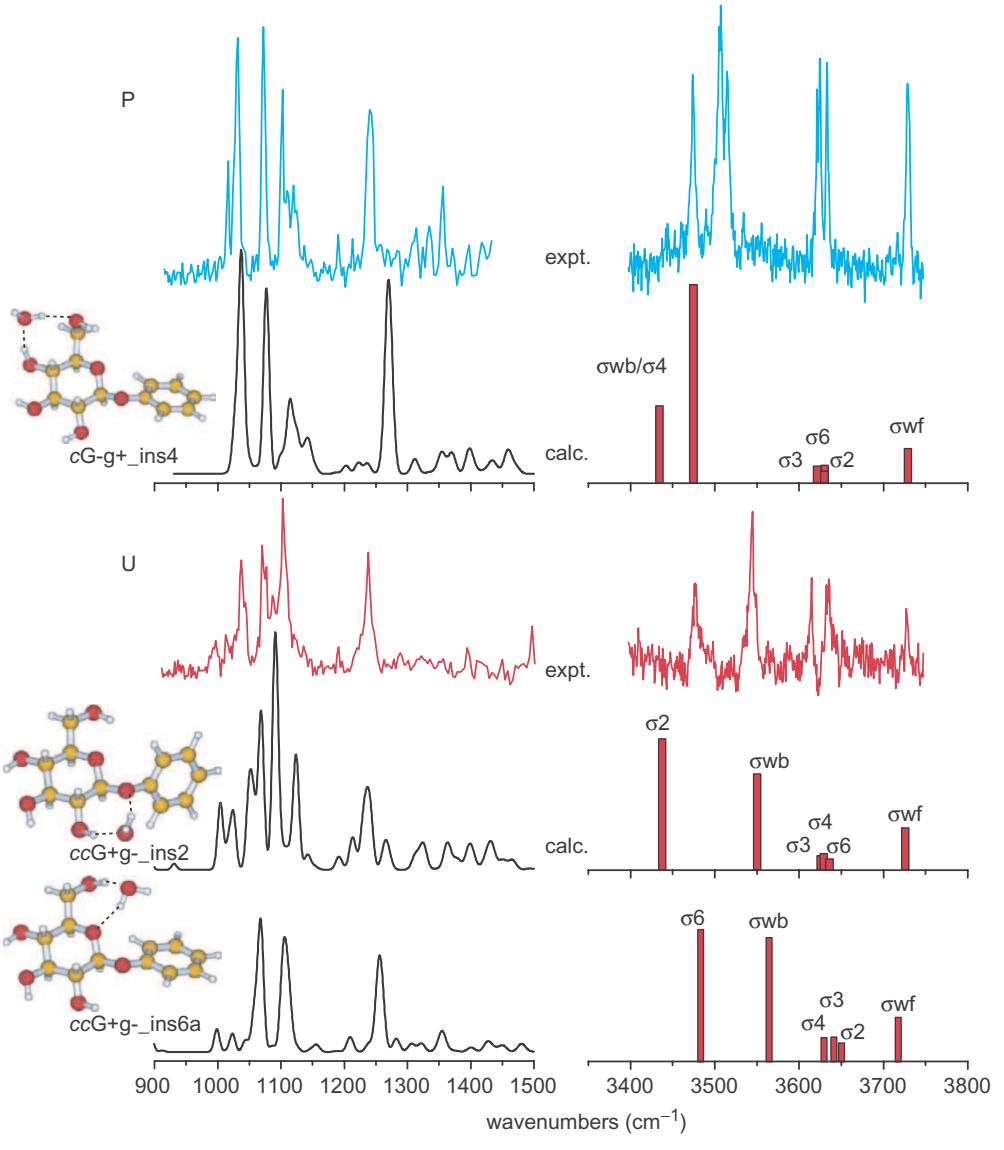


Figure 14. Experimental and calculated mid- and near-IR spectra of the two mono-hydrate complexes, βphGlc-W, labelled P and U in figure 13 [25].

three computed lowest energy mono-hydrate structures. Their experimental near-IR spectra each present the characteristic 'signature' of an insertion complex displaying two intense bands strongly shifted towards lower wavenumbers, symmetric and antisymmetric modes comprising the coupled vibrations, σn and σwb ('wb' = water-bound) associated with the two H-bonds in the chain OHn→OHw→O, and a single unshifted band at high wavenumber, σwf, associated with the 'free' OH of the bound water molecule. The correspondence between the experimental IR spectra associated

with the complex P, figure 14(a), and those calculated for the global minimum energy hydrate structure, cG–g+_ins4, is compelling – and also very striking since the 'clockwise' conformation cG–g+ (shown in scheme 6) of the carbohydrate moiety lies ~10 kJ mol^{-1} above the 'counter-clockwise' global minimum energy conformation of the 'bare' carbohydrate, βphGlc ccG+g–, see table 1. Not surprisingly, the cG–g+ conformer is not populated in the 'bare' carbohydrate; clearly *water binding drastically changes the preferred conformation of the sugar*.

The structural assignment of the less abundant mono-hydrate of βphGlc, U, is not quite as clear-cut. Either of the two lowest-lying structures located just above the global minimum, both of which *do* incorporate the favoured conformation of the bare carbohydrate, ccG+g–, with the water molecule inserted between OH2 and O1 or OH6 and O_{ring}, can provide a reasonable but not quite perfect match between the experimental and computed spectra. What is clear-cut is the changed conformation of the more abundant hydrate structure, P.

The galactoside hydrates follow a very similar pattern of behaviour[5] [25]. The strongly populated mono-hydrate, P, of βphGal has been assigned to an insertion complex with the water molecule located between the OH6 and O_{ring} of the cG–g+ conformer to form the structure cG–g+_ins6. The structure of the other strongly populated hydrate U is, like that of βphGlc, a little less clear. Certainly it incorporates the ccG+g– conformer but the water insertion could be either between OH6 and O_{ring} (more likely) or OH2 and O1, analogous to the candidate structures for conformer U of the glucoside hydrate (see figure 14). Once again hydration of the carbohydrate favours a conformational switch to the 'clockwise' orientation, cG–g+; in the bare carbohydrate this conformation lies 6.6 kJ mol^{-1} above the global minimum and its population is only just detectable, see table 1 and figure 6.

The minimum-energy conformation of αphMan already corresponds to the 'clockwise' orientation, cG–g+, see figure 5 and table 1, and like the hydrated complexes of βphGlc and βphGal, this is also the conformational structure adopted by its mono-hydrate [25]. The R2PI spectrum of αphMan-W1 reveals the population of a single conformer only and its near-IR spectrum matches best the spectrum computed for the global minimum energy structure, cG–g+_ins 4, see figure 15. Since the relative energies computed for the two nearest alternative structures lie ~6 kJ mol^{-1} higher, the population of a single conformer reinforces its proposed assignment to the global minimum.

A further aid to support the proposed structural assignments is provided by the correlation summarized in figure 16, which collects all the R2PI spectra of βphGlc, βphGal, αphMan and their mono-hydrated complexes.

Strikingly, the UV R2PI spectral band origins of all the observed conformations incorporating a hydrogen bond from OH2 to the phenoxy oxygen of the UV chromophore (O1), are located above ~36 830 cm^{-1} (indicated by a vertical dashed line). Band origins of structures which lack this interaction are all found at lower wavenumbers and in each case include the most populated mono-hydrate structures. Table 2 provides a summary of their computed relative energies [25].

[5]Though its R2PI spectrum is much more congested and other structures may also be populated. *Ab initio* calculations indicate five low-lying hydrate structures lying within 1.5 kJ mol^{-1} of the global minimum [25].

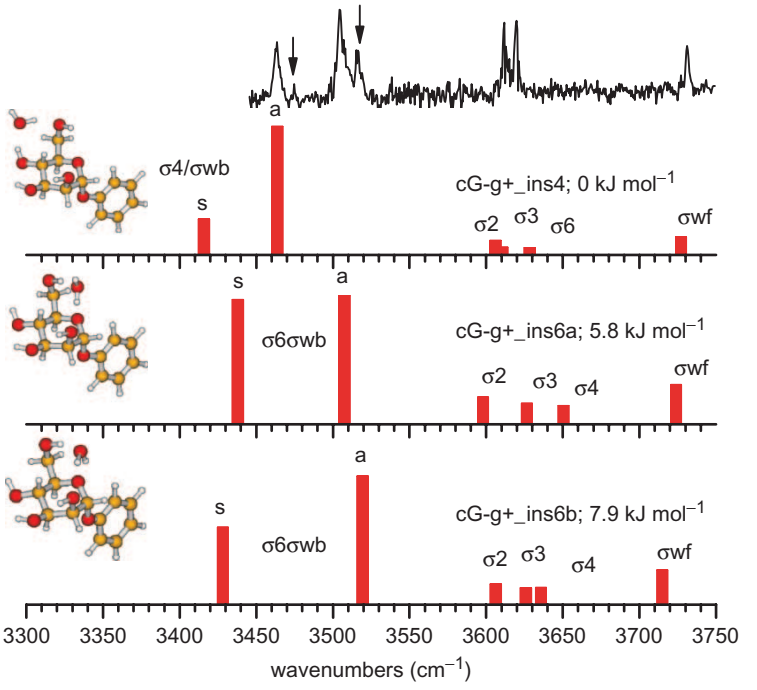


Figure 15. Experimental and computed near-IR spectra of the hydrated complex, αphMan-W1 [25]. 's' and 'a' represent the symmetric and antisymmetric vibrations of the coupled modes, σn and σwb, associated with the insertion structures. (The two additional features indicated by arrows are most likely associated with combination bands, involving a low wavenumber, intermolecular vibrational mode.)

6.3. *Co-operativity and conformational selectivity*

Given the variety of conformational landscapes presented by the bare monosaccharides, αphMan, βphGlc and βphGal, their clear preference for the cG–g+ conformation when they are hydrated by a single water molecule is remarkable. In the case of αphMan, the preference could have been expected since the cG–g+ conformation is also preferred in the bare molecule but the situation is quite different for βphGal and βphGlc (cf. table 1) where their preferred structures and conformational landscapes are considerably altered. The strong conformational selectivity can be understood in terms of co-operative inductive interactions which enhance the strength of the hydrogen-bonded network of linked OH groups. In the hydrates of αphMan and βphGlc, the water preferentially inserts at position 4 and the chain includes the sequence $OH2\rightarrow OH3\rightarrow OH4\rightarrow OH_w\rightarrow OH6\rightarrow O_{ring}$; in βpGal-W, the water inserts at position 6 to generate the sequence $OH2\rightarrow OH3\rightarrow OH4\rightarrow OH6\rightarrow OH_w\rightarrow O_{ring}$. *In each case, the preferred binding site corresponds to the weakest of the intramolecular H-bonds in the host carbohydrate, replacing it with two strong intermolecular H-bonds.*

There are close parallels with the stabilization identified in the minimum-energy conformation of the β(1,4) linked disaccharide, βBnLac (discussed in section 5.2) where the chain is extended to include seven OH groups. They lock into a clockwise orientation circling around the molecular framework, see figure 17(a), and generate a rigid scaffold across the glycosidic bond through the linked pair of inter-ring

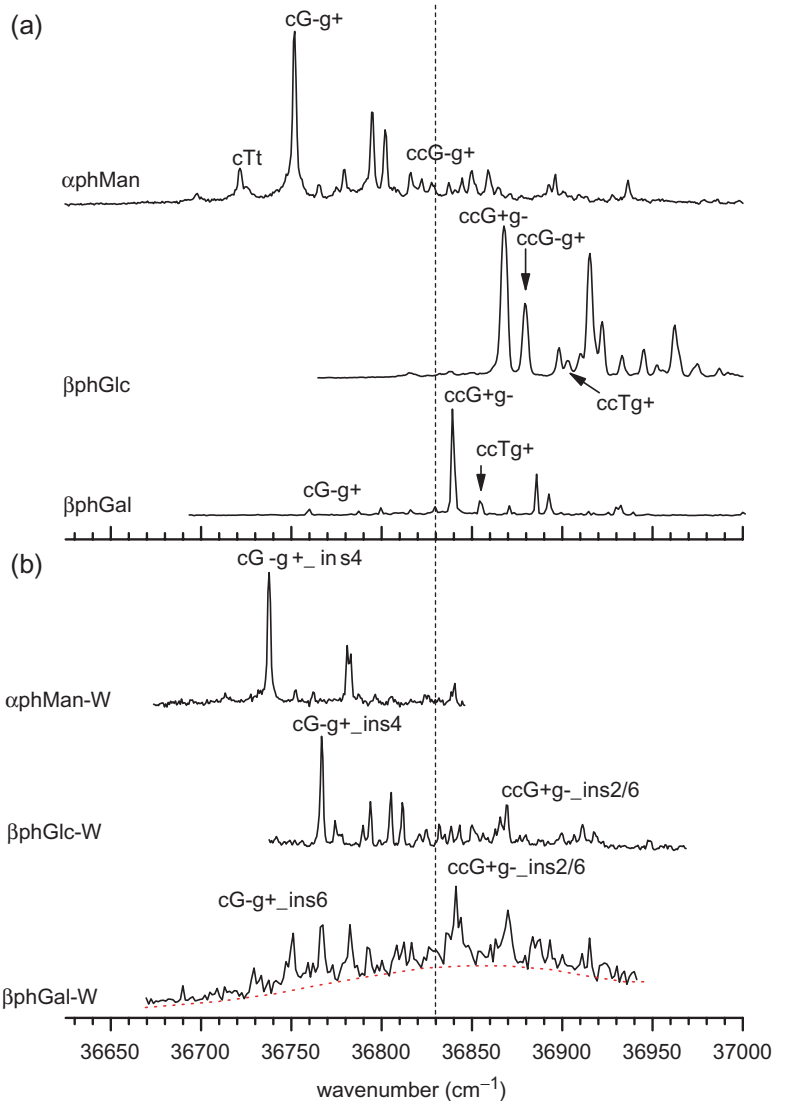


Figure 16. R2PI spectra of (a) bare monosaccharides and (b) their hydrated complexes [25]. The vertical dashed line marks the separation between the spectral origins of conformers associated with (above the line) and without (below) an OH→ O1 hydrogen bond.

hydrogen bonds. The cG–g+ conformations of the Gal and the Glc components needed to achieve this arrangement parallel the conformations adopted in their favoured monohydrate structures. This is illustrated in figure 17(b) which superimposes the most stable structure of hydrated βphGal over that of the unhydrated lactoside, βBnLac. The water molecule in βphGal-W plays the same role in stabilizing the cG–g+ conformation of the galactosyl unit as the OH6 group in the glucosyl component of the lactoside. Indeed, the *intra*molecular (Gal) hydrogen bond geometries are so similar in βphGal-W and βBnLac that their OH2, OH3 and OH4 groups superimpose almost perfectly.

Table 2. Calculated relative energies (kJ mol^{-1}) of mono-hydrated structures discussed in the text. The numbers in parentheses indicate the energy ordering for the most stable structures [25].

		αphMan-W	βphGal-W	βphGlc-W
cG–g+	ins4	0 (1)	14.0	0 (1)
	ins6a	5.8 (2)	0.9 (2)	6.2
	ins6b	7.9 (3)	0.3 (4)	6.7
ccG + g–	ins2	19.3	0.6 (3)	0.9 (2)
	ins6a	22.9[a]	0 (1)	1.1 (3)
	ins6b	25.7[a]	1.5 (5)	1.2 (4)

[a]These conformations feature a phenyl ring orientation stabilized by the presence of the water molecule. In each of the most stable, cG–g+, conformations of αphMan and αphMan-W, the phenyl ring is oriented perpendicular to the sugar ring; in its ccG + g–_ins6a/b complexes, however, there is a parallel orientation and the water bridges between the OH6 group and the phenyl ring.

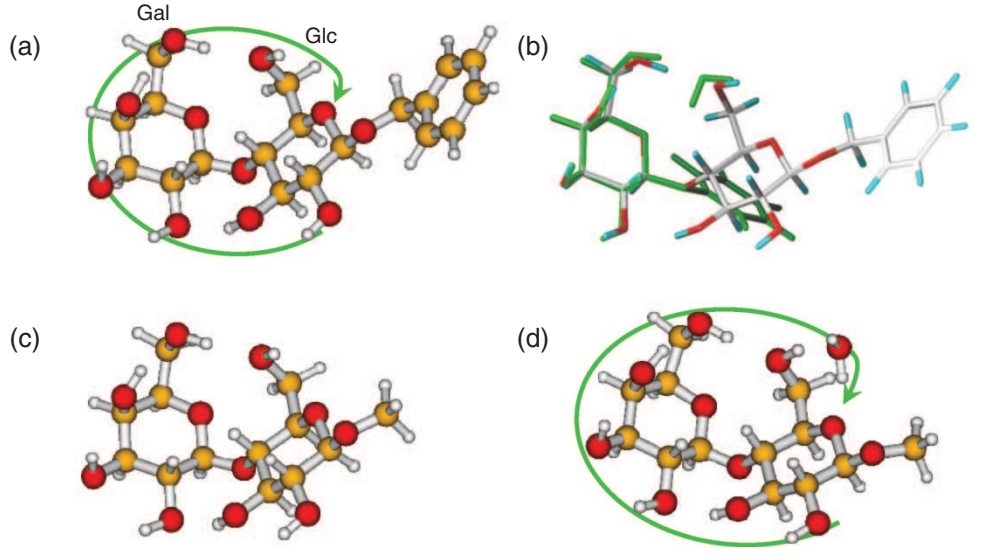


Figure 17. (a) Minimum energy structure of benzyl β-lactoside (βBnLac); (b) superposition of the minimum energy structures of βBnLac and the hydrated galactoside, βphGal-W; (c) and (d) computed minimum energy structures of methyl β-D-lactoside and its hydrate.

A further indication of the enhanced co-operativity created by the extended H-bonded network in the β(1–4) disaccharide is revealed through the near-IR spectral comparisons presented in figure 18. The universal shift to lower wavenumber of *all* the OH stretching vibrations in βBnLac would appear to reflect a stronger co-operative effect in the disaccharide (where there are seven links in the chain) than in the component monosaccharide–water complexes (where there are only five). In particular, in the absence of any additional co-operativity, the bands σ_2(Gal) and σ_3(Gal) would be expected to appear at the same wavenumbers in the disaccharide and in the mono-hydrate since they are involved in the same *intra*molecular hydrogen bonds in both cases.

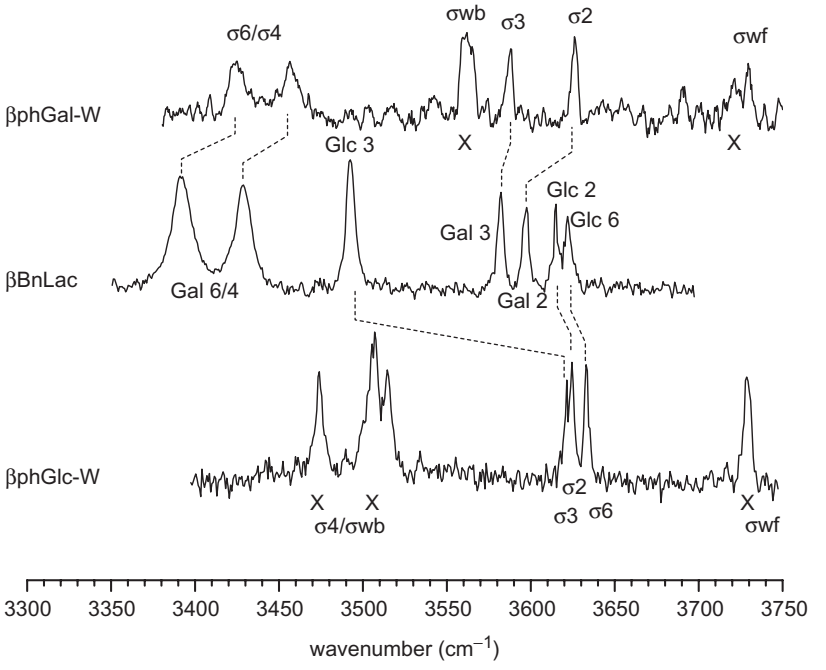


Figure 18. Comparison of the IR spectra associated with the minimum energy, cG–g+, structures of βphGal-W, βphGlc-W and βBnLac (where both the Gal and Glc groups adopt the cG–g+ conformation). The lines in the spectra of the monosaccharide-water complexes marked by a cross correspond to stretching modes of OH groups that are not present in the disaccharide.

Suppose the disaccharide were hydrated: what effect would this have on its structure, and in particular, the geometry about the glycosidic linkage? Figures 17(c) and (d) compare the computed minimum-energy structures of methyl β-lactoside, which has a similar conformational landscape to βBnLac (revealed at a reduced computational cost) and its mono-hydrate. The water molecule inserts between the most loosely H-bonded groups – in this case between the OH6 and O_{ring} sites on the Glc moiety of the disaccharide (figure 17d). Again, as with the hydrated monosaccharides, there is a clear preference for the bound water molecule to replace the weakest *intra*molecular hydrogen bond by two strong *inter*molecular bonds: (the next most stable hydrate lies $\sim$6 kJ mol^{-1} higher in energy). The disaccharide structure, and perhaps those of β(1–4) linked oligosaccharides in general, retains the fully co-operative *cis* configuration of its two hydroxymethyl groups, and the two strong hydrogen bonds connecting OH3(Glc) to OH2(Gal) and OH6(Gal) to OH6(Glc) are conserved.

6.4. *Mono-hydrated complexes: xylose and fucose*

Xylose, unlike Man, Gal and Glc, is not blessed with an exocyclic hydroxymethyl group, which reduces the range of alternative conformational choice; as discussed in section 4.3 its two anomers both adopt their global minimum, counter-clockwise (cc) conformations and no other structure is detectably populated. Despite their similar H-bonded

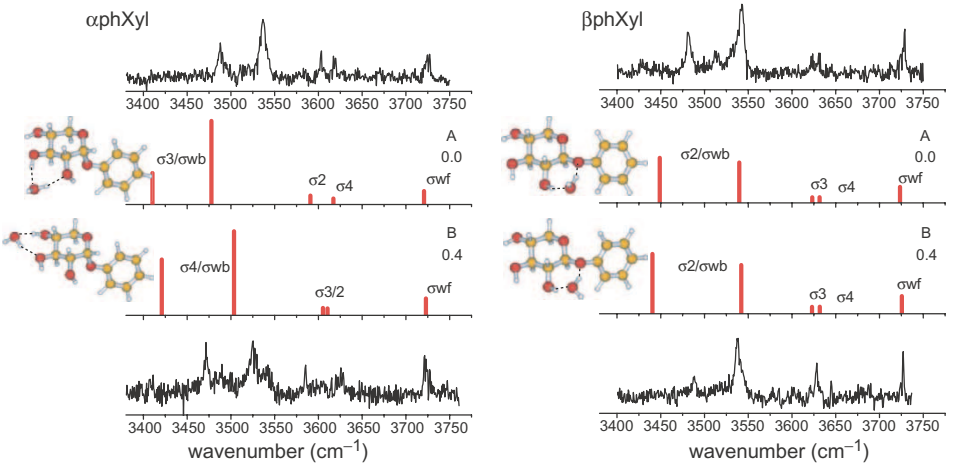


Figure 19. Experimental and calculated IR ion dip spectra of the singly hydrated α and β anomers of phXyl [30]. Relative energies of the two pairs of structures, A and B, are in kJ mol^{-1}. In the α-anomer the water molecule inserts into the OH2,3 or OH3,4 sites but in the β-anomer where the OH2→O1 site becomes the most 'vulnerable', it is this site that is selected.

sequences, OH4→OH3→OH2→O1, their IR spectra are quite distinct, however [30]. In the α-anomer, the cc orientation of the peripheral OH groups favours hydrogen bonding between OH2 and O1 and the vibrational band, σ2, shifts towards low wavenumber; in the β-anomer, σ2 is the *least* shifted band. The distinction persists in their mono-hydrates and following the pattern shown by αphMan, βphGlc and βphGal, water again 'goes for the weakest link'. This leads to a remarkable degree of selectivity. In βphXyl the molecule inserts between OH2 and O1 (in two alternative orientations, see figure 19) to create the structure cc_ins2 incorporating the H-bonded sequence OH4→OH3→OH2→W→O1 but in the α-anomer, the weakest links correspond to OH4→OH3 or OH3→OH2. These two alternative sites are now the ones selected for insertion, generating the extended 'cc_ins 4' sequence, OH4→W→OH3→OH2→O1 (or the alternative 'cc_ins 3' sequence), see figure 19 [30]. The selective behaviour may also help to explain the preference for populating the β-anomer in aqueous solution [51, 52].

The global minimum-energy structure adopted by the α-anomer of L-fucose, αphFuc (where the hydroxymethyl is now replaced by a methyl group), is quite similar to that of αphXyl. The counter-clockwise (cc) orientation of the peripheral OH groups[6] is retained, which once again favours H-bonding between OH2 and O1. In the hydrate, the site of weakest interaction favours insertion of a water molecule between OH3 and OH2 and as expected this structure corresponds to the global minimum and is much the most strongly populated in the free jet expansion. A second, more weakly populated structure can also be detected, which carries the IR spectral signature of an *addition* complex. This presents only one strongly shifted band rather than the characteristic

[6]When the sugar is viewed in the 'standard' orientation with C6 at the top and the ring oxygen atom at the top right, rather than the orientation used in the figures.

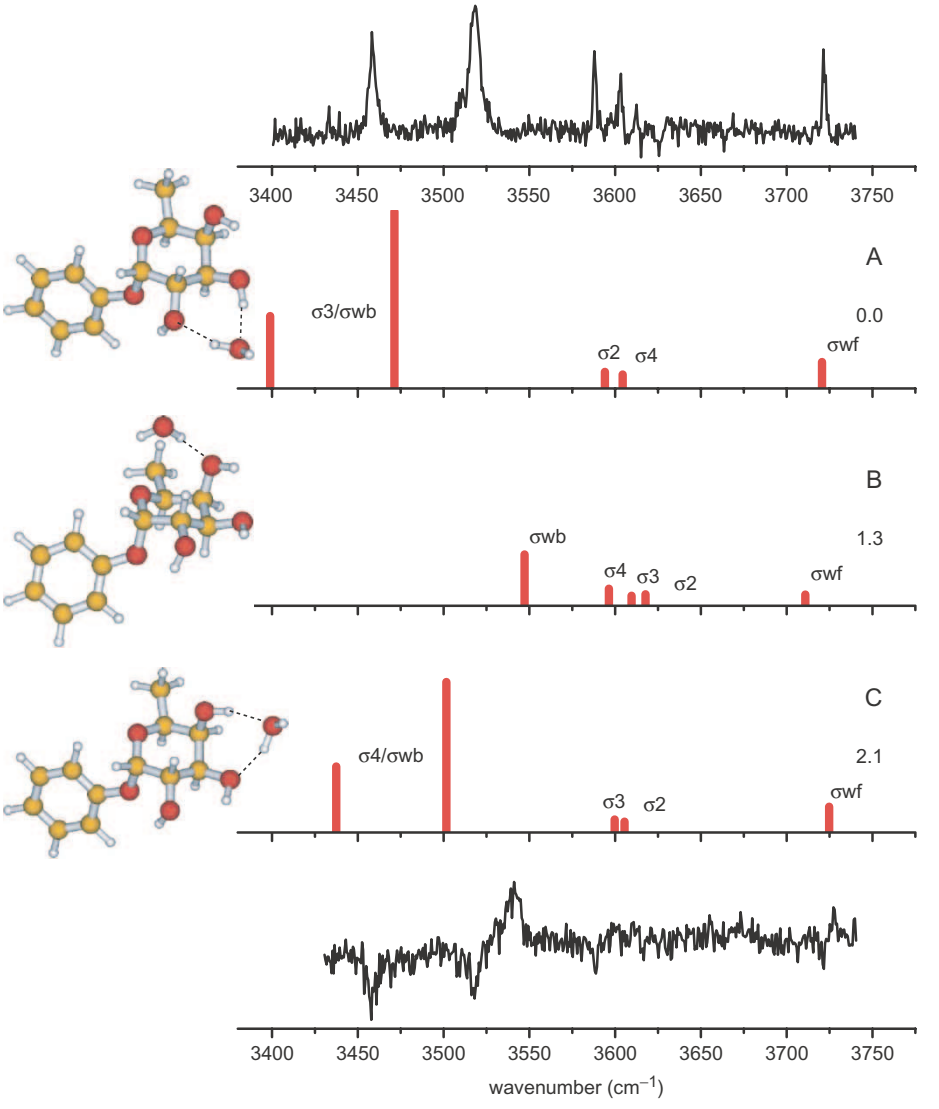


Figure 20. Experimental and computed IR ion-dip spectra of the singly hydrated complexes of phenyl α-L-fucopyranoside [28], monitored in the αphFuc$^+$ fragment ion channel: (the 'dips' in the lower spectrum are generated through predissociation of the more strongly populated, insertion complex, A). The relative energies of structures A and B are in kJ mol^{-1}.

doublet presented by an insertion complex and its IR spectrum fits that calculated for the second lowest-lying hydrate structure, cc_add 4 [28], see figure 20.

6.5. *Some concluding remarks*

The patterns of behaviour revealed through studies of the preferred mono-hydrated saccharide structures present a 'microcosm' of those that might be anticipated in

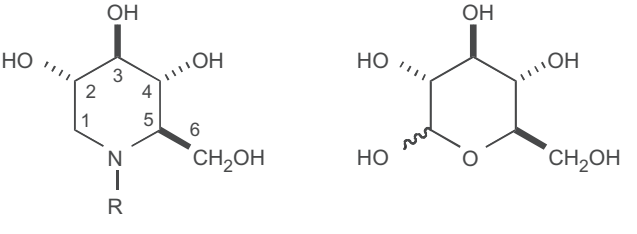
molecular recognition processes, when these are controlled primarily by H-bonded interactions. To be effective and reproducible the processes must involve a balance between conformational and structural *flexibility* and *specificity*. Multivalent recognition processes in particular require a degree of flexibility to ensure optimum contact – but not too much or 'any key will fit'. Specificity is needed to ensure selective recognition – 'the key must present the right recognition features'. These two qualities are reflected in the mono-hydrate structures selected by the five monosaccharides so far studied. Glucose and galactose present sufficient flexibility to relax from their preferred ccG + g– structures into the cG–g+ structure already favoured in mannose, to facilitate water binding at specific sites dictated by the rule, 'go for the weakest link'. The same rule is followed by xylose and fucose, where the need for flexibility is greatly reduced. Of course, more data are needed on a wider range of biomolecular hydrate structures before too much dogmatism sets in but the studies of water binding completed so far do suggest a means of identifying 'hot-spots' for carbohydrate recognition, e.g. the OH4–OH6 sites on mannose and the α-anomers of L-fucose and xylose, the OH2–O1 site on the β-anomer of xylose, and the OH6–O_{ring} site on galactose.

7. Using sugars: imino sugars and peptide mimics

7.1. *Sugar mimics: imino sugars*

Imino sugars, or polyhydroxylated alkaloids, are sugar mimics in which the ring oxygen is replaced by a nitrogen. They form an important class of natural and man-made drugs and have a wide range of proven and potential biological activities [53–56]. The deoxynojirimycin (DNJ) family, stereochemical mimics of glucose (scheme 7), are powerful inhibitors of α- and β-glucosidases [53–55]. As such, they can interfere with viral protein coat assembly and show promise as antiviral drugs, for example exhibiting anti-hepatitis B activity [56].

N-Alkylation of DNJ results in another powerful biological activity: inhibition of ceramide glucosyltransferases [57, 58]. Taking advantage of this activity, *N*-butyl DNJ is currently used as a treatment for Gaucher's disease, a lysosomal storage disorder. It works via inhibition of glycosphingolipid biosynthesis to counter a hereditary ceramide β-glucosidase deficiency. Unfortunately, the mechanism of inhibition is not well understood in either of these biological activities. It is thought that in the former,



[R = CH_2Ph]*N*-BnDNJ

[R = H] DNJ

glucose

Scheme 7. Molecular structures of deoxynojirimycin (DNJ) and glucose.

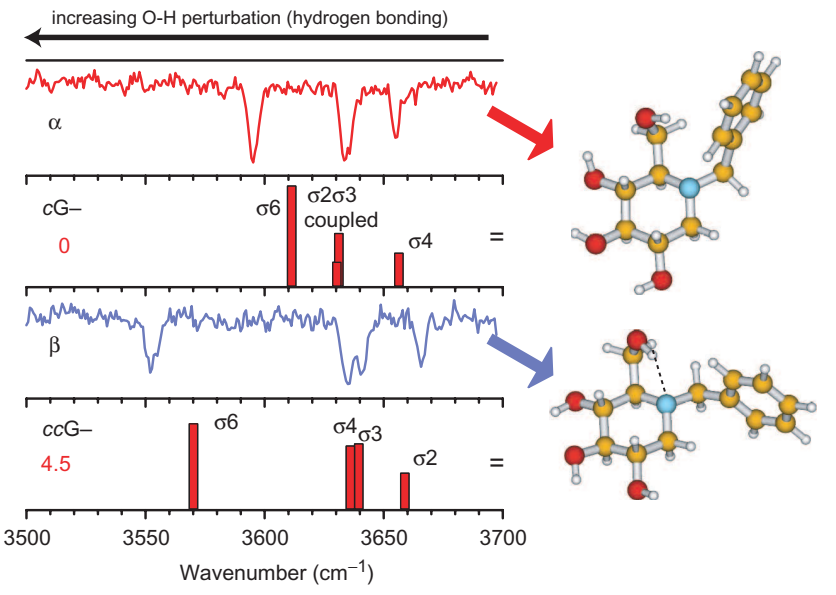


Figure 21. IR ion-dip spectra of the α and β conformers of *N*-BnDNJ shown with their best matches for conformational assignment [60]. Relative energies are listed in kJ mol^{-1}. All low-energy conformers have a G– orientation of the hydroxymethyl group.

anti-glycosidase activity, the DNJ is protonated while in the latter, anti-transferase activity, it is most likely not [59]. A better understanding of the structures should help elucidate mechanistic understanding on the atomic level and ultimately aid the design of better drugs.

The intrinsic conformational preferences of *N*-benzyl DNJ (*N*-BnDNJ), known to have anti-ceramide glucosyltransferase activity [59], are revealed in figure 21. Three distinct conformers, one major (α), one medium (β) and one minor (β'), are present in the gas phase and all three can be assigned to structures in which the exocyclic hydroxymethyl group is in the G–, *gauche* orientation, axial to the ring nitrogen [60].

The assignment of the major conformer, α, to the *c*G– conformer shown in figure 21 with OH6 hydrogen bonding into the π cloud of the chromophore ring is straightforward. Assignment of the minor conformers, β and β', is less clear; their R2PI (not shown here) and IR spectra both indicate conformers with very similar structures and the assignment shown for β in figure 21 represents the most favoured [60]. In all of the possible alternative assignments for the β/β'pair the G– orientation of the hydroxymethyl group is retained, but now there is a hydrogen bond from OH6 to the ring nitrogen. This shifts the band σ6 to $\sim$3550 cm^{-1} which lies lower than any bare monosaccharide OH vibrational band observed to date, apart from the minor *c*G–g+ conformer of βphGal, where a particularly strong OH4 $\rightarrow$ OH3 interaction shifts the band σ4 to 3520 cm^{-1} (see figure 6). The ring nitrogen can be thought of as the driving force in conformational selection in the imino-sugar, orienting the relatively flexible hydroxymethyl group into a G– orientation to create the very favourable OH6–N interaction. (In the α conformer, the G– orientation is stabilized by an interaction with the chromophore ring.)

The conserved G– orientation in *N*-BnDNJ differs from the preferred orientation in simple glucosides (but not in their mono-hydrated complexes). In bare βphGlc, only ~25% of the observed population adopts a G– orientation; the most populated conformation (~70%) is G+ (see figure 4) [24]. In bare βphGal, the preference for G+ is even stronger (~90%) [27]. The innate preference for the G– orientation in the imino sugars may well contribute to their strong binding compared to the natural glucoside substrate. Unfortunately no crystal structures in which *N*-alkylated DNJ is present as a tightly bound inhibitor have been reported but studies using rigid analogues with fixed geometry at this position have suggested that a G– orientation may be an important structural requirement for potent α-glucosidase inhibition [56]. This conjecture is supported by the only two reported crystal structures of biological complexes containing (non-alkylated) DNJ as a tightly bound inhibitor (one to an α- and one to a β-glucosidase): both show the hydroxymethyl group of DNJ with a G– orientation [61, 62].

A strategy has also been developed to examine the structure of DNJ itself, but in its *protonated* form thought to be necessary for anti-glycosidase (and hence antiviral) activity [59]. In these experiments, neutral DNJ is complexed with phenol in the gas phase expansion [63, 64] (to provide a proton donor) and the complex is photo-ionized via two-photon excitation. The ionized complex undergoes an efficient internal proton transfer from the phenol to the DNJ to generate a protonated DNJ-phenoxy radical complex, $DNJ.H^+\ldots$ OPh: its IR spectrum can be probed by monitoring the appearance of the photo-fragment $DNJ \cdot H^+$ generated via tunable near-IR photon absorption. Provided the presence of the phenoxy radical does not significantly affect the conformation of the $DNJ \cdot H^+$ moiety, the IR spectrum will reflect the conformation of protonated DNJ.

Figure 22 shows the IR photodissociation spectrum of the protonated DNJ-phenoxy complex together with those of its calculated lowest-energy structures. The appearance of a group of bands between 3630 and 3690 cm^{-1} in the experimental spectrum and the absence of any bands between 3500 and 3630 cm^{-1} clearly suggest that all the OH groups are at best only participating in very weak hydrogen bonding, ruling out a *trans* orientation of the hydroxymethyl group and excluding possible OH bonding with the phenoxy radical; either of these situations would result in OH stretching frequencies shifted below ~3620 cm^{-1}. The measured spectrum is consistent with an assignment to the calculated lowest-energy conformer, which has a G+ orientation of the hydroxymethyl group [64].

7.2. *Mimicking peptide secondary structure: carbopeptoids*

A tremendous diversity of protein function springs from a limited number of secondary structural elements including α-helices, β-sheets and turns. To be able to mimic this diversity using Nature-like, yet unnatural, building blocks is a fascinating challenge with potentially important benefits for drug design [65, 66]. Carbopeptoids, oligomeric chains made from alternating sugar and amino acid moieties have shown promise as synthetically accessible, flexible scaffolds for use in peptidomimetics [67–69]. As such, knowledge of and control over their conformational preference is of key importance.

Scheme 8 shows the monomer and dimer structures of a typical carbopeptoid building block based upon an alternating (2,5)*cis*-tetrahydrofuran unit. NMR-NOE studies of its tetramer based on the protected monomer unit **2**, have suggested the

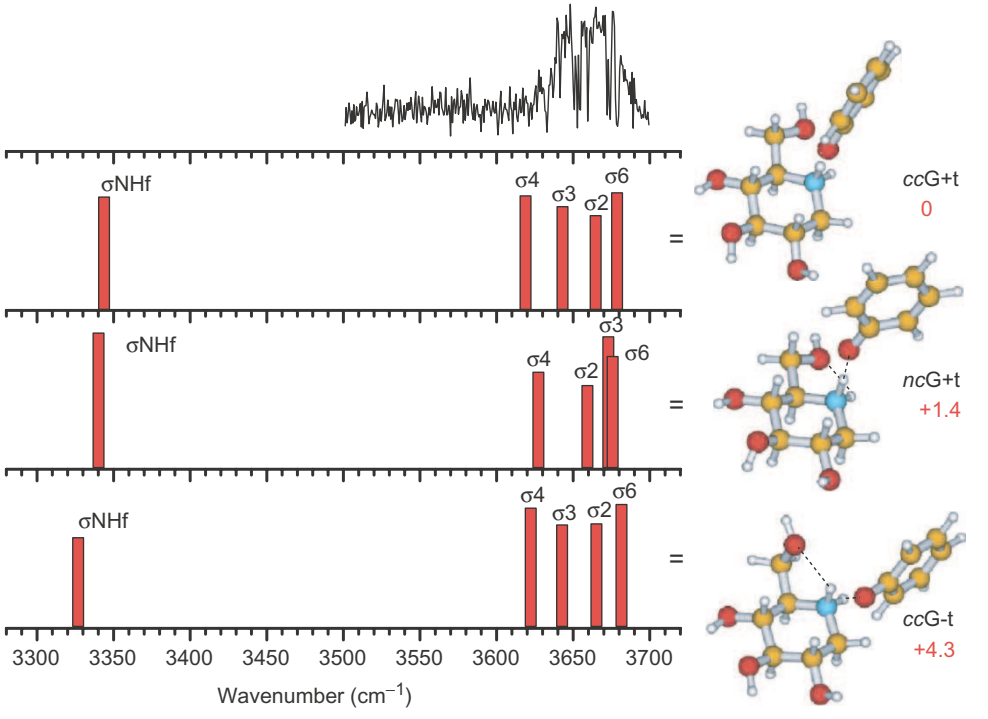
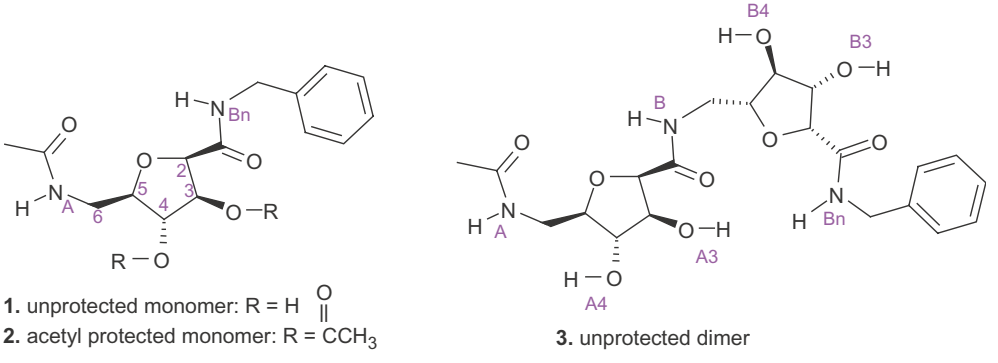
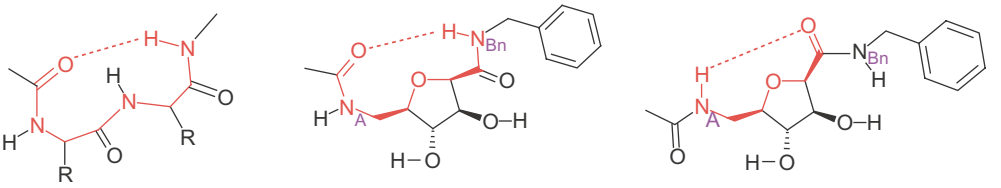


Figure 22. Measured IR photodissociation spectrum of the DN-J $H^+ \ldots$ OPh complex together with its computed low-energy conformers and spectra [64]. Relative energies are in kJ mol^{-1}.



Scheme 8. Carbopeptoid molecular structures.

presence in a chloroform solution of a repeated β-turn type structure, which coils the molecule into a helix [68, 70, 71]. Subsequent molecular dynamics simulations, however, have indicated the population of a (restricted) range of tetramer conformations (at 300 K), including the β-turn helical structure, which is populated for only ~15% of the time in a chloroform solution and is solvent dependent [72].



Scheme 9. Comparision of β-turn structures in peptides and in carbo-peptoid mimics.

Ab initio calculations on both the unprotected (**1**) and the protected (**2**) monomer unit have also identified several low-energy conformations, each stabilized by β-type 10-membered (C_{10}) turns (see scheme 9). The minimum energy conformations (at 0 K) of both (**1**) and (**2**) lie significantly lower in energy (> 13 kJ mol^{-1}) than the most stable non-β-turn forming conformers. Their near-infrared ion-dip spectra recorded in a free jet rare gas expansion confirm the intrinsic conformational preference for a β-type turn structure [73], see figure 23.

The calculations also provide an interesting 'spin-off'. In the lowest-energy conformers of **1** and **2**, the benzyl 'tag', used as a UV chromophore, folds back over the rest of the carbopeptoid molecule providing additional stabilization; thus, it cannot be considered to be structurally benign in this system. Calculations on derivatives of **1** in which the benzyl group is replaced by a methyl group show that the ordering of conformer stability is predominantly unchanged, and the methylated global minimum is the analogue of the benzylated one, i.e. both contain the β-type turn. However, there is another low-energy conformer which contains an alternative hydrogen bond, from N_A to the carbonyl on the other side of the ring (thus forming an eight-membered (C_8) β-turn rather than the C_{10} turn, scheme 9). It lies at a relative energy of 5 kJ mol^{-1} in the methylated monomers (compared with 15 kJ mol^{-1} in the benzylated ones). This difference in relative stability ordering can be easily understood through examination of the computed conformations: in the benzyl 'tagged' C_8 structure the chromophore ring points away from the rest of the molecule, in contrast to the folded C_{10} global minimum conformation, and it does not provide additional stabilization.

The results for the carbopeptoid units contrast with those for βBnLac and the monosaccharides, where the benzyl or phenyl tags do not interact significantly with the sugar units in the low-energy conformations and the chromophore has little influence on the conformational landscape – confirmed by parallel computational studies of the methylated analogues. Thus, while it is clear that the presence of spectroscopic tags *may* influence conformational selection, these effects can be guarded against by examination of the low-energy conformations. In *N*-BnDNJ, where there *is* a stabilizing interaction between the hydroxymethyl of the sugar ring and the chromophore in some of the observed conformers, the benzylated imino sugar is, in fact, bioactive and therefore relevant to study.

Ab initio calculations on the unprotected carbopeptoid dimer **3** (scheme 8), which now includes a second sugar and a third peptide unit, identify low-energy conformers exhibiting alternative hydrogen bonding arrangements that are somewhat more stable than double β-turn conformations although molecular mechanics calculations on

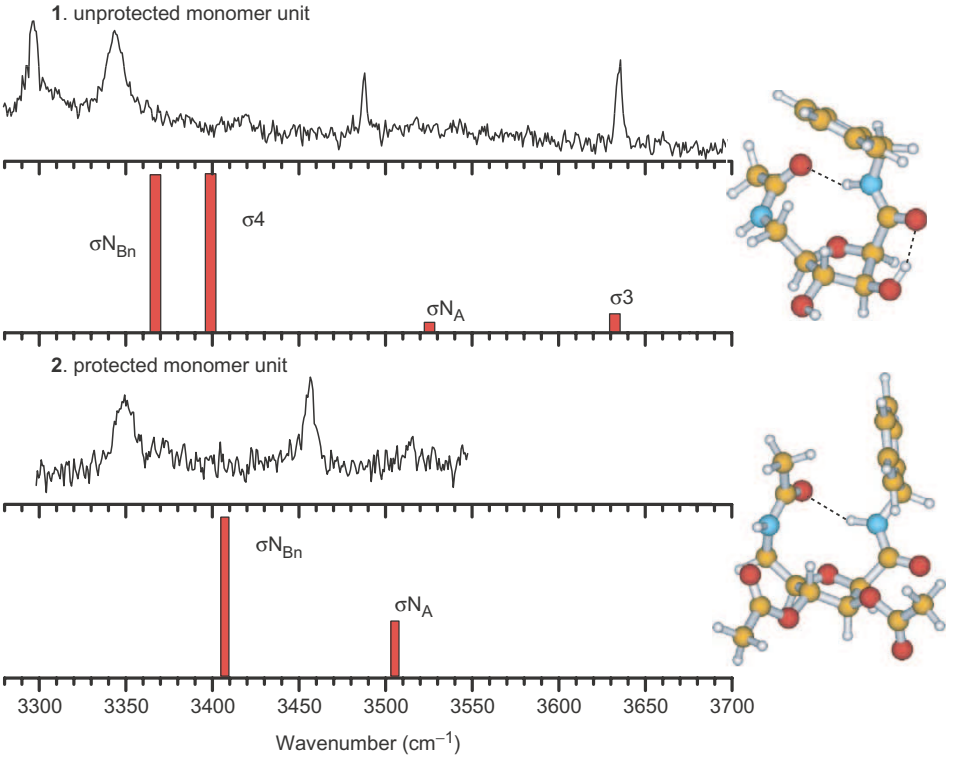


Figure 23. Experimental and computed near-IR spectra of carbopeptoid monomer units **1** and **2** together with their proposed conformational assignments [73]; both exhibit a hydrogen bond between N_{Bn} and a carbonyl oxygen, to create C_{10}, β-turn type structures.

extended (16 unit) chains suggest the repeating β-turn type structure does become the most stable in longer chains, allowing the molecule to coil into a helical structure [68].

Overall, these results indicate that while the *cis*-tetrahydrofuran based carbopeptoid unit is in some ways a relatively robust peptidomimetic building block, exhibiting a β-turn structure when isolated in the gas phase, its precise conformation is quite flexible and depends to a great extent on its environment, including *intra*molecular interactions. The abundance of hydrogen-bond donating and accepting groups in these molecules leads to a rich conformational landscape. In many ways this reflects the natural situation: the functional folds of proteins are often only marginally stable and short amino acid sequences adopt a variety of conformations in peptides and proteins, though within this there are known conformational preferences.

8. Challenges and opportunities

Raymond Dwek, in a seminal review article written a decade ago [4], identified two key issues in glycobiology: (1) 'How are carbohydrates involved in recognition events' and (2) 'How do they influence the properties of the proteins to which they are attached?'

An essential starting point in addressing the first issue includes characterization of the three-dimensional carbohydrate structures and of their dynamical flexibility. Characterizing the role played by their interaction with the ubiquitous solvent, water provides another critical step. Principal strategies that have been employed so far in pursuit of these aims have included X-ray crystallography (where feasible) in the solid state, a combination of NMR techniques and molecular dynamics calculations in solution, and mass spectrometry in the gas phase.

The current review has introduced a new strategy, again in the gas phase which like NMR and molecular dynamics in solution, also combines spectroscopy and computation but this time, electronic and vibrational spectroscopy of carbohydrate molecules and their hydrated complexes, conducted under molecular beam conditions, and *ab initio* calculation. Its early success has revealed a uniquely powerful means of characterizing carbohydrate conformations and structures and, in particular, the hydrogen bonded networks which they support (or which support them) as well as the specificity of their interaction with bound water molecules. The information that is being generated is beginning to complement in its small way, the information provided by the more 'traditional' methods and the success of the new strategy encourages optimism about its future application to characterize larger oligosaccharide structures.

Foremost among these are the core $(GlcNAc)_2(Man)_3$ *N*-linked glycan unit; the carbohydrate units and 'recognition antennae' that grow out of it; and the β(1–) glycosidic link between GlcNAc and the asparagine residue to which *N*-linked glycans are attached or in *O*-linked glycans, the common α(1–) link between GalNAc and serine or threonine.[7] This will lay the foundation for an exploration of the structural consequences of glycosylation on the neighbouring peptide structure, by comparing the conserved, local oligopeptide sequence, Asn-X-Thr/Ser, with and without its attachment to a short glycan chain, see figure 3 – an approach which will combine the successes now gained in the vibrational spectroscopy of oligosaccharides and oligopeptides and provide an entry into Dwek's theme (2). The arrival of techniques for the infrared spectroscopic interrogation of protonated glycopeptides generated at sufficient number densities to allow their vibrational spectroscopic interrogation, whether through electrospray ionization coupled with ion trapping [1] or the 'chemical proton transfer' strategy outlined in section 7.1, offers a direct connection to the gas-phase mass spectrometry of carbohydrates and glycopeptides and to the fragmentation of their molecular ions [74].

In 'real life' of course the structures are not static, but fortunately, improvements in dynamical predictions are also beginning to rise above the horizon. The availability of carbohydrate vibrational spectra in the mid-infrared fingerprint and far-infrared terahertz regions will allow better calibration of the current library of molecular force fields and should provide a stimulus to new developments in their accurate theoretical description, particularly of anharmonicity, where Gerber [75] and Clary [76] have already made very significant advances, and the coupled, strongly anharmonic low-frequency modes associated with conformational interchange, where the field is currently wide open.

[7]Perhaps it may be possible to understand why *N*-glycosylation always involves a β-link to Asn-X-Ser/Thr while *O*-glycosylation involves an α-link to Ser/Thr.

In 'real life' the carbohydrates also exist in a largely aqueous environment. The selectivity and structural consequences of water binding have so far only been investigated experimentally (in the gas phase) for singly hydrated complexes of monosaccharides. There is no physical constraint, however, on extending investigations to small oligosaccharide units to probe the influence of water binding on glyosidic linkages, or to multiply hydrated carbohydrates in general – it may not be the degree of hydration but rather the detailed explicit carbohydrate-water interactions (probed in gas phase experiments) that determine their structure and function [77]). Parallels between preferred water binding sites identified in the gas phase and the selectivity of protein-carbohydrate binding sites identified through X-ray crystallographic studies have already been demonstrated [25]. Connections between the gas phase and the condensed phase aqueous environment may become possible in the future through comparisons of Raman optical activity spectra recorded in aqueous and aprotic media [11, 12, 78, 79], and their 'gas phase' spectra, computed *ab initio*.

Carbohydrate-protein binding and many glycoprotein molecular recognition processes involve weak, multivalent interactions. The ease with which weakly bound intermolecular carbohydrate complexes can be prepared and interrogated in the gas phase under the rapid cooling conditions of a free jet expansion provides an obvious entry into their direct spectroscopic investigation. It will also be possible to explore the possible role of water-mediated binding through investigation of termolecular, carbohydrate-water-receptor assemblies since complex formation in the free jet need not be limited solely to bimolecular interactions. Furthermore, the interactions need not involve H-bonded interactions at all. Like a coin, or an argument, sugar molecules can present two sides, one 'decorated' with hydroxyl groups, favouring interaction with polar molecules; the other presenting a non-polar face that favours dispersive interactions with aromatic peptide residues, particularly phenylalanine. It should be possible to extend into the gas phase, for example, recent X-ray crystallographic investigations of the interaction between glucosides and artificial protein receptors [80] or the *ab initio* and NMR investigations of the interactions between fucose and benzene, and between methyl β-galactoside and phenol [81].

Finally, while the review has been focused very largely on the structures of small carbohydrates and the foundations for the future investigation of glycopeptide structures, it also provides a basis for future investigations of glycolipids in the gas phase and, while not a word has been said about the spectroscopic investigation of nucleic acid fragments, interested readers are strongly encouraged to take note of de Vries' brilliant pioneering investigations of guanosine [82].

Acknowledgements

The authors owe a great debt to Dr Ben Davis, Dr Antony Fairbanks and Dr David Gamblin, Professor George Fleet ('With these systems, whatever relates to shape is good to learn!') and Professor Naoki Asano, Professor Raymond Dwek and Dr Mark Wormald, for their inspiration, tutorial guidance, collaboration and, not least, the provision of precious, custom-synthesized carbohydrates. We also appreciate very much the many contributions that have been made by Professor Elaine Marzluff (Grinell College) during her sabbatical leave in Oxford, Dr Francis Talbot

(CEA, Saclay) who laid many of the early foundations, Dr Bo Liu who is now building on them, erstwhile undergraduate students Mr Alex Painter and Mr Theo Patsias, and our collaborators at FELIX in the FOM Institute at Rijnhuisen in The Netherlands for their generosity and help in running the mid-infrared experiments. Lastly, but certainly not least, we thank Professor J.-P. Schermann who some years ago asked us 'If we knew what fucose was?' None of the work we have described would have been possible without the generous support provided by the Leverhulme Trust (Grant No. F/08788D), the Royal Society for a USA Research Fellowship (RAJ) and a University Research Fellowship (LCS), the EPSRC, the CLRC Laser Support Facility, and the Physical and Theoretical Chemistry Laboratory.

References

[1] J. Oomens, N. Polfer, D. T. Moore, L. van der Meer, A. G. Marshall, *et al*., Phys. Chem. Chem. Phys. **7**, 1345 (2005).
[2] R. A. Jockusch, R. T. Kroemer, F. O. Talbot, L. C. Snoek, P. Çarçabal, *et al*., J. Amer. Chem. Soc. **126**, 5709 (2004).
[3] M. Martin-Pastor, A. Canales, F. Corzana, J. L. Asensio, and J. Jiménez-Barbero, J. Amer. Chem. Soc. **127**, 3589 (2005).
[4] R. A. Dwek, Chem. Rev. **96**, 683 (1996).
[5] R. Apweiler, H. Hermjakob, and N. Sharon, Biochim. Biophys. Acta **4**, 1473 (1999).
[6] M. E. Taylor and K. Drickamer, *Introduction to Glycobiology* (Oxford University Press, Oxford, 2003), p. 15.
[7] M. R. Wormald, A. J. Petrescu, Y. L. Pao, A. Glithero, T. Elliott, and R. A. Dwek, Chem. Rev. **102**, 371 (2002).
[8] A. Imberty and S. Perez, Chem. Rev. **100**, 4567 (2000).
[9] J. L. Asensio and J. Jiménez-Barbero, Biopolymers **35**, 55 (1995).
[10] H. J. Gabius, H.-C. Siebert, S. André, J. Jiménez-Barbero, and H. Rüdiger, Chem. Bio. Chem. **5**, 740 (2004).
[11] L. D. Barron, L. Hecht, I. H. McColl, and E. W. Blanch, Mol. Phys. **102**, 731 (2004).
[12] F. J. Zhu, N. W. Isaacs, L. Hecht, and L. D. Barron, J. Amer. Chem. Soc. **127**, 6142 (2005).
[13] L. C. Snoek, P. Çarçabal, R. T. Kroemer, J. P. Simons, J. M. Bakker, *et al*., Phys. Chem. Chem. Phys. **6**, 4546 (2004).
[14] S. B. Englesen, J. Koca, I. Braccini, C. H. Dupenhoat, and S. Perez, Carbohydr. Res. **276**, 1 (1995).
[15] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Chem. Phys. **122**, 084119 (2005).
[16] L. Hemmingsen, D. E. Madsen, A. L. Esbensen, L. Olsen, and S. B. Engelsen, Carbohydr. Res. **339**, 937 (2004).
[17] K. J. Naidoo and J. Y. J. Chen, Mol. Phys. **101**, 2687 (2003).
[18] K. J. Naidoo and J. W. Brady, J. Amer. Chem. Soc. **121**, 2244 (1999).
[19] E. G. Robertson and J. P. Simons, Phys. Chem. Chem. Phys. **3**, 1 (2001).
[20] T. S. Zwier, J. Phys. Chem. A **105**, 8827 (2001).
[21] J. P. Simons, C. R. Chimie **6**, 17 (2003).
[22] R. P. Weinkauf, J.-P. Schermann, M. S. de Vries, and K. Kleinermanns, Eur. J. Phys. D **20**, 309 (2002).
[23] J. P. Simons (Ed.) *et al*., *Bioactive Molecules in the Gas Phase*, Phys. Chem. Chem. Phys. **6**, 2543–2885 (2004).
[24] F. O. Talbot and J. P. Simons, Phys. Chem. Chem. Phys. **4**, 3562 (2002).
[25] P. Çarçabal, R. A. Jockusch, I. Hünig, L. C. Snoek, R. T. Kroemer, *et al*., J. Amer. Chem. Soc. **127**, 11414 (2005).
[26] See for example, G. I. Czonka, I. Kolossváry, P. Császár, K. E. Éliás, and I. G. Czismadia, J. Mol. Struct. (THEOCHEM) **395–396**, 29 (1997).
[27] R. A. Jockusch, F. O. Talbot, and J. P. Simons, Phys. Chem. Chem. Phys. **5**, 1502 (2003).
[28] P. Çarçabal, T. Patsias, I. Hünig, L. C. Snoek, and J. P. Simons, Phys. Chem. Chem. Phys. **8** (2006): DOI: 10.1039/b514301b.
[29] D. J. Becker and J. B. Lowe, Glycobiology **13**, 41 (2003).
[30] I. Hünig, A. J. Painter, R. A. Jockusch, P. Çarçabal, E. M. Marzluff, *et al*., Phys. Chem. Chem. Phys. **7**, 2474 (2005).
[31] M. L. Hayes, A. S. Serianni, and R. Barker, Carbohydrate Res. **100**, 87 (1982).

[32] N. W. H. Cheetham, P. Dasgupta, and G. E. Ball, Carohydrate Res. **338**, 955 (2003).
[33] E. W. Sayers and J. H. Prestegard, Biophys. J. **82**, 2683 (2002).
[34] N. S. Greenspan, Curr. Top. Microbiol. Immunol. **260**, 65 (2001).
[35] M. L. de la Paz, G. Ellis, M. Perez, J. Perkins, J. Jiménez-Barbero, and C. Vicent, Eur. J. Org. Chem. **5**, 840 (2002).
[36] M. L. de la Paz, C. Gonzalez, and C. Vicent, Chem. Comm. **5**, 411 (2000).
[37] M. L. de la Paz, J. Jimenez-Barbero, and C. Vicent, Chem. Comm. **4**, 465 (1998).
[38] C. A. Hunter, Angew. Chem. Int. Ed. **43**, 5310 (2004).
[39] F. A. Quiocho, D. K. Wilson, and N. K. Vyas, Nature **404**, 340 (1989).
[40] D. B. Smithrud, E. M. Sanford, I. Chao, S. B. Ferguson, D. R. Carcanague, *et al*., Pure Appl. Chem. **62**, 2227 (1990).
[41] T. N. Bhat, G. A. Bentley, G. Boulot, M. I. Greene, D. Tello, *et al*., Proc. Natl. Acad. Sci. USA **91**, 1089 (1994).
[42] C. R. Robinson and S. G. Sligar, Protein Sci. **5**, 2119 (1996).
[43] R. Ravishankar, M. Ravindran, K. Suguna, A. Surolia, and M. Vijayan, Curr. Sci. **72**, 855 (1997).
[44] S. Elgavish and B. Shaanan, J. Mol. Biol. **277**, 917 (1998).
[45] R. Ravishankar, M. Ravindran, K. Suguna, K. A. Surolia, and M. Vijayan, Curr. Sci. **76**, 1393 (1999).
[46] C. Clarke, R. J. Woods, J. Gluska, A. Cooper, M. A. Nutley, and G. J. Boons, J. Amer. Chem. Soc. **123**, 12238 (2001).
[47] C. A. Bottoms, P. E. Smith, and J. J. Tanner, Protein Sci. **11**, 2125 (2002).
[48] A. Ben-Naim, Biophys. Chem. **101**, 309 (2002).
[49] K. Turton, R. Natesh, N. Thiyagarajan, J. A. Chaddock, and K. R. Acharya, Glycobiology **14**, 923 (2004).
[50] R. A. Jockusch, F. O. Talbot, R. T. Kroemer, and J. P. Simons, J. Phys. Chem. A **107**, 10725 (2003).
[51] R. Schmidt, M. Karplus, and J. W. Brady, J. Amer. Chem. Soc. **118**, 541 (1996).
[52] C. Höög and G. Widmalm, J. Phys. Chem. B **105**, 6375 (2001).
[53] N. Asano, R. J. Nash, R. J. Moilyneux, and G. W. J. Fleet, Tetrahedron: Asymmetry **11**, 1645 (2000).
[54] N. Asano, Glycobiology **13**, 93R (2003).
[55] A. Vasella, G. J. Davies, and M. Böhm, Curr. Opinion Chem. Biol. **6**, 619 (2002).
[56] N. Asano, M. Nishida, A. Kato, H. Kizu, M. Matsui, *et al*., J. Med. Chem. **41**, 2565 (1998).
[57] F. M. Platt, G. R. Neises, G. Reinkensmeier, M. J. Townsend, V. H. Perry, *et al*., Science **276**, 428 (1997).
[58] F. M. Platt, G. R. Neises, R. A. Dwek, and T. D. Butters, J. Biol. Chem. **269**, 8362 (1994).
[59] T. D. Butters, L. A. G. M. van den Broek, G. W. J. Fleet, T. M. Krulle, M. R. Wormald, *et al*., Tetrahedron: Asymmetry **11**, 113 (2000).
[60] R. A. Jockusch, F. O. Talbot, N. Asano, G. W. Fleet, and J. P. Simons, Phys. Chem. Chem. Phys. **6**, 5283 (2004).
[61] A. Hempel, N. Camerman, D. Mastrpaolo, and A. Camerman, J. Med. Chem. **36**, 4082 (l993).
[62] N. Asano, H. Kizu, K. Oseki, E. Tomioka, K. Matsui, *et al*., J. Med. Chem. **38**, 2349 (1995).
[63] N. A. Macleod and J. P. Simons, Phys. Chem. Chem. Phys. **6**, 2821 (2004).
[64] E. M. Marzluff, I. Hünig, N. A. Macleod, and J. P. Simons, Phys. Chem. Chem. Phys., to be submitted.
[65] S. H. Gellman, Acc. Chem. Res. **31**, 173 (1998).
[66] D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes, and J. S. Moore, Chem. Rev. **101**, 3893 (2001).
[67] K. C. Nicolau, H. Florke, M. G. Egan, T. A. Barth, and V. A. Estevez, Tetrahedron Lett. **36**, 1775 (1995).
[68] T. D. W. Claridge, D. D. Long, C. M. Baker, B. Odell, G. H. Grant, *et al*., J. Org. Chem. **70**, 2082 (2005).
[69] T. K. Chakraborty, P. Srinivasu, S. Tapadar, and B. K. Mohan, J. Chem. Sci. **116**, 187 (2004).
[70] M. D. Smith, T. D. W. Claridge, G. E. Tranter, M. S. P. Sansom, and G. W. J. Fleet, Chem. Comm. 2041 (1998).
[71] M. D. Smith, T. D. W. Claridge, M. S. P. Sansom, and G. W. J. Fleet, Org. Biomol. Chem. **1**, 3647 (2003).
[72] R. Baron, D. Bakowies, and W. F. van Gunsteren, Ang. Chem Int. Ed. **43**, 4055 (2004).
[73] R. A. Jockusch, F. O. Talbot, G. W. J. Fleet, and J. P. Simons, to be submitted.
[74] A. Dell and H. R. Morris, Science **291**, 2351 (2001).
[75] R. B. Gerber, G. M. Chaban, S. K. Gregurick, and B. Brauer, Biopolymers **68**, 370 (2003).
[76] D. C. Clary and T. F. Miller, Phys. Chem. Chem. Phys **6**, 2563 (2004).
[77] S. B. Englesen, C. Monteiro, C. H. de Penhoat, and S. Perez, Biophys. Chem. **93**, 103 (2001).
[78] A. F. Bell, L. Hecht, and L. D. Barron, J. Amer. Chem. Soc. **116**, 5155 (1994).
[79] N. A. Macleod, P. Butz, J. P. Simons, G. H. Grant, C. M. Baker, *et al*., Phys. Chem. Chem. Phys. **7**, 1432 (2005).
[80] M. Mazik, H. Cavga, and P. G. Jones, J. Amer. Chem. Soc. **127**, 9045 (2005).
[81] M. C. Fernández-Alonso, F. J. Cañada, J. Jiménez-Barbero, and G. Cuavas, J. Amer. Chem. Soc. **127**, 7379 (2005).
[82] E. Nir, I. Hünig, K. Kleinermanns, and M. S. de Vries, Chem. Phys. Chem. **5**, 131 (2004).