Hydrogen Bonding and Cooperativity in Isolated and Hydrated Sugars: Mannose, Galactose, Glucose, and Lactose

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · SEPTEMBER 2005

Impact Factor: 12.11 · DOI: 10.1021/ja0518575 · Source: PubMed

J|A|C|S
ARTICLES
Published on Web 07/26/2005

Hydrogen Bonding and Cooperativity in Isolated and Hydrated Sugars: Mannose, Galactose, Glucose, and Lactose

Pierre Çarçabal,[†] Rebecca A. Jockusch,[†] Isabel Hünig,[†] Lavina C. Snoek,[†] Romano T. Kroemer,[‡] Benjamin G. Davis,[§] David P. Gamblin,[§] Isabelle Compagnon,[||] Jos Oomens,[||] and John P. Simons*,[†]

Contribution from the Chemistry Department, Physical and Theoretical Chemistry Laboratory, Oxford University, South Parks Road, Oxford OX1 3QZ United Kingdom, Sanofi-Aventis, Centre de Recherche de Paris, 13 quai Jules Guesde, BP14, 94403 Vitry-sur-Seine, France, Chemistry Department, Chemistry Research Laboratory, Oxford University, 12 Mansfield Road, Oxford OX1 3TA United Kingdom, and FOM Institute for Plasmaphysics Rijnhuizen, Edisonbaan 14, NL-3439 MN, Nieuwegein, The Netherlands

Received March 23, 2005; E-mail: john.simons@chem.ox.ac.uk

Abstract: The conformation of phenyl-substituted monosaccharides (mannose, galactose, and glucose) and their singly hydrated complexes has been investigated in the gas phase by means of a combination of mass selected, conformer specific ultraviolet and infrared double resonance hole burning spectroscopy experiments, and ab initio quantum chemistry calculations. In each case, the water molecule inserts into the carbohydrate at a position where it can replace a weak intramolecular interaction by two stronger intermolecular hydrogen bonds. The insertion can produce significant changes in the conformational preferences of the carbohydrates, and there is a clear preference for structures where cooperative effects enhance the stability of the monosaccharide conformers to which the water molecule chooses to bind. The conclusions drawn from the study of monosaccharide−water complexes are extended to the disaccharide lactose and discussed in the light of the underlying mechanisms that may be involved in the binding of carbohydrate assemblies to proteins and the involvement, or not, of key structural water molecules.

Introduction

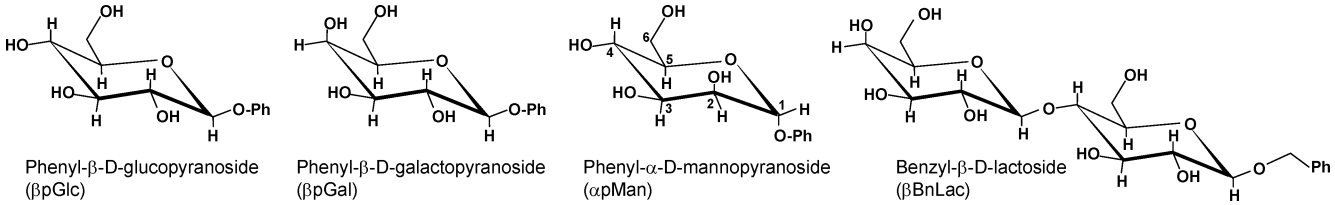
An extensively studied, but still far from being fully understood, function of carbohydrates is their ability to mediate biological processes, such as molecular recognition within living systems, for example, through the glycan foliage provided by N-linked glycopeptides.[1] Hydrogen bonding is one of the key components in the machinery of carbohydrate molecular recognition[2−4] and, more generally, in the maintenance of their preferred conformational structures. Intramolecular hydrogen bonds within the linked monosaccharide units and between them, and intermolecular hydrogen bonds between carbohydrates and neighboring molecules, for example, water or noncovalently bound biomolecules, can compete or work together in a cooperative way to stabilize favored conformations. This "matrix" of neighboring hydroxy groups can also create cooperative hydrogen-bonded networks,[5−8] which can effectively stabilize and strongly rigidify some conformations and also play a major role in molecular recognition.[3,9−12] Many studies have identified water not simply as a spectator solvent but as an active participant in carbohydrate molecular recognition processes.[2,13−23] The influence of this ubiquitous biological

[†] Chemistry Department, Physical and Theoretical Chemistry Laboratory, Oxford University, South Parks Road, Oxford OX1 3QZ United Kingdom.
[‡] Sanofi-Aventis, Centre de Recherche de Paris, 13 quai Jules Guesde, BP14, 94403 Vitry-sur-Seine, France.
[§] Chemistry Department, Chemistry Research Laboratory, Oxford University, 12 Mansfield Road, Oxford OX1 3TA United Kingdom.
[||] FOM Institute for Plasmaphysics Rijnhuizen, Edisonbaan 14, NL-3439 MN, Nieuwegein, The Netherlands.

(1) Helenius, A.; Aebi, M. *Annu. Rev. Biochem.* **2004**, *73*, 1019−1049.
(2) Hunter, C. A. *Angew. Chem., Int. Ed.* **2004**, *43*, 5310−5324.
(3) Greenspan, N. S. *Curr. Top. Microbiol. Immunol.* **2001**, *260*, 65−85.
(4) Jeffrey, G. A. *Food Chem.* **1996**, *56*, 241−246.
(5) Parra, R. D.; Bulusu, S.; Zeng, X. C. *J. Chem. Phys.* **2003**, *118*, 3499−3509.
(6) Missopolinou, D.; Panayiotou, C. *J. Phys. Chem. A* **1998**, *102*, 3574−3581.
(7) Masella, M.; Flament, J. P. *Mol. Simul.* **2000**, *24*, 131−156.
(8) Gung, B. W.; Zhu, Z. H.; Everingham, B. *J. Org. Chem.* **1997**, *62*, 3436−3437.
(9) de la Paz, M. L.; Ellis, G.; Perez, M.; Perkins, J.; Jimenez-Barbero, J.; Vicent, C. *Eur. J. Org. Chem.* **2002**, 840−855.
(10) de la Paz, M. L.; Gonzalez, C.; Vicent, C. *Chem. Commun.* **2000**, 411−412.
(11) de la Paz, M. L.; Jimenez-Barbero, J.; Vicent, C. *Chem. Commun.* **1998**, 465−466.
(12) Di Cera, E. *Chem. Rev.* **1998**, *98*, 1563−1591.
(13) Quiocho, F. A.; Wilson, D. K.; Vyas, N. K. *Nature* **1989**, *340*, 404−407.
(14) Bhat, T. N.; Bentley, G. A.; Boulot, G.; Greene, M. I.; Tello, D.; Dallacqua, W.; Souchon, H.; Schwarz, F. P.; Mariuzza, R. A.; Poljak, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1089−1093.
(15) Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G. J. *J. Am. Chem. Soc.* **2001**, *123*, 12238−12247.
(16) Ben-Naim, A. *Biophys. Chem.* **2002**, *101*, 309−319.
(17) Bottoms, C. A.; Smith, P. E.; Tanner, J. J. *Protein Sci.* **2002**, *11*, 2125−2137.
(18) Robinson, C. R.; Sligar, S. G. *Protein Sci.* **1996**, *5*, 2119−2124.
(19) Smithrud, D. B.; Sanford, E. M.; Chao, I.; Ferguson, S. B.; Carcanague, D. R.; Evanseck, J. D.; Houk, K. N.; Diederich, F. *Pure Appl. Chem.* **1990**, *62*, 2227−2236.
(20) Elgavish, S.; Shaanan, B. *J. Mol. Biol.* **1998**, *277*, 917−932 (PDB ID = 1AXZ).
(21) Ravishankar, R.; Ravindran, M.; Suguna, K.; Surolia, A.; Vijayan, M. *Curr. Sci.* **1997**, *72*, 855−861 (PDB ID = 2tep).

10.1021/ja0518575 CCC: $30.25

Chart 1



solvent on carbohydrate conformational landscapes needs to be understood, if their interaction with biological molecules is to be properly modeled.

To date, the main approaches to determining the conformation of carbohydrates have been based on X-ray crystallography and NMR measurements, and on molecular modeling using empirical molecular mechanics force fields (FF).[24] These approaches have limitations, most of them related in some way to the property that makes carbohydrates so interesting and challenging to study, their conformational flexibility. To validate a force field, comparison with experimental data in model systems is necessary,[25−28] but comparisons with NMR data may be compromised by ensemble-averaging, and both NMR and crystallographic data may be influenced by environmental effects, associated with the solvent (NMR) or the crystal lattice.[27,29] Vibrational Raman spectroscopy,[30] circular dichroism[31] (VCD), and Raman optical activity[32] (ROA) measurements in aqueous solution and Fourier transform infrared (FTIR) measurements in nonpolar solvents[33] have provided precious insights into carbohydrate conformations in solution, but the role of the solvent on the vibrational frequencies of the solute has to be understood quantitatively before structural information can be extracted or modeling methods calibrated accurately. FF calculations can include environmental effects, but, again, experimental data are required to validate the parameters used to describe the properties of the solute, the solvent, and their mutual interaction.

One way of circumventing these problems exploits conformer specific, mass-selected infrared laser spectroscopy conducted at low temperatures in the gas phase, to probe the "vibrational signatures" of monosaccharides and small oligosaccharides initially free of the environment and then interacting with it in a controlled way within well-defined, finite-size hydrated molecular complexes, coupled with ab initio computational chemistry to translate these signatures into their molecular and conformational structures.[34] We have already used this strategy to investigate the conformation of the monosaccharide derivatives phenyl β-D-galactopyranoside[35] (βpGal) and phenyl β-D-glucopyranoside[36] (βpGlc), and the disaccharide formed by their combination, benzyl β-lactoside[37] (βBnLac). Two monohydrates of βpGlc have also been identified, and tentative conformational and structural assignments were suggested.[38] Experimental data obtained on finite size systems such as these, supported by quantum chemical theory, could provide a basis for testing and calibrating carbohydrate FF's, which in turn could be used with greater confidence to model larger molecular systems that cannot be treated with computationally time-consuming, quantum chemistry calculation. Following this aspiration and motivated by the particular importance of the monosaccharide mannose, a key building block in glyco-conjugate structures, we have conducted the first combined infrared spectroscopic and ab initio study of the conformational and structural landscapes of phenyl α-D-mannose (αpMan) and its monohydrate complex (αpMan-W). We have also investigated the hydrated complex of βpGal (βpGal-W), a complex that has eluded spectroscopic detection until now. Prompted by these results, a series of new ab initio calculations, spectroscopic measurements, and structural assignments of the monohydrates of βpGlc (βpGlc-W) have been undertaken as well as a preliminary computational study of the hydration of lactose. The structures of the (phenyl or benzyl tagged) sugars under investigation here are shown in Chart 1.

The results reveal how water comes into play to "shape" to its own advantage the conformation of sugars and allow a much clearer picture to be drawn of the relevant processes. In this picture, hydrogen bonding and cooperativity are the driving forces: the near-infrared measurements provide a "spectroscopic indicator" of their manifestation. Comparisons with carbohydrate−protein X-ray crystal structures suggest that in some cases proteins recognize the preferred hydrated carbohydrate conformational structures. The carbohydrate water binding modes identified here are either conserved, with structural water molecules present in the crystal at the same positions identified in this gas-phase study, or the water molecule is displaced in favor of interactions with the protein at the same sites on the carbohydrate.

(22) Ravishankar, R.; Ravindran, M.; Suguna, K.; Surolia, A.; Vijayan, M. *Curr. Sci.* **1999**, *76*, 1393−1393 (PDB ID = 2tep).
(23) Turton, K.; Natesh, R.; Thiyagarajan, N.; Chaddock, J. A.; Acharya, K. R. *Glycobiology* **2004**, *14*, 923−929 (PDB ID = 1UZY).
(24) Wormald, M. R.; Petrescu, A. J.; Pao, Y. L.; Glithero, A.; Elliott, T.; Dwek, R. A. *Chem. Rev.* **2002**, *102*, 371−386.
(25) Heramingsen, L.; Madsen, D. E.; Esbensen, A. L.; Olsen, L.; Engelsen, S. B. *Carbohydr. Res.* **2004**, *339*, 937−948.
(26) McNamara, J. P.; Muslim, A. M.; Abdel-Aal, H.; Wang, H.; Mohr, M.; Hillier, I. H.; Bryce, R. A. *Chem. Phys. Lett.* **2004**, *394*, 429−436.
(27) Perez, S.; Imberty, A.; Engelsen, S. B.; Gruza, J.; Mazeau, K.; Jimenez-Barbero, J.; Poveda, A.; Espinosa, J. F.; van Eyck, B. P.; Johnson, G.; French, A. D.; Louise, M.; Kouwijzer, C. E.; Grootenuis, P. D. J.; Bernardi, A.; Raimondi, L.; Senderowitz, H.; Durier, V.; Vergoten, G.; Rasmussen, K. *Carbohydr. Res.* **1998**, *314*, 141−155.
(28) Perez, S.; Kouwijzer, M.; Mazeau, K.; Engelsen, S. B. *J. Mol. Graphics* **1996**, *14*, 307.
(29) Woods, R. J. *Glycoconjugate J.* **1998**, *15*, 209−216.
(30) Arboleda, P. H.; Loppnow, G. R. *Anal. Chem.* **2000**, *72*, 2093−2098.
(31) Monde, K.; Taniguchi, T.; Miura, N.; Nishimura, S. I. *J. Am. Chem. Soc.* **2004**, *126*, 9496−9497.
(32) Wen, Z. Q.; Barron, L. D.; Hecht, L. *J. Am. Chem. Soc.* **1993**, *115*, 285−292.
(33) Hawley, J.; Bampos, N.; Aboitiz, N.; Jimenez-Barbero, J.; de la Paz, M. L.; Sanders, J. K. M.; Carmona, P.; Vicent, C. *Eur. J. Org. Chem.* **2002**, 1925−1936.
(34) Bio-active molecules in the gas phase. In *Phys. Chem. Chem. Phys.*; Simons, J. P., Bowers, M. T., Hobza, P., Weinkauf, R., Ashfold, M. N. R., Eds.; Royal Society of Chemistry: London, 2004; Vol. 6, pp 2543−2890.
(35) Jockusch, R. A.; Talbot, F. O.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1502−1507.
(36) Talbot, F. O.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3562−3565.
(37) Jockusch, R. A.; Kroemer, R. T.; Talbot, F. O.; Snoek, L. C.; Çarçabal, P.; Simons, J. P.; Havenith, M.; Bakker, J. M.; Compagnon, I.; Meijer, G.; von Helden, G. *J. Am. Chem. Soc.* **2004**, *126*, 5709−5714.
(38) Jockusch, R. A.; Kroemer, R. T.; Talbot, F. O.; Simons, J. P. *J. Phys. Chem. A* **2003**, *107*, 10725−10732.

Methods

The experimental and theoretical tools used in this work have been described elsewhere.[39] Here, we simply recall the basics and give the details specific to the systems examined.

Spectroscopic Tagging of Carbohydrates. Our experiments rely on the observation of resolved UV resonant two-photon ionization (R2PI) spectra of the molecules of interest. For that purpose, the carbohydrates have all been "spectroscopically tagged" to provide a near UV chromophore through a linkage to a phenyl group. Synthetic access to the newly surveyed αpMan system was gained through the use of a peracetylated mannosyl donor with phenol, followed by Zemplén deacylation in a highly efficient manner and with excellent stereocontrol for the α-anomer product.[40−42] The tagged β-saccharides, βpGal and βpGlc, were purchased from Sigma Chemical Co. and used without further purification.

Fortunately, the perturbation introduced by addition of the chromophore (at the anomeric position O1; see Chart 1) appears to be benign because calculations indicate that the conformations of the monosaccharides are virtually the same whether they are tagged or not, and, as importantly, the relative energy ordering of the most stable conformers does not appear to be affected. Calculations of the conformational landscape of (unsubstituted) glucose by ourselves and others[43] identified the same conformations for the most stable minima as were found earlier for βpGlc.[36] Additional calculations on α-D-mannose and β-D-galactose (without the phenyl ring) have also confirmed the negligible influence of the phenyl chromophore on their conformational structures and preferences. It should be noted, however, that tagging introduces some extra complexity, because a single conformation of the sugar moiety of the molecule may be associated with more than one orientation of the phenyl ring.

Gas-Phase Single and Double Resonance Spectroscopy. The monosaccharides were evaporated from an oven, located in a vacuum chamber at the exit of a pulsed supersonic expansion valve. The oven temperature, optimized for each system, was typically set between 110 and 150 °C for the observation of bare molecules and between 150 and 200 °C for the study of water complexes. The carrier gas, argon, was maintained at a backing pressure of ∼4 bar for the experiments on the bare molecules and ∼6 bar for the water complexes. Sugar molecules entrained in the expanding jet were cooled to rotational (vibrational) temperatures of ∼10 (50) K. Hydrated complexes were formed when the carrier gas was seeded with water vapor prior to the expansion (∼0.25% H_2O in Ar).

The molecules entered the detection chamber through a 1 mm diameter skimmer and were ionized via R2PI in the extraction region of a time-of-flight mass spectrometer. The ions were then accelerated toward a multichannel plate detector to provide mass-selected $S_1 \leftarrow S_0$ R2PI spectra of the (phenyl tagged) neutral molecules or their hydrated complexes. Individual conformers were separately identified through a double resonance UV−UV hole burning (UVHB) scheme.[39] A similar IR−UV hole burning double resonance scheme (IRHB) generated the corresponding electronic ground-state IR spectrum of each species.

Computational Strategy. In the previous study of the glucose monohydrate (βpGlc-W),[38] alternative structures were constructed by adding the water molecule systematically to low-energy, thermally populated conformers of the bare βpGlc molecule at favorable insertion geometries (where the water participates in two hydrogen bonds). This procedure left open the possibility of bias in the structural survey and was particularly problematic if the conformation of the observed hydrated sugar was not derived from a populated low energy conformation of the nonhydrated molecule. To ensure that the conformational landscapes of each sugar, both free and hydrated, had been adequately sampled, subsequent exploration of their potential energy surfaces was performed with conformation searching using the Monte Carlo Multiple Minimum (MCMM) method as implemented in the MacroModel suite of programs (version 8.5).[44] The MMFF94s molecular mechanics force field[45] was used for 10 000 steps. All structures generated with energies within 25 kJ mol^{-1} of the minimum were examined, and, from these, multiple conformers were selected for higher level, electronic structure theory calculations (based on relative energy and the desire to explore a diversity of conformers); 55, 39, 15, and 22 structures were selected for βpGlc-W, βpGal-W, αpMan, and αpMan-W, respectively. Some of these only differed from each other by the orientation of the phenyl chromophore and/or the orientation of the water OH group acting as an H-bond acceptor from the saccharide. Fewer conformers of αpMan and its hydrate were considered as these systems presented only a single stable chromophore orientation, making their potential energy surfaces somewhat simpler than those the β-monosaccharides, βpGlc and βpGal, which featured two positions of the phenyl of similar stability.

The higher-level electronic structure theory calculations were performed using the Gaussian03 software package,[46] following the well-established procedures described elsewhere.[39] Optimized structures, transition states, and harmonic frequencies were calculated using hybrid method density functional theory at the B3LYP/6-31+G* level of theory. The low-lying structures, whether "bare" or hydrated, invariably adopted a $^4C_1$ ring conformation. Relative total energies, ΔE_0, given in what follows are from single-point energy calculations performed on the B3LYP-optimized geometries at the MP2/6-311++G** level of theory and include B3LYP/6-31+G* zero-point energy corrections. The harmonic frequencies of all of the OH stretching modes were scaled by a factor of 0.9734, an empirical scaling factor, which has been used in previous studies of carbohydrate conformation.[35−37]

Assigning Conformers and Structures. The spectral, conformational, and structural assignments are based upon a combination of several factors, which collectively "home in" either on a unique assignment or on 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 R2PI spectra) 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.

In the case of the structural determination of the hydrated complexes, the prediction of the vibrational frequencies for the stretching modes of the OH groups involved as donors in intermolecular H bonds between the carbohydrate and the water molecule turns out to be less satisfactory than for the OH groups involved in intramolecular interactions within the carbohydrate or for the free OH of the water molecule. Several possible sources of error can be identified: the difficulties in properly describing intermolecular interactions, the use of a single scaling factor for all OH stretches, and the influence of basis set superposition error. Nevertheless, even though the absolute accuracy of the prediction may sometimes be found wanting, the vibrational patterns featured by these modes are generally well reproduced, and, together with the other criteria mentioned above, assignments can be made with a degree of confidence.

To name the various conformers and structures consistently and as simply as possible, we have adopted a convention slightly different from that followed in our previous publications.[35−38] The three ring

(39) Robertson, E. G.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1−18.
(40) Lee, Y. S.; Rho, E. S.; Min, Y. K.; Kim, B. T.; Kim, K. H. *J. Carbohydr. Chem.* **2001**, *20*, 503−506.
(41) Zemplen, G. *Ber.* **1920**, 60, 1555−1564.
(42) Montgomery, E. M.; Richtmyer, N. K.; Hudson, C. S. *J. Am. Chem. Soc.* **1942**, *64*, 690−694.
(43) Momany, F. A.; Appell, M. D. *Abstr. Pap. Am. Chem. Soc.* **2003**, *226*, 080-CARB.
(44) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440−467.
(45) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 730−748.
(46) Frisch, M. J.; et al. *Gaussian 03*, revision B.03; Gaussian, Inc.: Pittsburgh, PA, 2003.

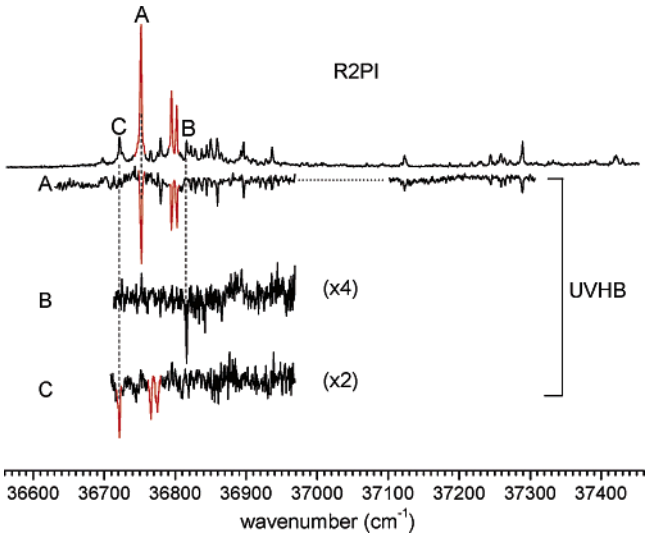


Figure 1. Resonant two-photon ionization (R2PI) and UV hole-burning (UVHB) spectra of αpMan isolated in the free jet expansion. The UV spectroscopic signature of a phenyl-monosaccharide is highlighted in red.[49]

hydroxy groups can interact together in a clockwise[47] chain of intramolecular H bonds OH2→OH3→OH4,[48] which we will label "c", or in a counterclockwise[47] chain OH4→OH3→OH2, which we will label "cc". The conformation of the hydroxymethyl group is labeled according to the previous convention adopted in ref 38. The labeling convention is exemplified in Figure 9.

Results and Conformational Assignment

αpMan. Bare Molecule. The R2PI and UVHB spectra of αpMan shown in Figure 1 reveal three distinct conformers, labeled as A, B, and C. Conformers A and C clearly exhibit similar vibronic structures dominated by a strong peak, followed at ∼50 cm^{-1} to higher wavenumbers by a doublet, highlighted in red in Figure 1.[49] Comparison with the previous studies of βpGlc and βpGal identify this pattern as the typical UV spectroscopic signature of an individual conformer of a phenyl-monosaccharide.[35,36,38] According to ground-state frequency calculations, the doublet most likely corresponds to torsional motions of the phenyl ring (which probably do not change much in the electronic state S_1). This would explain why the vibronic structure is apparently insensitive to the molecular conformation of the sugar. Assuming similar transition moments for the three conformers, A is by far the most strongly populated.

The IRHB spectra of each conformer are displayed in Figure 2 and compared with those calculated for each of the four lowest lying computed conformers.

Conformer A is the most abundant, which immediately suggests its assignment to the global minimum energy conformation, cG-g+, an assignment that is supported by the observed IR spectral pattern. The apparent dip in the band at 3620 cm^{-1} is the result of decreased IR power due to atmospheric water absorption at this wavelength. The asymmetrically broadened

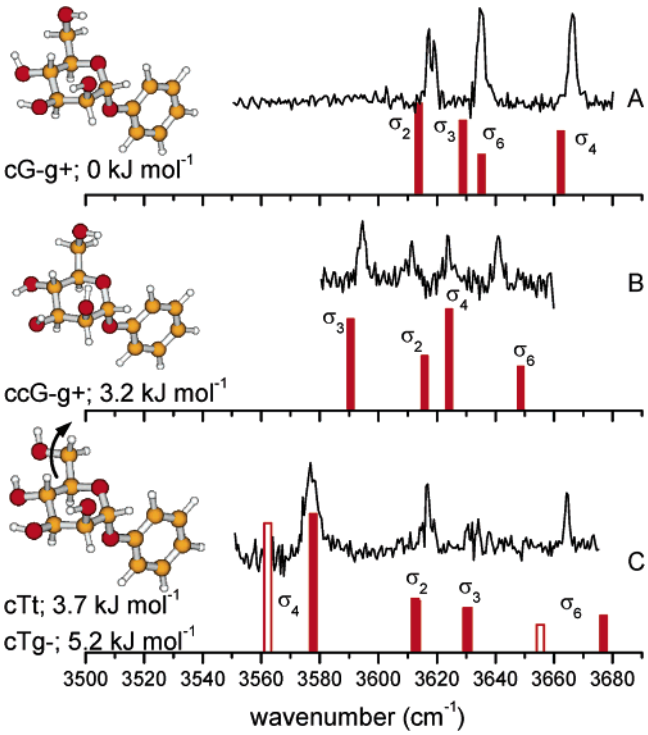


Figure 2. Experimental IRHB spectra of the three observed conformers of αpMan together with the calculated structures, relative energies, and spectra of the proposed conformational assignments. For conformer C, the calculated spectra of both possible conformational assignments are shown. The full and empty sticks correspond to the cTt (structure shown) and cTg- (structure not shown) conformations, respectively, which only differ by the torsion depicted by the arrow; their vibrational modes σ_2 and σ_3[48] are superposed. The intensity of the bands σ_2 of conformer A and σ_3 of conformer C is affected by atmospheric water IR absorptions. Larger size structures can be seen in Figure 9.

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[a]

	αpMan	βpGal	βpGlc
cG-g+	0 (1)	6.6 (4)	9.7
ccG-g+	3.2 (2)	15.3	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	4.8 (2)	4.5 (3)
ccG-g-	8.7	6.0 (3)	10.5

[a] ZPE corrections have been calculated from the B3LYP/6-31+G* frequency calculation. The numbers in parentheses indicate the energy ordering of the lowest lying conformers.

absorption feature centered at ∼3640 cm^{-1} suggests a contribution from two overlapping absorption features, and the calculated spectrum predicts two bands of medium intensity in this region, although the calculated splitting is somewhat greater. The relative intensity of the band at the highest wavenumber, centered at 3670 cm^{-1}, is enhanced by the power of the IR source, which increases toward higher wavenumbers.

The IR spectrum of the minor conformer B displays all four predicted IR bands, and their pattern accords well with the calculated spectrum of the second lowest lying conformer, ccG-g+; it is located at an energy at 3.2 kJ mol^{-1} above the global minimum, see Table 1, consistent with its low relative abundance.

Comparison of the IRHB spectrum associated with the second minor conformer, C, with calculated spectra strongly suggests

(47) The "clockwise" (c) and "counterclockwise" (cc) directions are defined using the orientation of the standard representation of monosaccharides where the ring oxygen is at the right top corner of the sugar ring as shown all along this report (see Figure 9). For example, the conformation labeled previously "ttt" in βpGlc and βpGal becomes "cc", while "ggg" becomes "c".

(48) The carbon atoms of the sugar ring are numbered as indicated for αpMan in Chart 1. OHn is the hydroxy group attached to the Cn. σ_n is the stretching vibration of OHn.

(49) The fact that the pattern cannot be easily seen on the UVUV spectrum of conformer B is due to the weakness of the signal associated with the doublet.

assignment to a cT conformation because of the presence of the strongly shifted band at 3575 cm^{-1}. The two best candidate conformers, cTt and cTg-, are shown here. The spectra predicted for either possible assignment do not allow a clear choice between them, and, given the small calculated energy gap between the two (1.5 kJ mol^{-1}), a definitive assignment to one or other of the possibilities should be resisted. These two conformers differ only in the orientation of the OH6 group; the H-bonded network remains the same in both cases. They are separated by a first-order saddle point, which generates a calculated zero-point potential energy barrier of only 1.4 kJ mol^{-1} (going from cTg- to cTt), but because the measurements were performed on molecules vibrationally cooled by collisions in the supersonic expansion down to temperatures of a few tens of Kelvins (1 kJ $mol^{-1} \approx 120$ K), the barrier is high enough for the conformers not to interconvert freely at these temperatures, despite their very minor structural difference.

In addition to the harmonic frequencies and IR absorption intensities, the calculations also provide the normal-mode analysis, which allows the assignment of each observed vibration as being predominantly due to the stretching motion of a particular OH group. The influence of hydrogen bonding on the vibrational mode can be recognized by considering the OH groups that are acting as donors. The easiest effect to measure is a shift of the OH stretching mode toward lower wavenumber, the strongest interactions leading to the largest shifts. Hydrogen bonding can also be revealed through increased intensity and line-width broadening (although power saturation of the vibrational transitions as well as fluctuations of the IR source power may obscure these effects). The feature labeled σ_4, for example, appears as a narrow band at a high wavenumber in conformer A, where the clockwise orientation of OH4 handicaps its interaction with either of its neighbors, O6 or O3. In conformer B, however, the OH4 group flicks from a clockwise to a counterclockwise orientation, facilitating its closer interaction with O3 and a consequent shift in σ_4 by ~40 cm^{-1} to lower wavenumber. In conformer C, the band shifts even further, by ~85 cm^{-1} to lower wavenumber, and it displays the enhanced intensity and width associated with hydrogen bonding. In either of the two possible assignments for C, the rotation of the hydroxymethyl group from a G- to a T orientation, coupled with the return of the neighboring OH to a clockwise orientation, brings OH4 and O6 into close proximity, creating a stronger OH4→O6 interaction. This evolution of the spectral shift of the vibration σ_4 with the conformation is consistent with the calculated distances between H4 and the oxygen atom with which it interacts. For conformer A, the distance H4−O6 is 3.2 Å, in conformer B the distance H4−O3 is 2.5 Å, and in either possible assignment of conformer C the H4−O6 distance is 2.0 Å.

There is a striking difference between the most stable conformation adopted by αpMan,[50] cG-g+, where the OH2 group is at an axial position, and by βpGlc and βpGal, ccG+g- (see Figure 9),[35,36] where the OH2 is equatorial. In these two cases, the equatorial position of OH2 and the β-anomeric configuration (O1 equatorial) favor an interaction between OH2 and O1, provided OH2 is oriented counterclockwise. The axial position of OH2 and the α-anomeric configuration (O1 axial) in αpMan prevent this arrangement. Instead, the peripheral OH groups switch to a clockwise conformation to facilitate the cooperatively interacting chain, OH2→OH3→OH4. The G-g+ orientation of the hydroxymethyl group, as adopted by conformer A of αpMan, is clearly stabilized by its role in elongating this clockwise H-bond chain.

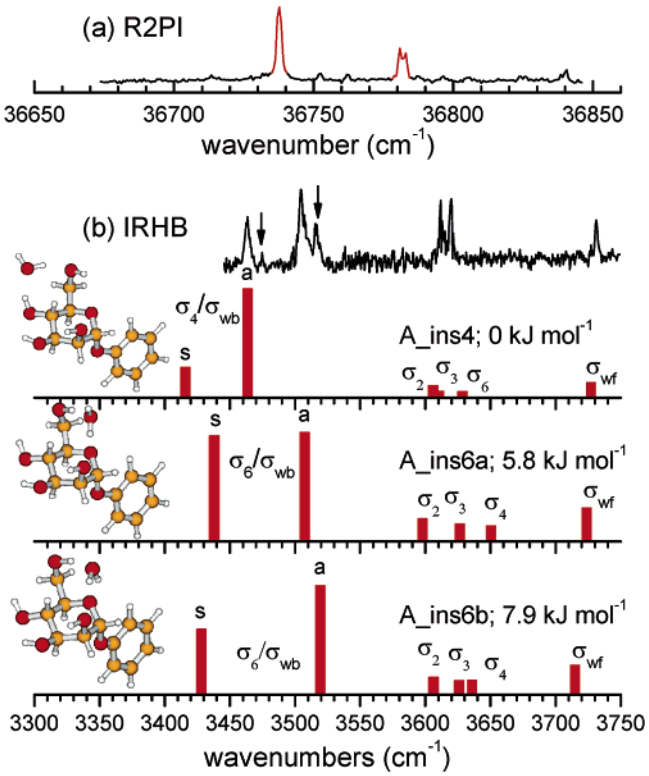


Figure 3. (a) R2PI spectrum of αpMan-W. The lines highlighted in red are those for which IRHB measurements have been made. (b) IRHB and calculated spectra and structures of αpMan-W. σ_{wb} and σ_{wf} refer to the bound and free OH stretching modes of the water molecule; the symmetric and asymmetric stretching modes of the groups involved in the intermolecular hydrogen bonds are labeled "s" and "a", respectively. Arrows indicate additional weak absorption bands, not predicted by the calculations.

αpMan-W Complex. The R2PI spectrum recorded by monitoring the ion signal in the mass channel of the αpMan-W complex shown in Figure 3a continues to display the characteristic vibronic structure discussed earlier (an intense band followed by two of lesser intensity ~50 cm^{-1} above the band origin). Its simplicity suggests the primary population of only a single hydrated conformer structure in the supersonic expansion, and IRHB measurements confirm this; each of the major UV features selected by the "probe" laser generates the same IR spectrum, shown in Figure 3b. The other, weaker UV bands are too weak to allow reliable IRHB measurements to be made, and there is evidence for associating some of these bands with larger clusters that fragment following ionization.

Although the IR spectra of each of the three most stable calculated structures of αpMan-W could all match qualitatively the observed IRHB spectrum, the relative energies of the second and third are calculated to lie ≥5 kJ mol^{-1} above the global minimum, see Table 2, high enough to make their population unlikely and consistent with the observation of a single

(50) In a very recent computational investigation of the conformation of mannose (without the phenyl chromophore) [Appell, M.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2005**, *340*, 459−468], the conformer tg-ggTt has been identified as the most stable conformation of the α anomer, but the authors did not consider the tg-ggG-g+ conformer − the unsubstituted analogue of the g-ggG-g+ conformer of αpMan reported here (labeled cG-g+). Using the same level of theory (B3LYP/6-311++G**), however, we find the tg-ggG-g+ conformer to be more stable than tg-ggTt, by 2.5 kJ mol^{-1}.

Table 2. Calculated ΔE_0 (kJ mol^{-1}) of the Monohydrated Structures Discussed in the Text (Numbers in Parentheses Indicate the Energy Ordering for the Most Stable Structures)

		αpMan-W	βpGal-W	βpGlc-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-	ins 2	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 all of the most stable conformations of αpMan and αpMan-W, the phenyl ring is perpendicular to the sugar ring (Figures 2, 3, and 9), but in the ccG+g-_ins6a/b complexes, the water bridges between OH6 and the phenyl ring, which is parallel to the sugar ring.

conformer, only. In addition, the spectrum calculated for the most stable structure reproduces the experimental one best, reinforcing its assignment to the insertion complex, A_ins4 (cG-g+_ins4). The conformation of the mannose sugar in this minimum energy structure is virtually identical to that of the lowest energy conformer of unhydrated αpMan. In bare αpMan, the band σ_4 appears at the highest wavenumber (see Figure 2) because of the structural restraints, which makes the interaction between OH4 and OH6 the weakest of those present in this conformer. In the monohydrate, the water molecule inserts between the OH4 and OH6 groups and replaces this poor interaction with two "better" hydrogen bonds,[51] and σ_4 is now located at the low wavenumber end of the IR spectrum. The bound OH group of the water molecule is the second "better" H bonded group, and its vibration is coupled with σ_4. In consequence, the two strongly perturbed bands observed at 3462 and 3504 cm^{-1} are associated with the symmetric and asymmetric modes of these coupled stretching vibrations. Similar vibrational couplings are observed in the IR spectra of most of the sugar−water complexes presented hereafter; their pattern provides a general "signature" of a hydrated insertion structure. Some additional weak absorption bands, which are not predicted in the calculated spectrum of the A_ins4 structure or indeed of any other of the computed conformational structures, are indicated by the arrows in Figure 3b. Similar bands have also been observed in the IR spectra of other bimolecular systems[38,52] and are often associated either with combination bands between OH stretch and intermolecular low frequency vibrations or combination/overtone bands made stronger by resonances with the OH stretching modes. It is also possible that the extra lines originate from other conformers of the complexes where the water molecule inserts at the same position but with a slightly different orientation. Only the vibrations of the modes involved in intermolecular hydrogen bonds would be sensitive to such minor differences.

***β*pGal. Bare Molecule.** The conformational landscape of βpGal is quite different from that of αpMan. Two conformers were observed and identified in an earlier investigation.[35] The most abundant, conformer A, was assigned to the most stable ccG+g- structure. In the course of the present study, another minor conformer, C, has been identified in the R2PI spectrum, see Figure 4a, and assigned to the conformer cG-g+ (which

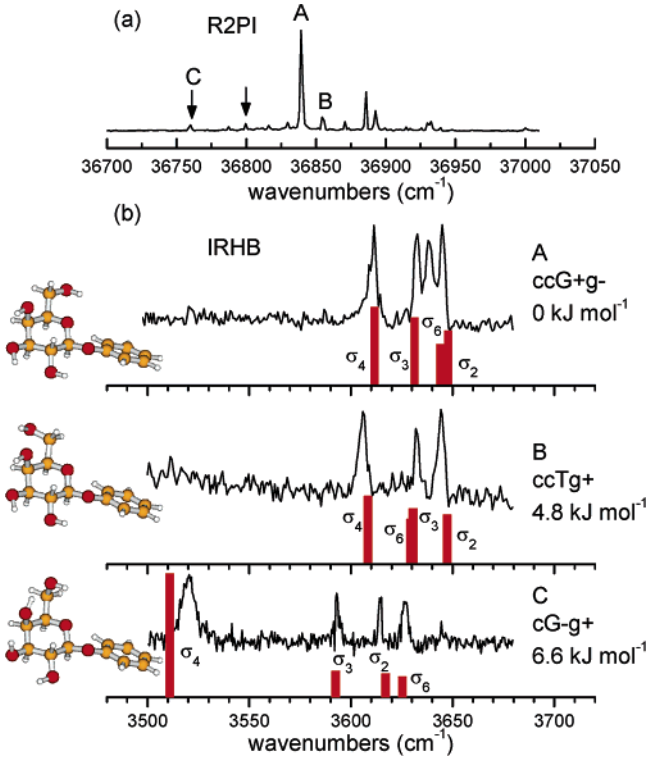


Figure 4. (a) R2PI spectrum of βpGal. The arrows mark the lines associated with the newly identified conformer C. (b) IRHB of the three observed conformers of βpGal and calculated spectra of their respective conformational assignment.

corresponds to the most stable structure in αpMan). Its experimental and computed IRHB spectra and structure are shown in Figure 4b along with the computed and observed spectra of all of the populated conformers.

Their assignments correspond to three of the four lowest lying conformers listed in Table 1, and their relative populations reflect the rank order of the relative energies. The minor conformer C is assigned to the structure cG-g+, which corresponds to the major, lowest energy conformer in αpMan. Because the OH4 group has an axial orientation in βpGal, the cG-g+ conformation in conformer C brings OH4 and OH6 into close proximity substantially enhancing the OH4→O6 interaction; the σ_4 mode is substantially shifted to lower wavenumber and is associated with an intense and broad absorption band.

Figure 5a compares the R2PI spectra of βpGal, αpMan, and βpGlc. Strikingly, the spectra of all of the observed conformers in which there is a hydrogen bond from OH2 to the phenoxy oxygen of the chromophore are located above 36 830 cm^{-1} (indicated by a vertical dashed line), whereas those lacking this interaction are all found at lower wavenumbers. This correlation can (and will) be used to support structural assignment of other hydrated sugar conformations.

***β*pGal-W Complexes.** The R2PI spectrum of hydrated βpGal is shown in Figure 5b, where it can be compared with the spectra generated by each of the other two hydrated sugars, αpMan-W and βpGlc-W. It is much more congested than that of αpMan-W or βpGlc-W, which suggests the population of a number of hydrate structures generating several overlapping R2PI spectra.[53]

(51) To appreciate whether one H bond is "better" than another, we consider both the $H_{donor}-O_{acceptor}$ distance and the directionality of the noncovalent interaction, which are provided by the DFT-optimized geometry.

(52) Gerhards, M.; Unterberg, C.; Gerlach, A.; Jansen, A. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2682−2690.

(53) The broad absorption indicated by the red dotted line below the resolved R2PI features shown in Figure 5b probably includes further unresolved vibronic progressions associated with P,U or additional, more weakly populated hydrate structures.

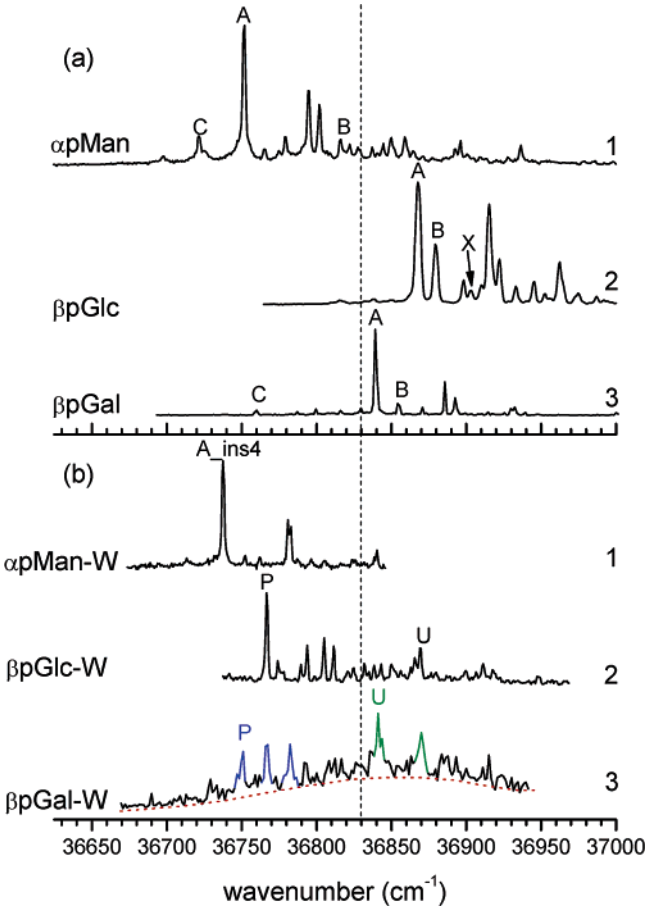


Figure 5. R2PI spectra of (a) the bare monosaccharides and (b) their hydrated complexes. The vertical dashed line marks the separation between the spectral origins of conformers associated with (above the line) and without (below) an OH → $O_{phenoxy}$ hydrogen bond.

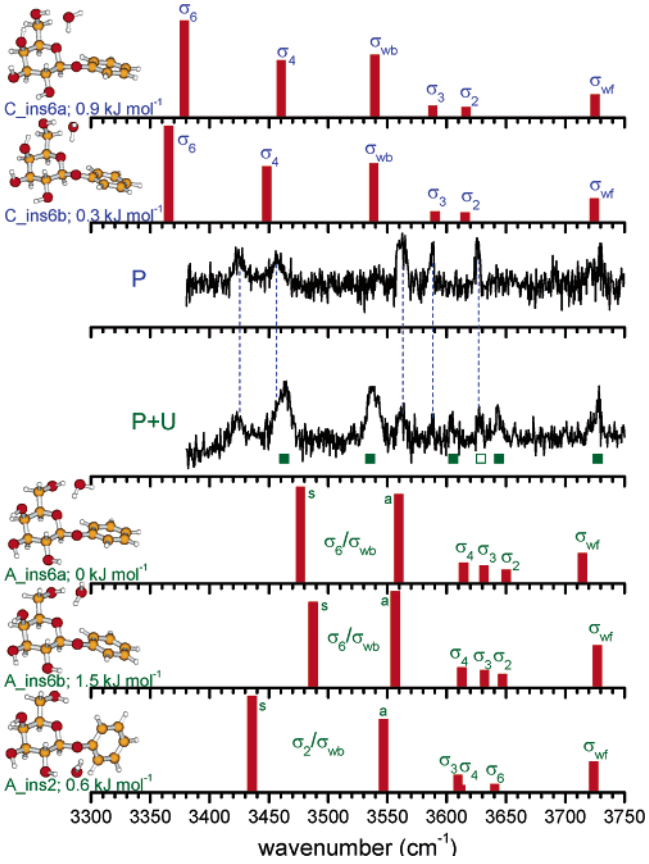


Figure 6. Experimental and calculated IRHB spectra of *β*pGal-W. P and P+U are the IRHB spectra associated with the lines respectively highlighted in blue and green on the R2PI spectrum in Figure 5b. The lines marked by squares on the spectrum P+U are discussed in the text.

This is consistent with the closely spaced relative energies calculated for the low-lying structures of *β*pGal-W, five of which lie within 1.5 kJ mol^{-1} of the global minimum (see Table 2). A similar situation was encountered previously in an investigation of the hydrated clusters of tryptophan,[54,55] where the spectral overlap was so severe that it prevented the resolution of their individual spectra through UV (or IR) hole-burning spectroscopy. In the present case, however, it has proved possible to conduct IRHB measurements on the five strong and sharp lines highlighted in color in Figure 5b.

The three blue lines (Figure 5b) all generated the same IRHB spectrum, labeled P (Figure 6), while more complex IRHB spectra, labeled P + U, were generated from the R2PI features highlighted in green (Figure 5b). They reflect blended features in the R2PI spectrum, which generate overlapping contributions from P and from a second hydrate structure, U.

The calculated IR spectra of the five lowest lying structures of *β*pGal-W are shown in Figure 6 along with those associated with the experimental IRHB spectra, P and P + U. Spectrum P displays three strong and broad absorption bands at the lower wavenumber end of the spectrum, at 3423, 3456, and 3563 cm^{-1}. They indicate three OH groups acting as H donors in three good hydrogen bonds. The only low lying structures that allow this are C_ins6b or 6a, formed by insertion of the water molecule between OH6 and the ring oxygen in the cG-g+ conformer (the weakly populated conformation C in bare *β*pGal); they each generate a strongly H-bonded chain, OH4→OH6→OHw→Oring, and their calculated IR spectral patterns are both in fairly good agreement with the observed spectrum. The most strongly shifted band, σ_6 in *β*pGal-W (C_ins6a/b), is the least shifted band in the cG-g+ conformer of bare *β*pGal, which provides another illustration of the conclusion drawn from the analysis of αpMan and αpMan-W: the water molecule seeks to replace the weakest intramolecular hydrogen bond of the bare molecule by two strong intermolecular hydrogen bonds.

The IRHB spectrum of the second hydrate, U, is partly contaminated by absorption bands associated with the hydrate structure P because of the R2PI spectral overlap. Nonetheless, five of the six predicted IR bands of U can easily be differentiated from those of P, and they are identified in Figure 6 by the full squares. The sixth IR-active OH stretching mode could well be associated with the blending of two overlapping bands associated with structures P and U at 3626 cm^{-1} (identified by an empty square in Figure 6). This spectrum agrees well with the spectra calculated for either of the two low-lying hydrate structures, A_ins6a or A_ins6b, formed by insertion of the water molecule at the same position as in P, but this time into the lowest energy conformation of bare *β*pGal, ccG+g-. (Another possible explanation for the "missing line" could be the blending of two absorption bands of the same

(54) Carcabal, P.; Kroemer, R. T.; Snoek, L. C.; Simons, J. P.; Bakker, J. M.; Compagnon, I.; Meijer, G.; von Helden, G. *Phys. Chem. Chem. Phys.* **2004**, *6*, 4546−4552.
(55) Snoek, L. C.; Kroemer, R. T.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2130−2139.

conformer; in this case, the structure A_ins2, which still involves the ccG+g- conformer, would provide a better fit.)

The R2PI lines leading to the IRHB spectra P and P+U (Figure 5) lie on opposite sides of the empirical line that separates the conformers with an OH→$O_{phenoxy}$ interaction (lying above the line, conformer U) and those without this interaction (lying below the line, conformer P). This supports further the assignment proposed here from the IR spectral interpretation; two families of structures dominate the conformational population of βpGal-W: one in which the water binds to conformer A of the bare molecule, and a second one in which conformer C (very weakly populated in bare βpGal) seems promoted by the complexation with a single water molecule. However, recognizing the experimental and theoretical difficulties associated with the study of βpGal-W, a more definitive assignment of structures within the families is not justified.

***β*pGlc-W: A Conformational Reassignment.** βpGlc was the first carbohydrate monosaccharide to be investigated under jet-cooled conditions in the gas phase, allowing its three most stable conformers to be separately resolved and conformationally assigned.[36] Subsequently, R2PI, UVUV, and IRHB spectra of the monohydrated complexes of βpGlc were reported and two distinct structures were identified.[38] Their definitive conformational and structural assignment was handicapped, however, by the lack of convincing theoretical support. The theoretical conformational search in this original study was restricted to complexes involving only those conformations of the bare molecule that had been detected experimentally, but the present investigation of βpGal-W has revealed that water binding can stabilize alternative conformations that are unpopulated or only weakly populated in the bare monosaccharide. In view of this, the conformational landscape of βpGlc-W has been reinvestigated using the unrestricted conformational searching strategy described earlier. This has led to the prediction of a global minimum conformational structure that had been missed both in our original investigation[38] and, subsequently, by others.[56] In the newly identified most stable structure of βpGlc-W, the sugar adopts a cG-g+ conformation, which in the bare molecule is calculated to lie at an energy 10 kJ mol^{-1} above the global minimum structure, ccG+g- (Table 1). Not surprisingly, the cG-g+ conformer was not detected in experiments on nonhydrated βpGlc.[36]

Newly measured IRHB spectra of hydrated βpGlc, recorded in the mid-IR using the radiation provided by the free electron laser FELIX,[57] are shown in Figure 7 together with the near-IRHB spectra reported earlier[38] and a series of calculated spectra. The agreement between the computed spectrum associated with the (newly identified) global minimum structure of βpGlc-W, cG-g+_ins4,[58] and the experimental spectrum associated with the principal conformer P is very good over the whole range of

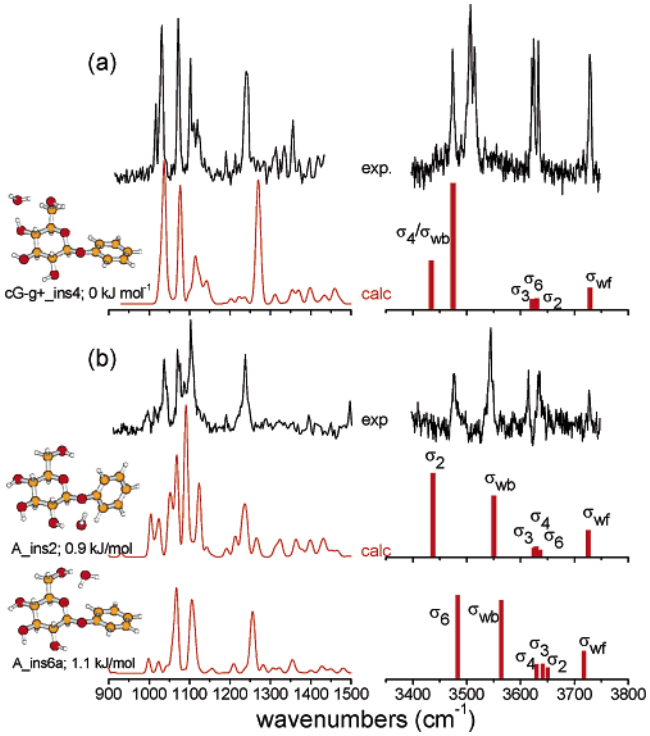


Figure 7. Experimental and calculated IRHB spectra of βpGlc-W recorded in the mid- and near-IR. As the density of IR-active modes in the mid-IR region is large, the calculated spectra are presented as convolutions of the stick spectra with a Gaussian line shape function of the same width as the bandwidth of the free electron laser, that is, 1% of the central wavelength.

the IR we have inspected, thus encouraging its confident assignment. The reliability of the (harmonic) DFT frequency calculations in the mid-IR (between 900 and 1700 cm^{-1}) is reinforced by the very good agreement between the calculated and experimental IRHB spectra of bare βpGlc, shown in Figure 8, following the same conformational assignments made earlier on the basis of near-IR spectra observations.[36] Most significantly, the assignment again demonstrates that water binding can drastically change the preferred conformation of the sugar.[59]

The observation of a substantial population of the cG-g+ conformer of βpGlc-W, a conformation not identified in similar experiments on the bare molecule and calculated to be quite high in energy (ΔE_0 = +9.7 kJ mol^{-1} in the bare molecule; see Table 1) and thus not significantly populated even at the relatively high temperature of the oven, suggests that significant conformational change may occur during or after formation of hydrated complexes in the jet. Although the binding of a single water molecule may provide enough energy to alter the saccharide conformation, it is also possible that larger hydrated clusters are formed early in the expansion and the singly hydrated clusters result from evaporative cooling coupled with conformational reorganization.

The structural assignment of the less abundant conformer, U, of βpGlc-W still remains ambiguous despite the new, more thorough theoretical survey and the extended mid-IRHB measurements that have been made since its first observation.[38]

(56) Momany, F. A.; Appell, M.; Strati, G.; Willett, J. L. *Carbohydr. Res.* **2004**, *339*, 553−567.
(57) Oepts, D.; van der Meer, A. F. G.; van Amersfoort, P. *Infrared Phys. Technol.* **1995**, *36*, 297.
(58) In a purely computational study of the unsubstituted monohydrate, Glc-W [Momany, F. A.; Appell, M.; Strati, G.; Willett, J. L. *Carbohydr. Res.* **2004**, *339*, 553−567], structures where the water binds to OH1 (a ccG-g+_ins1 structure) were identified as the most stable, but the cG-g+ type structure reported here was not considered. Optimization of βGlc-W without the phenyl ring in the cG-g+_ins4 configuration, conducted at the same level of theory used in that report (B3LYP/6-311++G**), leads to a structure lying at a relative energy 4.8 kJ mol^{-1} below the ccG-g+_ins1 structure. This result supports the view that the phenyl ring does not change, or indeed bias the conformational preferences of hydrated Glc.
(59) The conformational landscapes of amino acids can also be severely affected upon complexation with one water molecule as already shown in the case of phenylalanine [Lee, K. T.; Sung, J.; Lee, K. J.; Kim, S. K.; Park, Y. D. *J. Chem. Phys.* **2002**, *116*, 8251−8254] and tryptophan [Snoek, L. C.; Kroemer, R. T.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2130−2139].

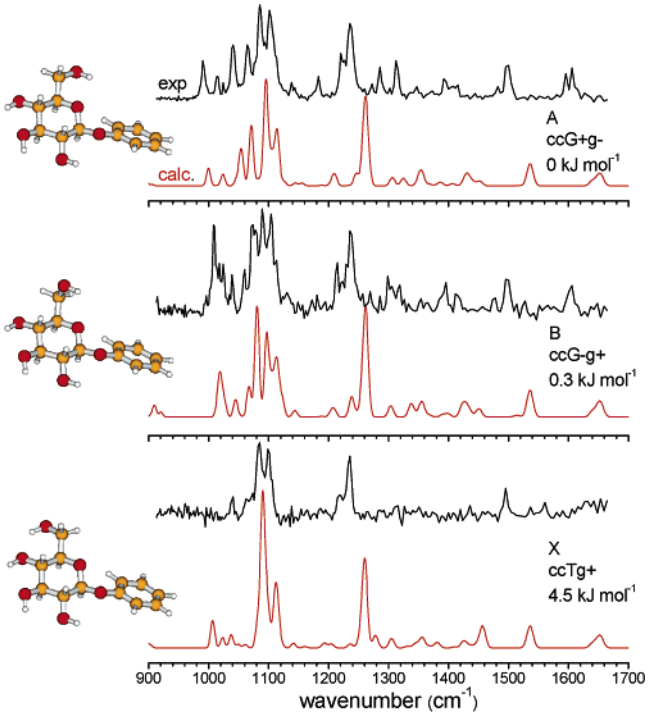


Figure 8. Mid-IRHB spectra of the three conformers of bare βpGlc.

Either of the two low energy structures, A_ins2 or A_ins6a, could provide a reasonable (although not quite satisfactory in either case) fit to the experimental spectra (in the mid and near-IR), and, given the small calculated energy difference between them (Table 2), a definitive assignment cannot be justified. It is always possible that there exists another structure that has eluded the conformational search; however, despite the computational effort to find other stable structures of βpGlc-W, better candidates for the conformational assignment of the minor hydrate U have not been found. The location of the R2PI band origin of the hydrate P (with a cG-g+ conformation) below 36 830 cm^{-1}, while that of U is located above it, provides further support, albeit empirical, for assigning U to a hydrated counterclockwise conformer.

Discussion

Cooperativity and Conformational Selectivity: Monosaccharides. Given the variety of conformational landscapes presented by each of the bare monosaccharides, their clear preference for the cG-g+ conformation when hydrated by a single water molecule is remarkable. In the case of αpMan, the preference might have been expected because the cG-g+ conformation is also preferred in the bare molecule. The situation is quite different, however, for βpGal and βpGlc. The cG-g+ conformer of βpGal (identified here for the first time) is calculated to lie ~6 kJ mol^{-1} above the global minimum and is only very weakly populated. In the case of βpGlc, its relative energy is even higher, ~10 kJ mol^{-1}, and its population has not been detected. In the monohydrated sugars βpGal-W and βpGlc-W, the structures incorporating the cG-g+ conformation are among the most stable and are the most populated. The binding of a single water molecule to the monosaccharides significantly changes their conformational landscapes and preferred structures. Figure 9 summarizes the conformations and structures identified in the present work. It is interesting to note that the conformational landscapes of the hydrated complexes appear to be simpler than those of the bare molecules and to be more uniform from one monosaccharide hydrate to another.

The strong conformational selectivity can be understood in terms of cooperative interactions, which augment the strength of a hydrogen-bonded network of linked OH groups. Considering the OH groups as polarizable dipoles, their hydrogen bonding, which is dominated by the electrostatic dipole−dipole interaction, will be enhanced by the inductive interaction between the dipole of the acceptor OH and the induced dipole of the donor OH. As the number of interacting OH groups increases, and with the increased strength of some of the individual H-bonds resulting from insertion of a water molecule, cooperativity acts to enhance the strength of the entire hydrogen-bonded network. In each monosaccharide, it is the cG-g+ conformation that provides the longest uninterrupted chain of hydrogen bonds in their monohydrated structures and the water molecule binds preferentially to this conformation to maximize the cooperative effect. The resulting hydrate structures can be considered as "locked" by cooperative hydrogen-bonded networks including both intra- and intermolecular hydrogen bonds. In the hydrates of αpMan and βpGlc, where the water inserts at position 4, the chain includes the sequence OH2→OH3→OH4→OH_w→OH6→O_{ring}; in βpGal-W, the water inserts at position 6 to generate the sequence OH2→OH3→OH4→OH6→OH_w→O_{ring}.

Cooperativity and Conformational Selectivity: Disaccharides. The conformation adopted by the isolated disaccharide benzyl β-lactoside (βBnLac) provides a further example of conformational "locking" by cooperativity. As shown in Figure 10a, its seven OH groups form an uninterrupted hydrogen-bonded network that circles around the molecular framework, locks the orientation of the OH groups, and, most significantly, generates a rigid scaffold across the glycosidic bond through the linked pair of inter-ring hydrogen bonds.[37] To achieve this, the Gal and the Glc components of the disaccharide each adopt the cG-g+ conformation, also favored in their monohydrate structures, allowing inter-ring hydrogen bonding and stabilization of the cis configuration (with respect to the two hydroxymethyl groups). This is illustrated in Figure 10d, which superimposes the most stable structures of hydrated βpGal and unhydrated βBnLac. The water molecule in βpGal-W plays the same role as the OH6 group of the Glc component of βBnLac, in stabilizing the cG-g+ conformation of the galactose unit. The intramolecular (Gal) hydrogen bond geometries are very similar in βpGal-W and βBnLac, and their OH2, OH3, and OH4 groups superimpose almost perfectly.

Ab initio calculations of methyl β-lactoside, which has a conformational landscape similar to βBnLac (revealed at a reduced computational cost), see Figure 10a,b, and its monohydrate generated the lowest energy hydrate structure shown in Figure 10c. The next most stable hydrate (not shown) is significantly higher in energy ($\Delta E_0 = +6.1$ kJ mol^{-1}). The clear preference identified in the hydrated monosaccharides, for a bound water molecule to replace the weakest intramolecular hydrogen bond by two strong intermolecular bonds, applies also to the disaccharide, and perhaps to oligo-saccharides in general; 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 10c). The disaccharide

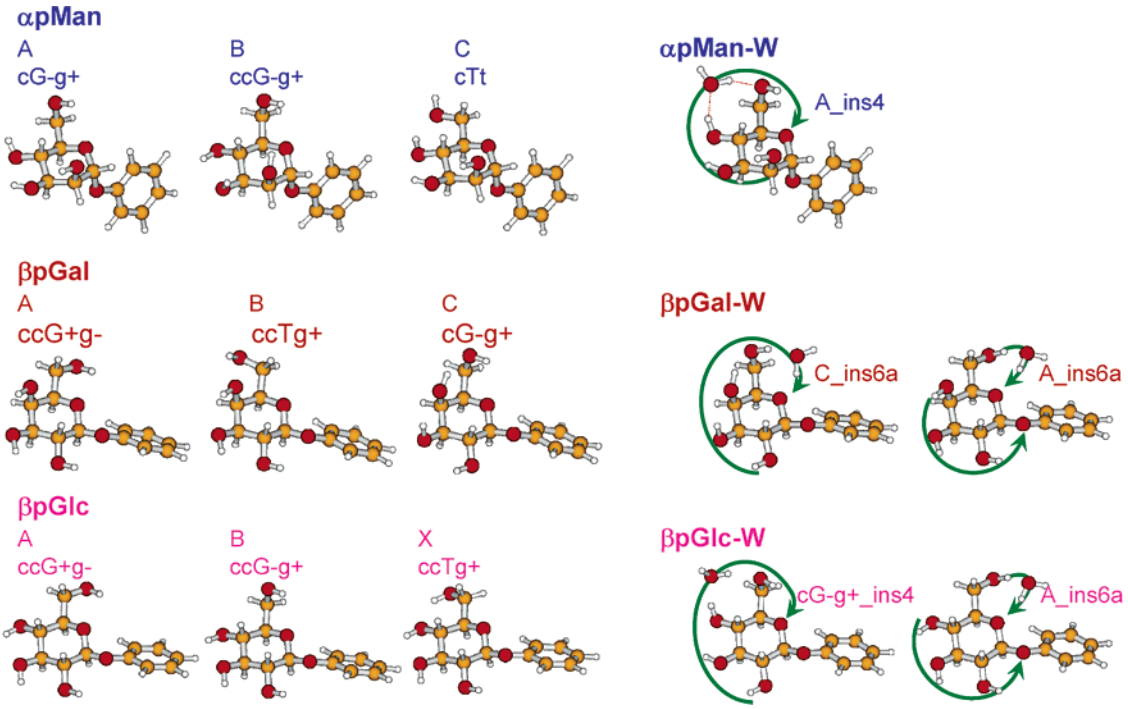


Figure 9. Comparison of the conformers discussed in the present work. The green arrows circling around the structures of the hydrated cluster schematically represent the cooperative H bond networks that stabilize the structures and favor fully cooperative conformations (structures A_ins4 of αpMan-W, C_ins6a of βpGal-W, and cG-g+_ins4 of βpGlc-W). The two conformers of βpGal-W and conformer A_ins6a of βpGlc-W are representative of the conformational families to which they belong and that have been identified and discussed in the text.

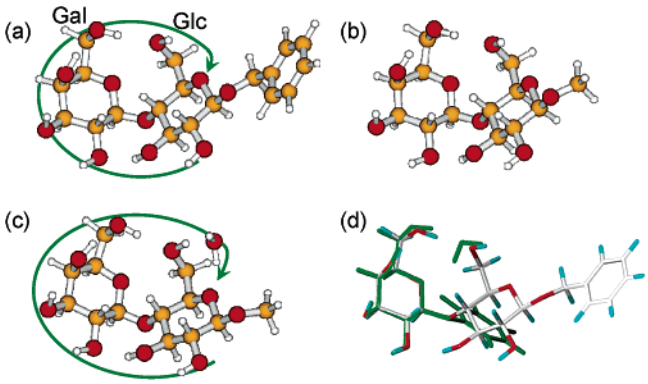


Figure 10. Most stable computed structures of (a) benzyl β-lactoside (the unique conformation observed in the gas phase),[37] (b) methyl β-lactoside, and (c) its monohydrate. (d) Superposition of the most stable structure of benzyl β-lactoside (colored) and the C_ins6 conformational structure of βpGal-W (green).

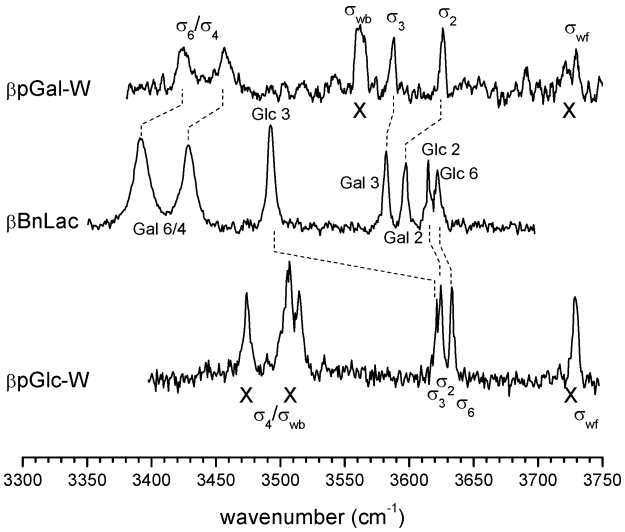


Figure 11. Comparison of the IR spectra of βpGal-W in the C_ins6 conformation (upper), βpGlc-W in the cG-g+_ins4 conformation (lower), and βBnLac (middle) where both the Gal and the Glc groups adopt the cG-g+ conformation.[37] The normal modes of βBnLac are labeled by the monosaccharide to which the corresponding OH belongs and its number within this monosaccharide. 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.

structure retains the fully cooperative cis configuration (enthalpically favored in the nonhydrate) of its two hydroxymethyl groups, and the two strong hydrogen bonds connecting OH3(Glc) to OH2(Gal) and OH6(Gal) to OH6(Glc) are still retained.

Indicators of Cooperative Hydrogen Bonding. Having presented a qualitative theoretical view of cooperativity and its structural consequences, is there clear experimental evidence of its manifestation? The experimental IR spectra of βBnLac, and of βpGal-W and βpGlc-W, each bound in their cG-g+ conformation to a water molecule inserted respectively at OH6 and OH4, are displayed for comparison in Figure 11. The most striking difference is the position of the absorption band associated with the stretching mode σ_3 of Glc, which is displaced from 3622 cm^{-1} in βpGlc-W to 3492 cm^{-1} in the disaccharide. This large shift reflects the strong bonding across the glycosidic linkage, OH3(Glc)→OH2(Gal); in the water complex βpGlc-W, the interaction OH3(Glc)→OH4(Glc) is much weaker. The two remaining bands in the water complex βpGlc-W, σ_2 and σ_6, are also displaced toward lower wavenumber in the disaccharide, although the shift is very much smaller. The H-bonded groups in βpGal-W display a similar pattern of behavior: bands σ_2-(Gal), σ_3(Gal), and the coupled, strongly hydrogen bonded pair σ_4/σ_6 all shift to lower wavenumbers in the disaccharide.

In the hydrated monosaccharides, the stabilizing chain only incorporates five interacting OH groups, but the disaccharide

is stabilized by a hydrogen-bonded chain of seven interacting OH groups. Consequently, the universal shift to lower wavenumber of all of the OH stretching vibrations would appear to reflect a stronger cooperative effect in the disaccharide than in the monosaccharide−water complexes. In particular, in the absence of any additional cooperativity, the bands σ_2(Gal) and σ_3(Gal) would be expected to remain at the same wavenumbers in the disaccharide and the monohydrate, because they are involved in the same intramolecular hydrogen bonds in both cases; see Figure 10d. This is a particularly interesting and valuable result because such observations could be used in other systems to provide a direct experimental signature of cooperativity. The potential for providing such experimental evidence demonstrates the advantage of observing well-resolved vibrational spectra of rotationally cold molecules, isolated, hydrated, or complexed, in the gas phase. Measurements in the near-IR, of the spectral shifts, bandwidths, and absorption intensities of OH (or perhaps NH) stretching modes, can be analyzed to understand how hydrogen-bonded networks operate to stabilize (or rigidify) particular structures and to probe directly the groups that play an active role in these networks as hydrogen bond donors and/or acceptors.

Protein Binding, Hydration, and Selectivity. The present work identifies some general trends in the preference for water to bind to "weak spots" in carbohydrates and points at cooperativity as an important phenomenon at the origin of the conformational choice/selection of monohydrated sugar molecules. With these results obtained on systems in conditions fairly remote from the biological medium, it would be fair to raise the question of the relevance of such studies: "How do these findings translate to larger scale systems, and can they be used to explain biological processes?" These questions are addressed here through some examples.

Inspection of the 14 known X-ray crystal structures of proteins that contain bound simple monosaccharide α-mannosides as ligands[60−70] reveals the "weak" OH4−OH6 binding site as a primary point of recognition in many carbohydrate-binding domains[60−62,65,67,69−71] (although in some cases, e.g., *Escherichia coli* FimH[71] and Jacalin,[60] the binding sites are deeper than others and other binding modes are also employed, e.g., OH3−OH4,[63] O2−O3,[68,72] Ca(II)−O3−O4,[63] O2−Ca(II)−O3−Ca(II)−O4[66]). Interestingly, disruption of the OH4−OH6 weak spot (as well as primary binding at OH2−OH3) also leads to a G+ OH6−W−O_{ring} hydration mode in snowdrop lectin,[68] akin to one of those seen in βpGal-W (A_ins6a/b). In the same way that monosaccharide hydration modes aim to generate cooperativity, nature also appears to exploit similar facets of globally cooperative hydrogen bond recognition in its ability to bind simple monosaccharide ligands.

Other sugars also show hydration trends in nature that correlate with those found here. A strong water-mediated hydrogen bond between O6 and Leu86 plays an important role in the binding of galactose to the lectin from *Erythrina corallodendron*;[20] the Gal OH6 site identified through our hydration studies is one that nature has also selected. In the structure of peanut agglutinin with bound Gal-β(1,3)GalNAc (the Thomsen−Friedenreich Antigen structure),[21,22] the Gal moiety binding and, indeed, the molecular origin of specificity are mediated by a network of critical protein-bound water molecules within which the terminal Gal moiety adopts a G+ conformation, an orientation of OH6 we observe in both nonhydrated and singly hydrated βpGal. In the recently published X-ray structure of *Erythrina cristagalli* lectin, in which only the galactose moiety of lactose is bound as a ligand, there are no direct hydrogen bonds between the lectin and disaccharide;[23] instead, interactions are mediated by a set of structural water molecules. In this structure, Gal adopts a G+ conformation that allows the OH6 to hydrogen bond to a key binding water molecule.

In the structure of the archetypal α-mannoside binding protein, lectin conconavalin A (ConA),[67] methyl α-mannopyranoside is bound by the protein through direct hydrogen bonding primarily to OH6 and OH4, thus unpicking the weak intramolecular hydrogen bond of mannose, while the remaining hydroxyls remain largely exposed to the solvent. ConA thus achieves selectivity for mannose and glucose, in part by virtue of the fact that it replaces the weakest of the intramolecular hydrogen bonds in those saccharides (OH4−OH6) with beneficial interactions with the protein, much as water does in monohydrate complexes determined here for the first time. Our hydration studies thus appear to be a predictor of potential protein binding modes. If the point of first hydration represents the first vulnerable point that a given binding protein unpicks to recognize its cognate ligand, the second hydration point the second, and so on, the implications of such a model are profound and suggest that recognition of carbohydrates by proteins is not simply a case of matching hydrogen bond donors and acceptors and hydrophobic patches on sugars with those in proteins but, instead, should be considered in reference to the solvated ligand and the associated equilibria. This idea has been clear for some time, but until now there has been no direct experimental evidence to point to preferred hydration points on the ligand.

We would tentatively suggest that two general modes of carbohydrate binding are therefore exploited by nature: those in which carbohydrate-bound water molecules are replaced by direct hydrogen bonds and those in which the bound water molecules are preserved. Although our initial data are too limited to draw definitive conclusions, it would appear that in either

(60) Bourne, Y.; Astoul, C. H.; Zamboni, V.; Peumans, W. J.; Menu-Bouaouiche, L.; Van Damme, E. J. M.; Barre, A.; Rouge, P. *Biochem. J.* **2002**, *364*, 173−180 (PDB ID = 1kuj).
(61) Bourne, Y.; Roussel, A.; Frey, M.; Rouge, P.; Fontecillacamps, J. C.; Cambillau, C. *Proteins* **1990**, *8*, 365−376 (PDB ID = 1lob).
(62) Moothoo, D. N.; Canan, B.; Field, R. A.; Naismith, J. H. *Glycobiology* **1999**, *9*, 539−545 (PDB ID = 1bxh).
(63) Ng, K. K. S.; Drickamer, K.; Weis, W. I. *J. Biol. Chem.* **1996**, *271*, 663−674 (PDB ID = 1rdl, 1rdm).
(64) Ng, K. K. S.; Kolatkar, A. R.; Park-Snyder, S.; Feinberg, H.; Clark, D. A.; Drickamer, K.; Weis, W. I. *J. Biol. Chem.* **2002**, *277*, 16088−16095 (PDB ID = 1kwu).
(65) Loris, R.; Van Walle, I.; De Greve, H.; Beeckmans, S.; Deboeck, F.; Wyns, L.; Bouckaert, J. *J. Mol. Biol.* **2004**, *335*, 1227−1240 (PDB ID = 1ukg).
(66) Sudakevitz, D.; Kostlanova, N.; Blatman-Jan, G.; Mitchell, E. P.; Lerrer, B.; Wimmerova, M.; Katcoff, D. J.; Imberty, A.; Gilboa-Garber, N. *Mol. Microbiol.* **2004**, *52*, 691−700 (PDB ID = 1uqx).
(67) Naismith, J. H.; Emmerich, C.; Habash, J.; Harrop, S. J.; Helliwell, J. R.; Hunter, W. N.; Raftery, J.; Kalb, A. J.; Yariv, J. *Acta Crystallogr., Sect. D* **1994**, *50*, 847−858 (PDB ID = 5cna).
(68) Hester, G.; Kaku, H.; Goldstein, I. J.; Wright, C. S. *Nat. Struct. Biol.* **1995**, *2*, 472−479 (PDB ID = 1msa).
(69) Pratap, J. V.; Jeyaprakash, A. A.; Rani, P. G.; Sekar, K.; Surolia, A.; Vijayan, M. *J. Mol. Biol.* **2002**, *317*, 237−247 (PDB ID = 1j4u).
(70) de Souza, G. A.; Oliveira, P. S. L.; Trapani, S.; Santos, A. C. O.; Rosa, J. C.; Laure, H. J.; Faca, V. M.; Correia, M. T. S.; Tavares, G. A.; Oliva, G.; Coelho, L.; Greene, L. J. *Glycobiology* **2003**, *13*, 961−972 (PDB ID = 1mvq).
(71) Hung, C. S.; Bouckaert, J.; Hung, D.; Pinkner, J.; Widberg, C.; DeFusco, A.; Auguste, C. G.; Strouse, R.; Langermann, S.; Waksman, G.; Hultgren, S. J. *Mol. Microbiol.* **2002**, *44*, 903−915 (PDB ID = 1kiu).
(72) Hester, G.; Wright, C. S. *J. Mol. Biol.* **1996**, *262*, 516−531 (PDB ID = 1niv).

strategy nature seeks to "unpick" weak hydrogen-bonding interactions first: these weaker elements can be identified through our selective hydration method. The initial inspection of bound monosaccharide structures in proteins suggests that, although complete intramolecular hydrogen-bonding networks are rare in liganded form, the stronger interactions remain and the weakest are "unpicked" first. Hotspots for "unpicking" include those identified here by hydration: the OH4−OH6 region of mannose and the OH6−O_{ring} region of galactose.

Concluding Remarks

The structural investigation of a series of singly hydrated monosaccharides, isolated in the gas phase, has revealed the role of cooperativity in achieving the selective population of specific carbohydrate molecular conformations. The selectivity is created through maximizing the sequence of interacting hydrogen-bonded OH groups surrounding the carbohydrate; their cooperative interaction enhances the strength of the hydrogen-bonded network, stabilizing the conformation. The conclusions reached through the study of monosaccharides also apply to the disaccharide lactose and may be translated to larger systems in the future.

Cooperativity is already known to be an important factor in molecular recognition processes of large molecular assemblies.[3,4,12,73] To date, the water-mediated binding of carbohydrates by lectins has been considered in the context of structurally conserved water molecules associated with protein[20] rather than the evolution of proteins to recognize the given water shroud that each carbohydrate bears. From the early results gained here, this might equally be considered as being a function of the sugar rather than the protein. If the complete water shroud of a sugar can be determined, this might provide a general and broadly applicable method of ligand-fitting analysis of carbohydrates into proteins, something that to date has not been possible due to the unclear role played by water intermediaries in many structures.

(73) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123*, 12658−12663.

Taken together with our previous studies on a nonhydrated lactoside,[37] which showed the usefulness of considering nonhydrated structures as models for deeply bound ligands in carbohydrate-binding proteins, these results suggest that certain selectively hydrated structures, such as those determined here for mannosides, may provide good models for the recognition of partially solvated structures by proteins that have shallow binding sites (as is found in most typical lectins). Thus, through simple increase in hydration levels, we can begin to explore minimalist models of ligands in both deeply bound (desolvated) and shallowly bound (part or even fully solvated) environments, and better understand the binding modes that nature employs.

Acknowledgment. We are most grateful to Professor Gerard Meijer and Dr. Gert von Helden of the Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, for giving us access to their experimental setup at the FOM Institute Rijnhuizen, The Netherlands. We also gratefully acknowledge the support by the Stichting voor Fundamenteel Onderzoek der Materie (FOM) in providing the required beam time on FELIX and highly appreciate the skillful assistance provided by the FELIX staff. This work was supported by the European Community − Research Infrastructure Action under the FP6 "Structuring the European Research Area" Program through the Integrated Infrastructure Initiative "Integrating Activity on Synchrotron and Free Electron Laser Science". Finally, we acknowledge the support provided by the EPSRC, the Royal Society (R.A.J., USA Research Fellowship; L.C.S., University Research Fellowship), the Leverhulme Trust (Grant No. F/08788D), the CLRC Laser Loan Pool, and the Physical and Theoretical Chemistry Laboratory at Oxford.

Supporting Information Available: Complete ref 46. Computed Cartesian coordinates and total energies of all of the structures listed in Tables 1 and 2 (PDF and TXT). This material is available free of charge via the Internet at http://pubs.acs.org.

JA0518575