DOI: 10.1002/chem.200800474

Conformational Choice and Selectivity in Singly and Multiply Hydrated Monosaccharides in the Gas Phase

Emilio J. Cocinero,[a] E. Cristina Stanca-Kaposta,[a] Eoin M. Scanlan,[b] David P. Gamblin,[b] Benjamin G. Davis,*[b] and John P. Simons*[a]

Abstract: Factors governing hydration, regioselectivity and conformational choice in hydrated carbohydrates have been examined by determining and reviewing the structures of a systematically varied set of singly and multiply hydrated monosaccharide complexes in the gas phase. This has been achieved through a combination of experiments, including infrared ion-depletion spectroscopy conducted in a supersonic jet expansion, and computation through molecular mechanics, density functional theory (DFT) and ab initio calculations. New spectroscopic and/or computational results obtained for the singly hydrated complexes of phenyl β-D-mannopyranoside (β-D-PhMan), methyl α-D-gluco- and α-D-galactopyranoside (α-D-MeGlc and α-D-MeGal), when coupled with those reported earlier for the singly hydrated complexes of α-D-PhMan, β-D-PhGlc and β-D-PhGal, have created a comprehensive data set, which reveals a systematic pattern of conformational preference and binding site selectivity, driven by the provision of optimal, co-operative hydrogen-bonded networks in the hydrated sugars. Their control of conformational choice and structure has been further revealed through spectroscopic and/or computational investigations of a series of multiply hydrated complexes; they include β-D-PhMan·$(H_2O)_{2,3}$, which has an exocyclic hydroxymethyl group, and the doubly hydrated complex of phenyl α-L-fucopyranoside, α-L-PhFuc·$(H_2O)_2$, which does not. Despite the very large number of potential structures and binding sites, the choice is highly selective with binding invariably "focussed" around the hydroxymethyl group (when present). In β-D-PhMan·$(H_2O)_{2,3}$, the bound water molecules are located exclusively on its polar face and their orientation is dictated by the (perturbed) conformation of the carbohydrate to which they are attached. The possible operation of similar rules governing the structures of hydrogen-bonded protein–carbohydrate complexes is proposed.

Keywords: carbohydrates · hydrogen bonds · IR spectroscopy · molecular recognition · solvent effects

Introduction

Carbohydrate molecular recognition by proteins is a key factor in many biological processes,[1] but the factors dictating selective carbohydrate–protein interactions are not easily discerned. Selectivity must depend in part upon the conformational structures of the interacting carbohydrate ligand and its receptor. In addition, because of their flexibility and the universality of hydrogen-bonded interactions in aqueous environments, their structures may be influenced by explicit hydration, dehydration or by the direct involvement of water at the receptor site.[2] Modelling and understanding protein–carbohydrate interactions presents a complex challenge.[3]

The great majority of monosaccharide carbohydrate structures have been determined in the condensed phase, either in crystalline environments using X-ray crystallography[4–6] or in solution through molecular dynamics (MD) calculations and NMR spectroscopy measurements.[7–9] Their inherent, unperturbed conformational structures, free of any environmental effects were until recently only accessible in silico, computed through molecular mechanics methods

[a] Dr. E. J. Cocinero, Dr. E. C. Stanca-Kaposta, Prof. J. P. Simons
Chemistry Department, University of Oxford
Physical and Theoretical Chemistry Laboratory
South Parks Road, Oxford OX1 3QZ (UK)
Fax: (+44) 1865-275410
E-mail: John.Simons@chem.ox.ac.uk

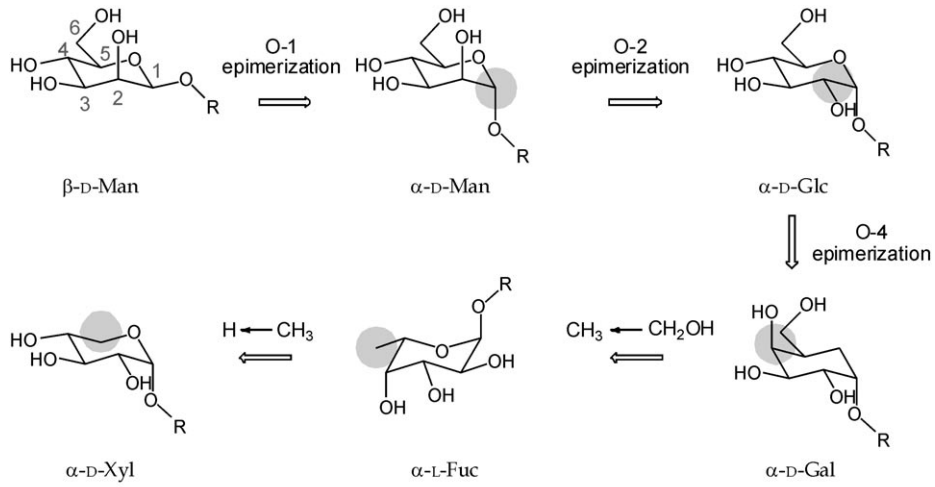
[b] Dr. E. M. Scanlan, Dr. D. P. Gamblin, Prof. B. G. Davis
Chemistry Department, University of Oxford
Chemistry Research Laboratory
12 Mansfield Road, Oxford OX1 3TA (UK)
Fax: (+44) 1865-275674
E-mail: Ben.Davis@chem.ox.ac.uk

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.200800474.

using a variety of force fields, through density functional theory (DFT) and ab initio calculations, or through Car–Parrinello MD calculations.[8–10] But a start has now been made in determining experimentally the inherent conformations of bare carbohydrates and the preferred conformations and structures of their singly hydrated complexes isolated in the gas phase.[11–14] The vibrational spectroscopic signatures of individually resolved conformers, initially isolated from the environment and then interacting with it in a controlled way within singly hydrated (and also molecular) complexes, have been obtained through mass-selected ultraviolet and infrared laser spectroscopy conducted at low temperatures in free jet expansions,[15] coupled with a combination of molecular mechanics, DFT and ab initio calculations, to enable conversion of their spectroscopic signatures into conformational and intermolecular structures.

In the gas phase, the preferred carbohydrate conformations, explicit water binding site(s) and their resulting intermolecular structures are dominated primarily by co-operative hydrogen bonding. Controlling factors that have been identified include the flexibility of their exocyclic hydroxymethyl group (present, for example, in glucose, galactose and mannose; Scheme 1), the relative orientations (axial versus equatorial) of their hydroxyl groups, and their anomeric configuration.[13,16,17] These factors can operate separately or collectively, adapting the carbohydrate conformation to optimize the sequence of intra- and inter-molecular hydrogen-bonded interactions in the hydrated complex.[11,13,16,17] The first investigations suggested a set of "working rules" based upon the preferred structures of a series of singly hydrated monosaccharides (see Scheme 1), including the β-anomer of phenyl[1] D-glucopyranoside (β-D-PhGlc, in which all the peripheral hydroxyl groups are oriented equatorially) and the β-anomer of phenyl D-galactopyranoside[16] (β-D-PhGal, in which OH4 is oriented axially); the α-anomer of phenyl D-mannopyranoside (α-D-PhMan, in which OH2 is oriented axially); and the α- and β-anomers of their truncated analogues, phenyl D-xylopyranoside[17] and L-fucopyranoside[13] (in which the exocyclic hydroxymethyl is altered to methyl, α/β-L-PhFuc, or to no substituent at all, α/β-D-PhXyl).

The present work sets out, firstly, to test the predictive power of these working rules by exploring experimentally and also computationally their operation in governing the preferred conformations and structures of a comprehensive panel of systematically varied monosaccharides, including the singly hydrated anomers β-D-PhMan, α-D-MeGal and α-D-MeGlc together with α-D-PhMan, β-D-PhGlc and β-D-PhGal; and secondly, to identify the working rules that may govern *multiply* hydrated carbohydrates by exploring the structural preferences of the two "trial" systems, β-D-PhMan·$(H_2O)_{2,3}$ and α-L-PhFuc·$(H_2O)_2$, with and without the exocyclic CH_2OH group. Collectively this panel of monosaccharides (Scheme 1) provides a powerful, structurally iterated series that can be used to map out key aspects of the hydration of these important biomolecules.

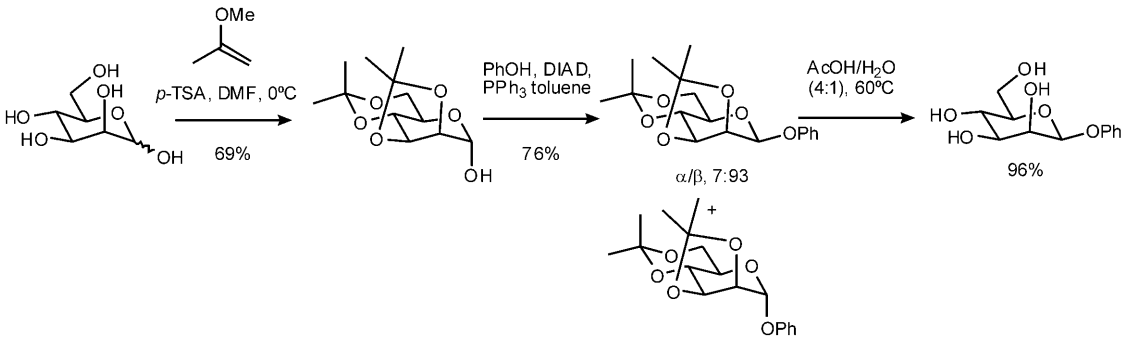


Scheme 1. Schematic representation of the structural inter-relations in the conformational panel used to probe the hydration of monosaccharide carbohydrates. R = phenyl, methyl.

Results and Discussion

Synthesis: Phenyl β-D-mannopyranoside (**1**), a key additional monosaccharide, was prepared in three steps starting from commercially available D-mannose using a highly β-selective Mitsunobu[18,19] glycosylation strategy (Scheme 2). Selective protection of the 2, 3, 4 and 6 hydroxyl groups was achieved through reaction of D-mannose with two equivalents of 2-methoxypropene in the presence of *para*-toluenesulfonic acid (*p*-TSA). The resulting kinetic diisopropylidene mannoside[20] was then reacted with phenol using triphenylphosphine and diisopropyl azodicarboxylate (DIAD) to furnish the *O*-phenyl compound as a mixture of α and β isomers (7:93). The high β selectivity is notable given the difficulties associated with the synthesis of β-mannosides and outstripped that associated with several other methods.[21] This first example of Mitsunobu mannosylation may therefore have general application and its wider synthetic utility is currently being explored. Pure β-mannoside was readily obtained in 71% isolated yield following separation from the α-anomer

[1] The phenyl group provides the ultraviolet chromophore required for detection by means of the resonant two-photon ionization step in the infrared ion-depletion (IRID) detection technique. DFT and ab initio calculations have confirmed that its presence has a negligible influence on the conformational landscape of small carbohydrates.[11]



Scheme 2. Synthesis of phenyl β-D-mannopyranoside.

by flash chromatography. Hydrolytic deprotection in acetic acid/water furnished the desired *O*-phenyl β-mannoside in 96% yield.

UV and IR spectroscopy of β-D-PhMan: Figure 1 shows the resonant two-photon ionization (R2PI) spectrum of β-D-PhMan recorded in the argon expansion, non-solvated and

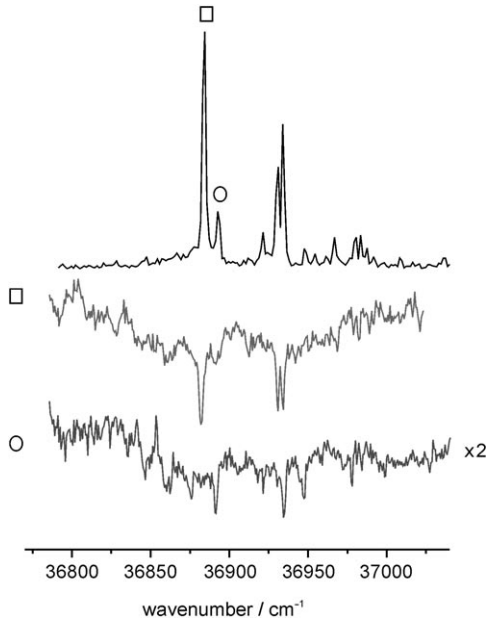


Figure 1. R2PI (upper trace) and UV–UV hole-burn spectra (lower traces) of β-D-PhMan; the symbols □ and ○ identify the vibronic bands selected for each hole-burn spectrum.

free of any interaction with the environment, together with its associated UV–UV hole-burn spectra, a double-resonance pump–probe technique that selectively depletes each selected conformer by transferring population out of its ground electronic state.[11–14] The hole-burn spectra indicate contributions from two distinct components with relative populations of around 4:1. Their corresponding IR–UV infrared ion-depletion (IRID) spectra, shown in Figure 2, establish their association with two alternative conformations, identified through comparison with the spectra computed

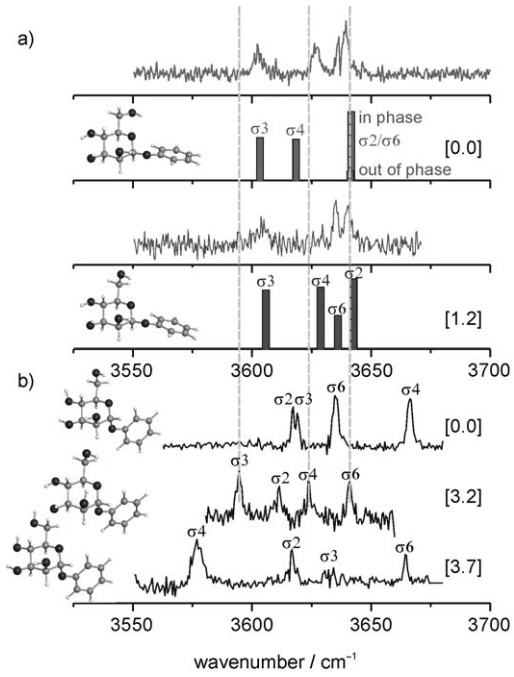


Figure 2. IRID spectra associated with the low-energy conformers of α- and β-D-PhMan (with relative energies (MP2/6-311++G**/ZPE corrected) in kJ mol^{-1} shown in square brackets). a) β-D-PhMan; experimental and calculated (stick spectra) vibrational spectra for the two lowest energy conformers, ccG+g$^-$ [0.0] and ccG−g$^+$ [1.2]; b) experimental vibrational spectra of the three lowest lying conformers of α-D-PhMan (cG−g$^+$) [0.0], (ccG−g$^+$) [3.2] and (cTt) [3.7].[16] The dotted grey lines indicate the positions of the O–H vibrational bands, σ3, σ4 and σ6 in α-D-PhMan (ccG−g$^+$).

through DFT and ab initio calculations as the two lowest energy conformers, ccG+g$^-$ ($E_{rel}=0$ kJ mol^{-1}) and ccG−g$^+$ ($E_{rel}=1.2$ kJ mol^{-1}), respectively; they differ only in the orientation of the hydroxymethyl group. (See the Experimental Section, last paragraph, for a definition of the conformational nomenclature.) Each one presents a counter-clockwise ("cc") hydrogen-bonded chain OH4→OH3→OH2→O1, encouraged by the relatively strong OH3(eq)→OH2(ax) interaction, which shifts the vibrational mode σ3 towards a lower wavenumber, around 3600 cm^{-1}. The two IR spectra of β-D-PhMan closely resemble the spectrum of the counter-clockwise ccG−g$^+$ conformer of the α-anomer, in which σ3

appears at $\approx 3595\ cm^{-1}$, but they clearly differ from those of the lowest or highest energy conformers of the α-anomer, both of which adopt clockwise ("c") conformations.

The structural origin of this difference, as might be expected, is the epimeric C1 centre. In the α-anomer its axial orientation increases the OH2···O1 separation from around 2.3 to around 3.9 Å, and the most strongly shifted vibrational mode in its lowest energy conformation (cG−g$^+$) becomes $\sigma 2 \approx 3618\ cm^{-1}$, reflecting a strong (reversed) hydrogen-bonding interaction, OH2(ax)→OH3(eq). This, together with the re-oriented hydroxymethyl group, facilitates the cooperative "clockwise" sequence OH2→OH3→OH4→OH6. The strong shift of σ2 in the ccG−g$^+$ conformer of α-D-PhMan to around $3610\ cm^{-1}$ (cf. the counter-clockwise ccG−g$^+$ and ccG+g$^-$ conformers of β-D-PhMan, in which $\sigma 2 \approx 3640\ cm^{-1}$) reflects the relative strengths of the OH2→O5 and OH2→O1 interactions in the α- and β-anomers, respectively.

Hydrated mannoside complexes of β-D-PhMan·$(H_2O)_{n=1-3}$: R2PI spectra recorded in the ion mass channels corresponding to β-D-PhMan·$(H_2O)_{n=1-3}$ are presented in Figure 3. One

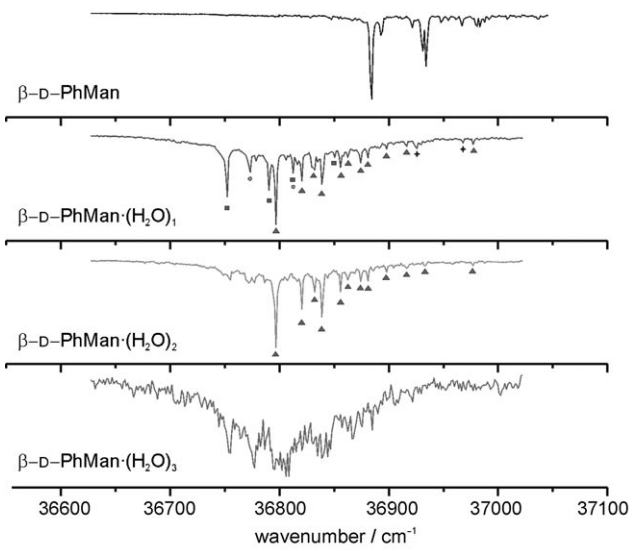


Figure 3. R2PI spectra of β-D-PhMan and its hydrated complexes recorded in the singly, doubly and triply hydrated ion mass channels. The symbols ■, ● and + shown above the vibrational bands in the $n=1$ mass channel identify spectra associated with three distinct monohydrate structures, while the symbol ▲ identifies bands associated with the dihydrate.

major and two minor singly hydrated structures were distinguished through ion-depletion experiments, with relative populations declining in the approximate ratio 10:3:1, but the doubly and triply hydrated complexes populated essentially single structures only; the weak features lying at wavenumbers below $36\,800\ cm^{-1}$ in the dihydrate arise through fragmentation of the trihydrate. Their associated IRID spectra are presented in Figures 4, 5 and 6 (below).

The monohydrate: The experimental vibrational spectrum associated with the most strongly populated monohydrate (shown in Figure 4b) is in excellent accord with the computed spectrum of its global minimum energy conformational structure, cG−g$^+$·ins(4,6), in which ins(4,6) indicates a water

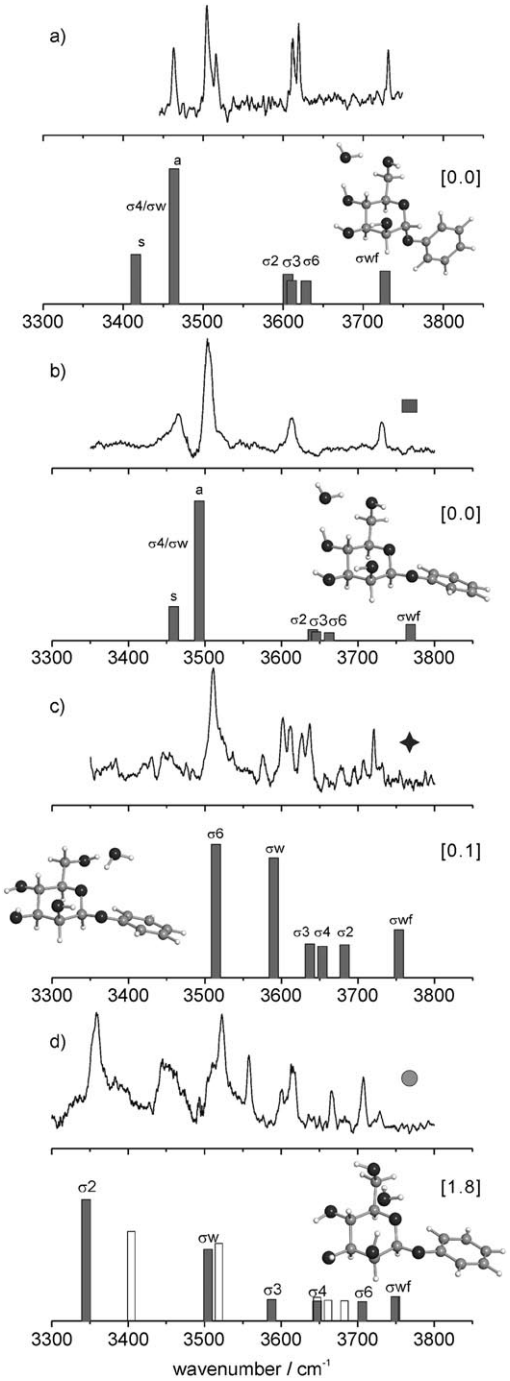


Figure 4. Experimental and computed IRID spectra of singly hydrated α- and β-D-PhMan; relative energies [kJ mol^{-1}] are shown in square brackets. a) α-D-PhMan·$(H_2O)_{n=1}$, cG−g$^+$·ins(4,6). b) β-D-PhMan·$(H_2O)_{n=1}$, cG−g$^+$·ins(4,6). c) β-D-PhMan·$(H_2O)_{n=1}$, ccG+g$^-$·ins(6,5). d) β-PhMan·$(H_2O)_{n=1}$, ccG−t·ins(2,6). The symbols ■, + and ● identify the spectra associated with the three distinct monohydrate structures labelled in the R2PI spectrum.

molecule inserted between OH4 (acting as a hydrogen-bond donor) and OH6 (the acceptor). It is also very similar to the corresponding spectrum of the singly hydrated α-anomer shown in Figure 4a. The strongly displaced doublet at low wavenumber, which is characteristic of an insertion complex,[11–13,16] is associated with the symmetric and anti-symmetric modes involving OH4 and the bound water molecule (W). Its structure requires a conformational switch of the co-operative hydrogen-bonding network from counter-clockwise (ccG+g$^-$) in the lowest energy conformer of the free carbohydrate to clockwise (cG−g$^+$) in the hydrate, OH2→OH3→OH4→*W*→OH6. The switch is facilitated by the rotated orientation of the exocyclic CH_2OH group, the insertion of the water molecule at the OH4–OH6 site, and the reversed OH2(ax)→OH3(eq) interaction. An identical conformational change is also favoured in the monohydrates of the β-anomers of glucose and galactose, β-D-PhGlc and β-D-PhGal.[16] In the α-anomer of mannose, however, in which the OH2(ax)···O1 distance is far greater (≈3.9 cf. ≈2.2 Å) and an OH2→O1 interaction is not possible, the clockwise cG−g$^+$ conformation is already favoured in the *free* carbohydrate. Therefore this conformation, which corresponds to the global minimum, is already set up to accommodate the insertion of a first water molecule at the OH4–OH6 site. Indeed it is the only monohydrate structure that is significantly populated in α-D-PhMan.[11,16]

The more-weakly populated hydrates associated with the IRID spectra (c) and (d) also contain "intruder bands" that appear because of the difficulty in achieving a complete separation of individual conformers when their UV bands overlap or are congested. The "unfilled" stick diagram shown in (d) corresponds to the IR spectrum computed for the intermediate energy structure, cG−g$^+$•ins(6,2) [1.2].

Comparisons between the experimental and computed vibrational spectra of the minor, more weakly populated monohydrate structures of β-D-PhMan (Figure 4c and d), although complicated by the appearance of "intruder" bands, predominantly support their association with the two counter-clockwise structures ccG+g$^-$•ins(6,5) and ccG−t•ins(2,6). The first assignment is favoured by the absence of any features at wavenumbers below that of the intense band at 3510 cm^{-1}, the appearance of the closely spaced cluster of features located between around 3600 and 3650 cm^{-1}, and the "free" OH_W band at ≈3720 cm^{-1}. Assignment of the other hydrate to the structure ccG−t•ins(2,6) is favoured by two distinguishing "fingerprints": the strongly displaced bands at around 3350 and 3510 cm^{-1} and the widely spaced quartet lying between around 3600 and 3730 cm^{-1}. The strong shift of σ2 to around 3350 cm^{-1} reflects the asymmetric disposition of the bound water molecule, which lies closer to OH2 than O6 ($OH2\cdots O_W$ = 1.79 Å, $OH_W\cdots O6$ = 1.89 Å). A possible contribution from a fourth hydrate structure, cG−g$^+$•ins(6,2), lying at a slightly lower energy is also suggested by the correspondence between its computed spectrum (unfilled bars) and some of the intruder bands.

The dihydrate: In marked contrast to the monohydrate, the relative energy of the doubly hydrated complex, β-D-PhMan•$(H_2O)_2$ was calculated to be approximately 15 kJ mol^{-1} below that of its nearest neighbour. Not surprisingly, its IRID spectrum is now free of "interfering" bands, indicating the population of only one predominant structure (Figure 5). All of its eight O–H vibrational bands are clearly

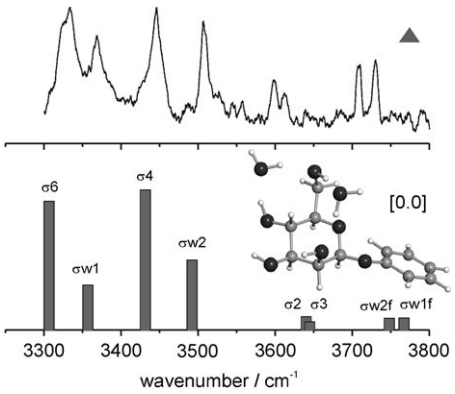


Figure 5. Experimental and computed IRID spectrum of the dihydrate β-D-PhMan•$(H_2O)_{n=2}$, cG−g$^+$•ins(4,6; 6,2). The symbol ▲ identifies the spectrum associated with the dihydrate structure labelled in the R2PI spectrum.

resolved and the experimental spectrum is in remarkably good accord with that computed for the minimum energy structure cG−g$^+$•ins(4,6; 6,2). Like the most favoured monohydrate structure, it again incorporates the cG−g$^+$ conformation, and the two water molecules, inserted separately and not as a water dimer, are located on each side of the hydroxymethyl group. The first water molecule occupies the (4,6) site favoured in the monohydrate and the second one bridges across the pyranose ring to the axial OH2 group to create the extended "virtuous circle" of hydrogen bonding, OH2→OH3→OH4→*W1*→OH6→*W2*→OH2. The powerful role of co-operativity in determining the conformational choice can be gauged by the very large spectral shift of σ6 from around 3640 to around 3300 cm^{-1} and the favoured population of a single structure. Its fragment ion R2PI peaks, recorded in the singly hydrated ion mass channel, were more intense than those associated with parent ions of the singly hydrated complexes (see Figure 3).

The trihydrate: When a third water molecule binds to β-D-PhMan, the structure it adopts is a logical extension of the structures assumed by the singly and doubly hydrated sugar. The first water molecule continues to insert into the OH4–OH6 site while the second *and* the third are inserted between OH6 and OH2, now as a water dimer, to conserve the virtuous cycle of hydrogen bonds adopted in the dihydrate OH2→OH3→OH4→*W1*→OH→*W2*→*W3*→OH2 and retain the cG−g$^+$ conformation of the hydrated monosaccharide. This structure has the lowest relative energy and generates an IR spectrum in good accord with the experimental spectrum shown in Figure 6a, particularly when the

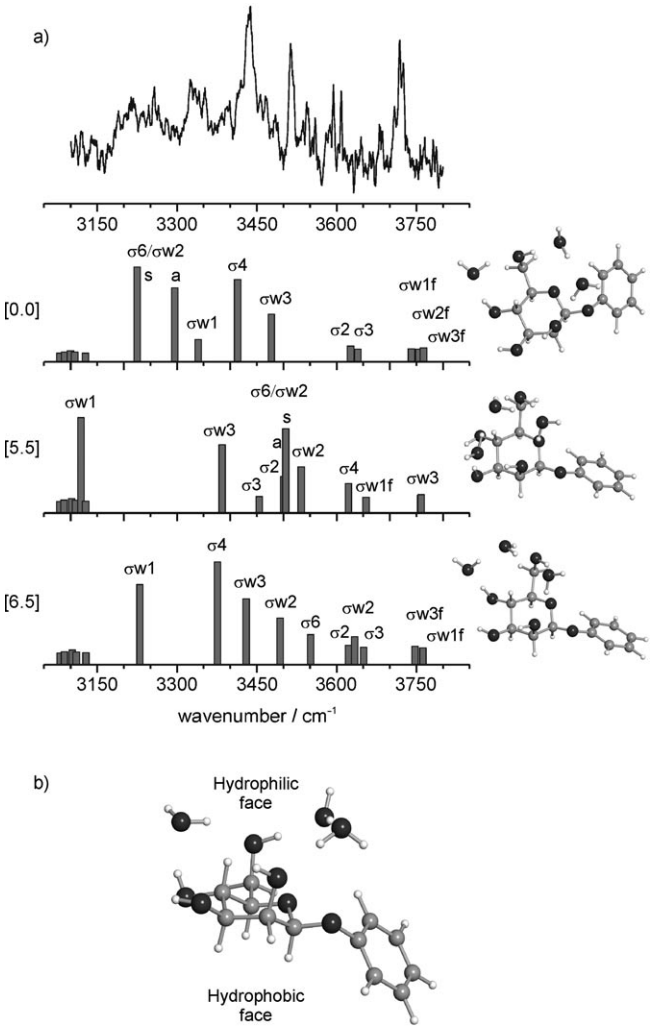


Figure 6. a) Experimental and computed IRID spectra of the trihydrate β-D-PhMan·$(H_2O)_{n=3}$, cG−g$^+$·ins(4,6; (6,2:6,2)). Relative energies [kJ mol^{-1}] are shown in square brackets. b) A side-on view of the trihydrate, revealing the contrast between the non-hydrated (apolar) and hydrated (polar) faces of the carbohydrate.

falling IR laser power at wavenumbers $<3300\,cm^{-1}$ is taken into account, and certainly in much better accord than those associated with the next lowest energy alternatives, located at energies around 5–6 kJ mol^{-1} higher, also shown for comparison. The side-on view of the trihydrate, shown in Figure 6b provides a striking illustration of the amphiphilic character of the carbohydrate. Its hydrophobic, apolar face remains "untouched" and the water molecules are all located exclusively on the hydrophilic, polar face.

α-L-PhFuc·$(H_2O)_2$: The hydroxymethyl group clearly provides the favoured binding site in both the singly and multiply hydrated complexes of β-D-PhMan; the same selectivity also operates in the singly hydrated complexes of β-D-PhGlc, β-D-PhGal and α-D-PhMan.[16] However, when the hydroxymethyl is replaced by a methyl group or a hydrogen atom with no change of relative configuration—for example, in fucose rather than galactose, or in xylose rather than glucose—the water molecule bound in the singly hydrated complex shifts further round the pyranose ring, occupying the (3,2) site in both α- and β-L-PhFuc,[13] the (3,2) or (4,3) sites in α-D-PhXyl, and the (2,1) site in β-D-PhXyl,[17] whereas the counter-clockwise hydrogen-bonding orientation in the host carbohydrate is retained. This continues to be the case in doubly hydrated α-L-PhFuc, but rather than binding the two water molecules individually (on each side of the hydroxymethyl group) as in β-D-PhMan, in the absence of a hydroxymethyl group they now prefer to bind as a water dimer, again inserting at the (3,2) site, to create the counter-clockwise hydrogen-bonded chain OH4→OH3→*W1*→*W2*→OH2. The computed structures and relative energies of the two lowest energy structures of doubly hydrated phenyl α-L-fucopyranoside are shown in Figure 7.

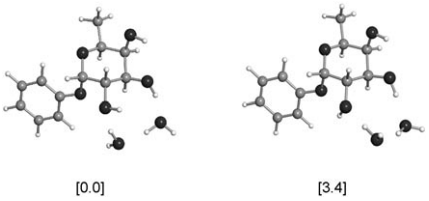


Figure 7. The computed (MP2/6-311++G**//B3LYP/6-311+G*) lowest energy structures of the dihydrate α-Ph-L-Fuc$(H_2O)_2$; their relative energies in kJ mol^{-1} are shown in square brackets. Each one binds a water dimer at the (3,2) site; there is also a structure at an intermediate energy that differs only in the orientation of the phenyl "tag". The L enantiomer was selected because of its natural occurrence (cf. the D enantiomers of Glc, Gal and Man).

Monosaccharide hydration, the working rules: In Glc, Gal and Man, the hydroxymethyl group always provides the preferred hydration site. Water binding also greatly enhances the relative stability of the clockwise conformation (cG−g$^+$ in the α- and β-anomers of glucose and in the β-anomers of galactose and mannose), and although structures retaining the original counter-clockwise ccG+g$^-$ carbohydrate conformation can be populated, the singly hydrated cG−g$^+$ structure either becomes one of the preferred conformations or *the* preferred conformation (see Table 1). This selectivity and conformational preference is displayed a fortiori in the doubly and triply hydrated complexes of β-D-PhMan. In the α-anomer of mannose, in which the axial orientation of OH2 increases the OH2···O1 distance to around 3.9 Å (cf. ≈2.2 Å in α-D-MeGlc, α-D-MeGal and β-D-PhMan or ≈2.5 Å in β-D-PhGlc and β-D-PhGal), there is no longer any possibility of a (counter-clockwise) OH2→O1 interaction. The clockwise cG−g$^+$ conformation (and binding site selectivity) is adopted both in its singly hydrated complex, α-D-PhMan (cG−g$^+$·ins4,6), *and* in the isolated carbohydrate.[12] Indeed, its hydration greatly reduces the relative stability of the ccG+g$^-$ conformation (see Table 1).

The behaviour of the α-anomer of galactose appears at first sight to be anomalous, since the hydrated complex of α-D-MeGal prefers to retain its ccG+g$^-$ conformation, though, like the hydrated β-anomer, the water molecule still inserts

Table 1. Calculated relative energies ($\Delta E(0)$ in kJ mol^{-1}) of the lowest lying cG−g$^+$ and ccG+g$^-$ conformers of unhydrated (W0) and singly hydrated (W1) complexes of the α- and β-anomers of glucose, galactose and mannose; energies determined through single-point calculations conducted at the MP2/6-311++G** level, based upon DFT structures (B3LYP/6-31+G* and B3LYP/6-311+G*). Numbers in brackets indicate their relative energy ordering.

	α-MeGlc		α-MeGal		α-PhMan	
	W0	W1	W0	W1	W0	W1
cG−g$^+$	>10	0.0 (1) ins4,6	7.1 (4)	6.9 (15) ins4,6	0.0 (1)	0.0 (1) ins4,6
ccG+g$^-$	0.4 (2)	3.9 (3) ins3,2	0.0 (1)	0.0 (1) ins6,5	5.7 (5)	>19 ins2,1
	β-PhGlc		β-PhGal		β-PhMan	
	W0	W1	W0	W1	W0	W1
cG−g$^+$	9.7 (4)	0.0 (1) ins4,6	6.6 (4)	0.3 (2) ins6,5	12.0 (13)	0.0 (1) ins4,6
ccG+g$^-$	0.0 (1)	0.9 (2) ins6,5	0.0 (1)	0.0 (1) ins6,5	0.0 (1)	0.1 (2) ins6,5

at the (6,5) site. The explanation lies in the axial orientation of OH4, which enhances the OH4(ax)→OH3(eq) interaction. This, together with the enhanced OH2(eq)→O1(ax) interaction in the α-anomer, favours retention of the "cc" chain, OH4→OH3→OH2→O1. In addition, retention of the G+g$^-$ orientation favours the ins(6,5) water binding motif OH6→*W*→O5. In contrast, in α-D-MeGlc, in which OH4 is oriented equatorially, the conformational switch from ccG+g$^-$ to cG−g$^+$, which removes the OH2→O1 interaction, is compensated by the insertion at the (4,6) site to create the alternative strongly bound motif OH2→OH3→OH4→*W*→O6.

The full set of preferred Glc, Gal and Man monohydrate structures is displayed in Figure 8. It includes those identified experimentally through a combination of vibrational spectroscopy coupled with DFT and ab initio calculations (β-D-PhGlc, β-D-PhGal and α/β-D-PhMan) and those predicted theoretically (α-D-MeGlc and α-D-MeGal). A clear pattern of selectivity and conformational preference is revealed. When the OH4 group is oriented equatorially, as in Glc and Man, the cG−g$^+$ conformation creates an optimised structure for reception of the bound water molecule at the (4,6) hydroxymethyl site—the weakest link in the original ccG+g$^-$ configuration—to maximise the co-operative hydrogen bonding. When the OH4 group is oriented axially, as in Gal, this is achieved by insertion of the water molecule at the (6,5) site.

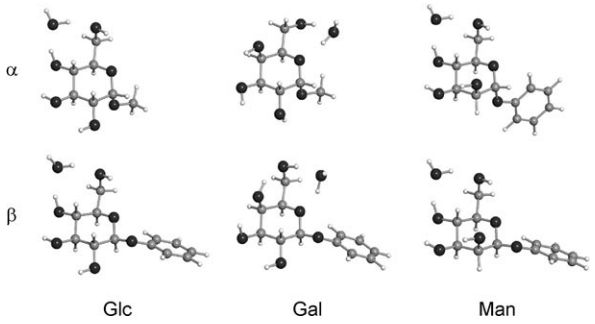


Figure 8. The preferred structures of the singly hydrated complexes of α- and β-anomers of Glc, Gal and Man.

These working rules continue to operate in multiply hydrated β-D-PhMan. In the dihydrate the second molecule binds on the opposite side of the hydroxymethyl group, at the (6,2) site, and in the trihydrate the (6,2) site is occupied by a water dimer.

Carbohydrate–protein binding: Given the propensity rules generated by optimisation of hydrogen-bonding networks in hydrated carbohydrate complexes, it is not unreasonable to anticipate the possible operation of similar rules in hydrogen-bonded protein–carbohydrate complexes. Indeed, an initial survey of published X-ray crystal structures of simple monosaccharide α-D-mannoside ligands[16] identified the O4–O6 binding site as a favoured, though not exclusive recognition point. The hydroxymethyl group also acts as a focal point in many other monosaccharide ligands, which bind to the protein at their O4–O6 or O6–O5 sites. Examples include: α-D-MeGlc bound at O6–O5 to Concanavalin A,[22] pea lectin,[23] *Lathyrus Ochrus* isolectin I,[24] *Pterocarpus angolensis* lectin[25] and to Jacalin;[26] α-D-MeGal bound at O6–O4 to Jacalin,[26] *Artocarpus hirsuta*[27] and to Jacalin;[28] and β-D-MeGal bound at O6–O5 to C-type animal lectins,[29] and peanut lectin.[30] When explicitly bound water molecules are conserved in protein–carbohydrate complexes, it may be because their presence maintains a favoured ligand conformation as well as providing a stronger overall protein–carbohydrate interaction.[31] When there are no conserved water molecules, the interaction between the protein and the ligand and between the displaced and surrounding water molecules will be the deciding factor. Finally, the clear distinction between the hydrophilic and hydrophobic faces displayed in the multiply hydrated carbohydrate β-D-PhMan·$(H_2O)_3$ raises interesting issues relating to molecular recognition. The hydrophobic face is "ready" for stacking interactions with aromatic residues[32,33] but the hydrophilic face is covered by water molecules with a structure directed by the conformation of the underlying carbohydrate, which may itself be altered by hydration. Which structure will the protein recognize?

Conclusion

The factors that govern selectivity and conformational choice in hydrated carbohydrates have been examined by determining the structures of a systematically varied set of singly and multiply hydrated monosaccharide complexes in the gas phase by using ultraviolet and infrared ion depletion spectroscopy conducted in a supersonic jet expansion and coupled with DFT and ab initio calculations. New results obtained for the singly hydrated complexes of phenyl β-D-man-

nopyranoside (β-D-PhMan), methyl α-D-gluco- and α-D-galactopyranoside (α-D-MeGlc and α-D-MeGal), when coupled with those reported earlier for the singly hydrated complexes of α-D-PhMan, β-D-PhGlc and β-D-PhGal, have created a comprehensive data set that reveals a systematic pattern of conformational preference and regioselectivity, driven by the provision of optimal, co-operative hydrogen-bonded networks in the hydrated sugars. In addition, despite the very large number of potential structures and binding sites, the choice is highly selective with binding invariably "focused" around the hydroxymethyl group. A survey of some of the structures listed in the protein databank suggests the possible operation of similar rules governing the structures of hydrogen-bonded protein–carbohydrate complexes.

Similar rules also continue to operate in the multiply hydrated carbohydrate β-D-PhMan·$(H_2O)_{2,3}$, in which the bound water molecules form a cyclic structure that incorporates the hydroxymethyl group. They are located exclusively on the polar face and their orientation is dictated by the (perturbed) conformation of the carbohydrate to which they are attached.

Experimental Section

Spectroscopy: Detailed descriptions of the experimental strategy have been published previously.[13] The carbohydrate samples were vaporised into a supersonic jet argon expansion using a laser desorption system constructed in-house, and their hydrated complexes were formed by seeding the argon carrier gas with water vapour (≈0.25%) and then stabilised in the high collision frequency region of the free jet expansion. The expanding jet passed through a 2 mm skimmer to form a collimated molecular beam which was crossed by one or two tunable laser beams in the extraction region of a linear time-of-flight mass spectrometer (Jordan). Mass-selected resonant two-photon ionization (R2PI) spectra were recorded using the frequency-doubled output of a pulsed Nd:YAG-pumped dye laser (Spectron 810/LambdaPhysik FL2002, 0.5 mJ pulse^{-1} UV for the monosaccharides and Continuum Powerlite II/Sirah PS-G, 3 mJ pulse^{-1} UV for their hydrated complexes, both operating at 10 Hz). Conformer-specific UV and IR spectra were recorded by employing UV–UV and IR ion-depletion (IRID) double-resonance spectroscopy. The UV–UV experiments utilised the frequency-doubled output of the pulsed Nd:YAG pumped dye lasers described above. The IR spectroscopy experiments employed radiation in the range 3100–3800 cm^{-1}, generated by difference frequency mixing of the fundamental of a Nd:YAG laser with the output of a dye laser in a $LiNbO_3$ crystal (Continuum Powerlite 8010/ND6000/IRP module) or by a tunable OPO/OPA laser system (LaserVision); all laser pulses were of an approximately 10 ns duration. The delay between the pump and the probe laser pulses was approximately 150 ns in both the IR ion-depletion and UV hole-burning experiments, whereas the delay time between opening the pulsed valve and triggering the UV laser was adjusted to optimise the generation of either "naked" or hydrated molecules in the supersonic expansion.

Computation: Conformational and structural assignments were made through comparison with calculation by using a combination of molecular mechanics, ab initio and density functional theory (DFT) methods as implemented in the MacroModel software (MacroModel v.8.5, Schrödinger, LLC21),[34] and the Gaussian 03 program package.[35] Initial structures were generated by a molecular mechanics conformational search using the Monte Carlo multiple minimisation method. A much-reduced set of conformational or hydrated complex structures was selected for DFT geometry optimisation (B3LYP/6-31+G* for the naked and B3LYP/6-311+G* for the hydrated systems) based upon their relative energies (with a cut-off at ≈15 kJ mol^{-1}) and comparisons with previous investigations of carbohydrate structures. More accurate energies were then calculated for these optimised structures at the MP2/6-311++G** level of theory. Zero-point corrections and harmonic vibrational frequencies for the energetically lowest lying conformers were obtained using the B3LYP structures; the frequencies computed for the O–H stretch modes (expressed in wavenumbers) were scaled by the factor 0.9734 for comparison with the experimental spectra. The full list of energies and structures for the lowest energy conformers is given as Supporting Information.

Notation: The nomenclature of the hydrated and unhydrated monosaccharide conformations has been described previously.[13] The designations, ccG+g$^-$, cTt, and so forth, indicate the conformation of the carbohydrate: briefly, "cc" ("c") indicates a counter-clockwise (clockwise) orientation of the peripheral OH groups, OH4→OH3→OH2→O1 (O1→OH2→OH3→OH4), and G+g$^-$ (Tt) indicates the gauche (*trans*) orientation of the exocyclic hydroxymethyl group and its terminal OH6 group, respectively. In the case of the hydrated structures the insertion position of the water molecule is indicated by adding "ins(position)" to the nomenclature of the bare molecule, for example, ins(4,6) indicates a water molecule inserted between OH4 (acting as a hydrogen-bond donor) and OH6 (the acceptor).

Acknowledgements

The authors thank Dr. Jann Frey, Dr. Lavina C. Snoek, Ms. Yoana Perez Badel and Prof. Luis A. Montero for their contributions to the studies presented here. We acknowledge the financial support provided by Engineering and Physical Sciences Research Council (EPSRC), the Leverhulme Trust (grant F/08788G), the Spanish Ministry of Education and Science (EJC), the Oxford Supercomputing Centre, the STFC Laser Loan Pool and the Physical and Theoretical Chemistry Laboratory.

[1] H.-J. Gabius, H.-C. Siebert, S. André, J. Jiménez-Barbero, H. Rüdiger, *ChemBioChem* **2004**, *5*, 740–764.
[2] C. Clarke, R. J. Woods, J. Gluska, A. Cooper, M. A. Nutley, G. J. Boons, *J. Am. Chem. Soc.* **2001**, *123*, 12238–12247.
[3] A. Laederach, P. J. Reilly, *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 591–597.
[4] A. J. Petrescu, S. M. Petrescu, R. A. Dwek, M. R. Wormald, *Glycobiology* **1999**, *9*, 343–352, and references therein.
[5] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, *J. Mol. Biol.* **1977**, *112*, 535–542; http://www.rcsb.org/pdb/home/home.do.
[6] F. H. Allen, O. Kennard, *Chemical Design Automation News*; **1993**, *8*, 31–37; http://www.ccdc.cam.ac.uk/products/csd/.
[7] M. R. Wormald, A. J. Petrescu, Y.-L. Pao, A. Glithero, T. Elliott, R. A. Dwek, *Chem. Rev.* **2002**, *102*, 371–386.
[8] V. Kräutler, M. Müller, P. H. Hünenberger, *Carbohydr. Res.* **2007**, *342*, 2097–2124.
[9] O. Coskuner, *J. Chem. Phys.* **2007**, *127*, 015101.
[10] L. Hemmingsen, D. E. Madsen, A. L. Esbensen, L. Olsen, S. B. Engelsen, *Carbohydr. Res.* **2004**, *339*, 937–948.
[11] J. P. Simons, R. A. Jockusch, P. Çarçabal, I. Hünig, R. T. Kroemer, N. A. Macleod, L. C. Snoek, *Int. Rev. Phys. Chem.* **2005**, *24*, 489.
[12] P. Çarçabal, I. Hünig, D. P. Gamblin, B. Liu, R. A. Jockusch, R. T. Kroemer, L. C. Snoek, A. J. Fairbanks, B. G. Davis, J. P. Simons, *J. Am. Chem. Soc.* **2006**, *128*, 1976.
[13] P. Çarçabal, T. Patsias, I. Hünig, B. Liu, C. Kaposta, L. C. Snoek, D. P. Gamblin, B. G. Davis, J. P. Simons, *Phys. Chem. Chem. Phys.* **2006**, *8*, 129.
[14] Carbohydrate structural library: http://physchem.ox.ac.uk/~jps.
[15] E. G. Robertson, J. P. Simons, *Phys. Chem. Chem. Phys.* **2001**, *3*, 1–18.

[16] P. Çarçabal, R. A. Jockusch, I. Hünig, L. C. Snoek, R. T. Kroemer, B. G. Davis, D. P. Gamblin, I. Compagnon, J. Oomens, J. P. Simons, *J. Am. Chem. Soc.* **2005**, *127*, 11414.
[17] I. Hünig, A. J. Painter, R. A. Jockusch, P. Çarçabal, E. M. Marzluff, L. C. Snoek, D. P. Gamblin, B. G. Davis, J. P. Simons, *Phys. Chem. Chem. Phys.* **2005**, *7*, 2474.
[18] O. Mitsunobu, Y. Yamada, *Bull. Chem. Soc. Jpn.* **1967**, *40*, 2380–2382.
[19] O. Mitsunobu, *Synthesis* **1981**, 1–28.
[20] M. Mella, L. Panza, F. Ronchetti, L. Toma, *Tetrahedron* **1988**, *44*, 1673–1678.
[21] J. J. Gridley, H. M. I. Osborn, *J. Chem. Soc. Perkin Trans. 1* **2000**, 1471–1491.
[22] S. J. Harrop, J. R. Helliwell, T. C. Wan, A. J. Kalb, L. Tong, J. Yariv, *Acta Crystallogr. Sect. A* **1996**, *52*, 143–155; (PDB ID = 1gic).
[23] M. B. Shevtsov, I. N. Tsygannik, to be published (PDB ID = 1hkd).
[24] Y. Bourne, A. Roussel, M. Frey, P. Rougé, J. C. Fontecilla-Camps, C. Cambillau, *Proteins: Struct., Funct., Bioinf.* **1990**, *8*, 365–376; (PDB ID = 1 Loa).
[25] R. Loris, A. Imberty, S. Beeckmans, E. V. Driessche, J. S. Read, J. Bouckaert, H. D. Greve, L. Buts, L. Wyns, *J. Biol. Chem.* **2003**, *278*, 16297–16303; (PDB ID = 1n3o).
[26] A. A. Jeyaprakash, G. Jayashree, S. K. Mahanta, C. P. Swaminathan, K. Sekar, A. Surolia, M. Vijayan, *J. Mol. Biol.* **2005**, *347*, 181–188; (PDB ID = 1s4).
[27] K. N. Rao, C. G. Suresh, U. V. Katre, S. M. Gaikwad, M. I. Khan, *Acta Crystallogr. Sect. A* **2004**, *60*, 1404–1412; (PDB ID = 1tp8).
[28] R. Sankaranarayanan, K. Sekar, R. Banerjee, V. Sharma, A. Surolia, M. Vijayan, *Nat. Struct. Biol.* **1996**, *3*, 596–603; (PDB ID = 1jac).
[29] A. R. Kolatkar, W. I. Weis, *J. Biol. Chem.* **1996**, *271*, 6679–6685; (PDB ID = 1afa).
[30] R. Ravishankar, K. Suguna, A. Surolia, M. Vijayan, *Acta Crystallogr. Sect. A* **1999**, *55*, 1375–1382; (PDB ID = 1qf3).
[31] S. Elgavish, B. Shaanan, *J. Mol. Biol.* **1998**, *277*, 917–932.
[32] J. Screen, E. C. Stanca-Kaposta, D. P. Gamblin, B. Liu, N. A. Macleod, L. C. Snoek, B. G. Davis, J. P. Simons, *Angew. Chem.* **2007**, *119*, 3718–3722; *Angew. Chem. Int. Ed.* **2007**, *46*, 3644–3648.
[33] E. C. Stanca-Kaposta, D. P. Gamblin, J. Screen, B. Liu, L. C. Snoek, B. G. Davis, J. P. Simons, *Phys. Chem. Chem. Phys.* **2007**, *9*, 4444–4451.
[34] F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrikson, W. C. Still, *J. Comput. Chem.* **1990**, *11*, 440–467.
[35] Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Wallingford CT, **2004**.

Received: March 14, 2008
Published online: August 20, 2008

