Modeling the tyrosine–sugar interactions in supersonic expansions: glucopyranose–phenol clusters†

*Imanol Usabiaga, Jorge González, Pedro F. Arnáiz, Iker León, Emilio J. Cocinero and José A. Fernández**

Dep. of Physical Chemistry, Fac. of Science and Technology, University of the Basque Country (UPV/EHU), Barrio Sarriena, s/n, 48940 Leioa, Spain

*Corresponding Author:

José A. Fernández, josea.fernandez@ehu.es

pone: +34946015387

fax: +34946013500

web: https://sites.google.com/site/gesemupv/

Abstract

Sugars are fundamental building blocks in living beings and their interaction with proteins play a central role in fundamental processes, such as energy storage and production, post-transductional modifications or immune response. Understanding such processes requires of a deep knowledge of the interaction mechanism at molecular level. Here we explore the interaction between α/β-Methyl-D-glucopyranose and β-Phenyl-D-glucopyranose with phenol, the residue of tyrosine, using a combination of mass-resolved laser electronic spectroscopy in jets and quantum mechanical calculations. The observed structures of different complexes are stabilized by a subtle equilibrium between several interactions, among which the weak dispersion forces may tilt the balance. Thus, the small structural changes introduced by the orientation of the anomeric substituent are amplified by the interaction with phenol, resulting in different number of detected isomers and in changes in the preferred solvation site. Furthermore, inclusion of the entropy for the calculated structure is advisable to understand the energetic reasons for the detection of a small set of experimental isomers.

INTRODUCTION

Despite the enormous variety of living beings on Earth, all of them share some common characteristics, inherited from the first unicellular organism inhabiting the planet. For example, all living organisms seem to base their metabolism in the existence of four main groups of biomolecules: amino acids, lipids, DNA bases and carbohydrates. These building blocks may be found isolated, playing roles as chemical messengers or may be forming superstructures (proteins, DNA, lipid membranes, glycoproteins… etc).[1]

Regarding the carbohydrates, they play a central role in energy storage and consumption, but also in immunity: the cell membrane is ornamented by glycans, polysaccharides attached to proteins and lipids, which are the cellular I.D. used by the immune system to classify the cells as friends or foes.[2, 3] The immune cells present numerous receptors that probe the identity of such glycans; how the polysaccharide accommodates in the receptor decides if an immune response is elicited or not.[4, 5] Such accommodation is not only determined by the sequence and nature of the component sugar units, but also by the ramifications and the kind of sugar-sugar connections, because the O-glicosidic linkages are very versatil: α/β 1-2, 1-3, 1-4… etc.[6, 7] In some cases, orientation of a simple hydroxyl substituent in a complex sequence of sugars may makes the difference, like for example in the case of the blood groups.[8] The carbohydrate-protein molecular recognition can also be associated to the presence of water molecules and ions, particularly Ca^{2+}. These "partners" not mere spectators but form part of the structure that is recognize by the protein.

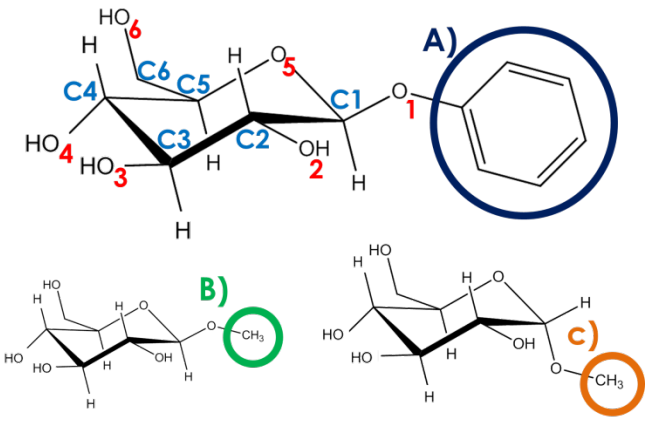
Thus, the molecular mechanism underlying the recognition process is complex and it has been subject of several studies in solution (using mainly NMR[9]) and gas phase (using laser spectroscopy [10, 11]). Among the latter, it is worthy to mention the pioneering work by Simons' group, which has mapped the conformational landscape of several mono-, di- and poly-saccharides using a combination of mass-resolved laser spectroscopy and quantum mechanical calculations.[12-16] From those studies the complexity of these systems becomes clear: except for the monosaccharides and a few disaccharides, they present unresolved electronic spectra that yield little information on their structure. Furthermore, the number of OH groups in each monomeric unit, which in addition are usually involved in cooperative hydrogen bond networks, results in congested spectra in the region of the OH/NH stretches, complicating the extraction of structural information from the IR/UV spectra, unless they are accompanied by high level quantum mechanical calculations.[14] A last complication is introduced by the high melting point and low vapor pressure of the sugars that makes necessary the use of laser desorption systems in their study, introducing an additional level of experimental complexity. To circumvent part of these difficulties, Simons's group demonstrated that it is possible to attach a chromophore to the

anomeric carbon of a sugar to obtain physical data on the sugar using mass-resolved laser spectroscopy. This addition does not introduce appreciable changes in the conformational landscape and it has, in addition, the advantage of blocking the mutarotation that interconverts the anomers (α or β) and removes the interactions with the OH1. This position is blocked in oligosaccharides in most cases, and therefore the results from these systems are easily compared with the interactions in oligossaccharides.[17] Finally, our group was able to study a pure sugar (without chromophore tag) via microwave spectroscopy thanks to UV ultrafast laser vaporization,[17] and later, we characterized the structure of monosaccharide for first time thanks to enriched isotopic samples.[18]

From the computational point of view the systems are not easy to tackle either: sugars usually present several possible conformations with very close stability and therefore the computational level must be carefully chosen to obtain quantitative results. Regarding the prediction of the IR spectra, small shifts in the OH stretching vibrations may result in very different assignments of the experimental data, and therefore the computational level must be carefully chosen to obtain quantitative results at an affordable computational cost.

Here we present the characterization of the interaction between glucopyranose derivatives and the residue of tyrosine (scheme 1), as a model system for the amino acid-sugar interaction. The methyl group locks the sugar in one of its two anomeric forms, which otherwise could interconvert and will allow us to explore the influence of the anomeric conformation in the intermolecular interactions. On the other hand, placing a phenyl ring as substituent will allow us to explore the competition between the sugar unit and the aromatic ring for the interaction with phenol.

A related system was previously studied by Stanca-Kaposta et al.[19] However, the inherent difficulties of the study resulted in a low S/N ratio, precluding a solid assignment of the experimentally detected isomers. The slight differences between the interacting molecules in their study and those presented here resulted in a significant improvement in S/N ratio, which allowed us to record clean, well-resolved spectra. Comparison of these results with predictions from accurate quantum mechanical calculations, yielded well-sustained assignments for the experimental spectra. Also, we introduce the concept of Gibbs binding free energy as a useful magnitude for the analysis of the energetics of the conformational landscape of complex systems. This theoretical approach may help tackling other systems of increasing complexity that have so far lied out of reach.



Scheme 1. A) Phenyl-β-D-Glucpyranose B) Menthyl-β-D-Glucpyranose C) Menthyl-α-D-Glucpyranose. Atom numbering (oxygen in red; Carbon in blue) in the three sugar derivatives studied in this work.

METHODS

Experimental

The experimental system is in essence similar to what was published in ref [20] and consists of a modified time of flight mass spectrometer equipped with a laser desorption/ionization (LDI) source attached to a pulsed valve (General Valve Inc.), three Nd/YAG-pumped dye lasers (Scanmate, Lambda Physic and Fine Adjustment Pulsare), an OPO system (LaserVision) to generate light in the IR region and electronics for synchronization and data collection and handling. The electronic spectroscopy of the systems was carried out using resonance-enhanced multiphoton ionization (REMPI, 1+1 photons), although most of the species present broad absorptions. Modification of the experimental set up to record the 2-color REMPI spectra did not result in an improvement and therefore, all the electronic spectra presented in this work were recorded using a 1+1 scheme.

Structural information on the systems studied was obtained from IR/UV double resonance experiments, recorded using the OPO system and one of the dye lasers. In those cases in which the broad emission of the OPO did not allow to discriminate close IR transitions, the Fine Adjustment Pulsare dye laser (~0.2 cm^{-1}, seeded ~0.001 cm^{-1} FWHM) was used instead.

Calculations

A critical aspect of this study was to ensure that the conformational landscape was thoroughly explored, as small, experimentally non-significant changes in the relative orientation of some chemical groups may result in shifts in the IR bands that may derive in incorrect assignments of the experimental spectra. Therefore, a first, automated exploration was carried out using molecular mechanics and two search algorithms: the "Large scales Low Mode" (which uses frequency modes to create new structures) and a Monte Carlo-based search, as implemented in Macromodel (www.schrodinger.com). In a second stage, the structures were inspected using chemical intuition, looking for alternatives which implicate small rotations of functional groups or of one of the molecules respect to the other. The whole group of candidate (~500 structures, 30kJ/mol) structures found using both methods. [clustering por similitudes se reduce a ~60] was subjected to full optimization at M06-2X/6-31+G(d). Then, those structures lying in a reasonable stability window of 30 kJ/mol were subjected to further re-optimization (de las 10 mas estables) at the M06-2X/6-311++G(d,p) calculation level as implemented in Gaussian 09,[21] which has proved to yield accurate results for systems of similar complexity.[22] A final normal mode analysis highlighted the validity of the optimized structures as true minima and allowed us to compute the zero point energy (ZPE). Thus, the energy values given in this work include the ZPE correction. This approach has been successfully used in systems of similar or even higher degree of complexity.[23-25] The 20 most stable structures for each system optimized at M06-2X/6-31+G(d) level may be found in the supporting information (Figures S10-S12)

Regarding the entropy and the Gibbs free energy, it was calculated using the output from the Gaussian calculations and the tools supplied by the NIST (http://www.nist.gov/mml/csd/informatics_research/thermochemistry_script.cfm), using the following equations[26]:

$$\Delta G_{monomers} = E_{Elec.} + E_{ZPE} + E_{ddH} - T \cdot \Delta S_{dS} \quad (1)$$

$$\Delta G_{Complex} = E_{Elec.} + E_{ZPE} + E_{BSSE} + E_{ddH} - T \cdot \Delta S_{dS} \quad (2)$$

$$\Delta G_B = \Delta G_{Complex} - \sum \Delta G_{monomer} \quad (3)$$

$$\Delta G_R = \Delta G^i_{Complex} - \Delta G_{min} \quad (4)$$

$$\Delta G^i_{RB} = [\Delta G^i_{Complex} - \sum \Delta G_{monomers}] - [\Delta G^{Min}_{Complex} - \sum \Delta G_{monomers}]$$

$$= \Delta G^i_{\mathrm{B}} - \Delta G^{Min}_{\mathrm{B}} \quad (5)$$

Equation (1) was used to obtain the Gibbs free energy for a given species using the stability of the species ($E_{Elec.}$), its zero-point energy ($E_{Elec.}$), the change in enthalpy with temperature (E_{ddH}), the temperature and the entropy (ΔS_{dS}). The two latter magnitudes were calculated using the tools at NIST web site. In the case of a complex (equation 2), it is necessary to include also de BSSE correction (E_{BSSE}) estimated using the counterpoise procedure.[27] Thus the change in Gibbs free energy due to cluster formation, ΔG_B, was obtained using (3), as the difference between the Gibbs free energy of the complex ($\Delta G_{Complex}$) and the constituent monomeric units ($\Delta G_{Monomer}$). This value allowed us to estimate the maximum temperature at which the species are stable. Then, the relative Gibbs free energy of binding of the conformational isomer *i* of a complex, ΔG_{RB}^{i} , may be defined as the difference between the binding free energy of the complex, ΔG_{B}^{i}, and that of the global minimum, ΔG_{B}^{Min}equation 5. Nevertheless, this value is the same as ΔG_R.

RESULTS

One of the difficulties of exploring the aggregation of sugars is their ability to adopt several conformations of very similar stability, because the flexibility their flexible OH groups. For example, α/β-Methyl-D-glucopyranose (α/β-MeGlc, Figure S1 of the supporting information) may adopt three conformations depending on the orientation of the hydroxymethyl flexible arm respect to the rest of the molecule, with ΔG_R of less than ~ 3 kJ/mol in the temperature window of our supersonic expansion. Thus, in principle one would expect a co-existence of several conformers of each molecular aggregate in the expansion, a fact that could seriously complicate the study.

Figure 1 collects the REMPI (resonance enhanced multiphoton ionization) spectra of phenol, β-Phenyl-D-glucopyranose (β-PhGlc), α-MeGlc···phenol (α-MeGlc·P), β-MeGlc···phenol (β-MeGlc·P) and β-PhGlc···phenol (α-PhGlc·P). As can be seen, the 0_0^0 transition of phenol appears at 36349 cm^{-1}, in good agreement with the data in the literature. The spectrum shows several hot bands (for example at 36273 and 36293 cm^{-1}), due to a non-optimal cooling. Actually, phenol was not introduced in the sample carrier, but produced by dissociation of β-PhGlc during the laser desorption process, and therefore the hot bands indicate that phenol appears with a significant amount of vibrational energy that is not completely cooled by the expansion. An enlarged version of Figure 1 may be found in the supplemental information (Figure S2), together with a detailed analysis of all the spectral features. IR/UV hole burnings of all the species studied in this work may be also found in the supporting information (Figures S4-S8)

The spectrum of β-PhGlc was already reported by F. O. Talbot and J. P. Simons,[28] who demonstrated the existence of three conformational isomers with 0_0^0 transitions at 36868, 36880 and 36903 cm^{-1} and that correspond to different orientations of the hydroxymethyl moiety connected to carbon 6 (See Figure S3 of the supplemental information).[29] The overall spectrum in Figure 1 presents several well-resolved transitions and it is in good accord with previous publications. By contrast, the spectra of α- and β-PhGlc·P are broad absorptions starting at ~ 36250 cm^{-1}; i.e., shifted to the red of the phenol 0_0^0 transition. Both spectra are very similar, although that of α-MeGlc·P exhibits several discrete features. Above 36700 cm^{-1} both spectra contain important contribution of fragmentation from higher-order clusters, and therefore, such region was avoided during recording of the IR/UV spectra.

Finally, the spectrum of β-PhGlc·P is also a broad absorption starting at ~36250 cm^{-1}. It gains intensity to the blue, presenting two strong peaks at 36892 and 36912 cm^{-1}. It is worthy to note that the complex has two chromophores: phenol and the β-PhGlc, although in these kind of

complexes the chromophore with the less energetic electronic transition is excited; depending on the cluster's structure, an excitation transfer from the chromophore with the bluish transition to the other one may exist.[25, 30]

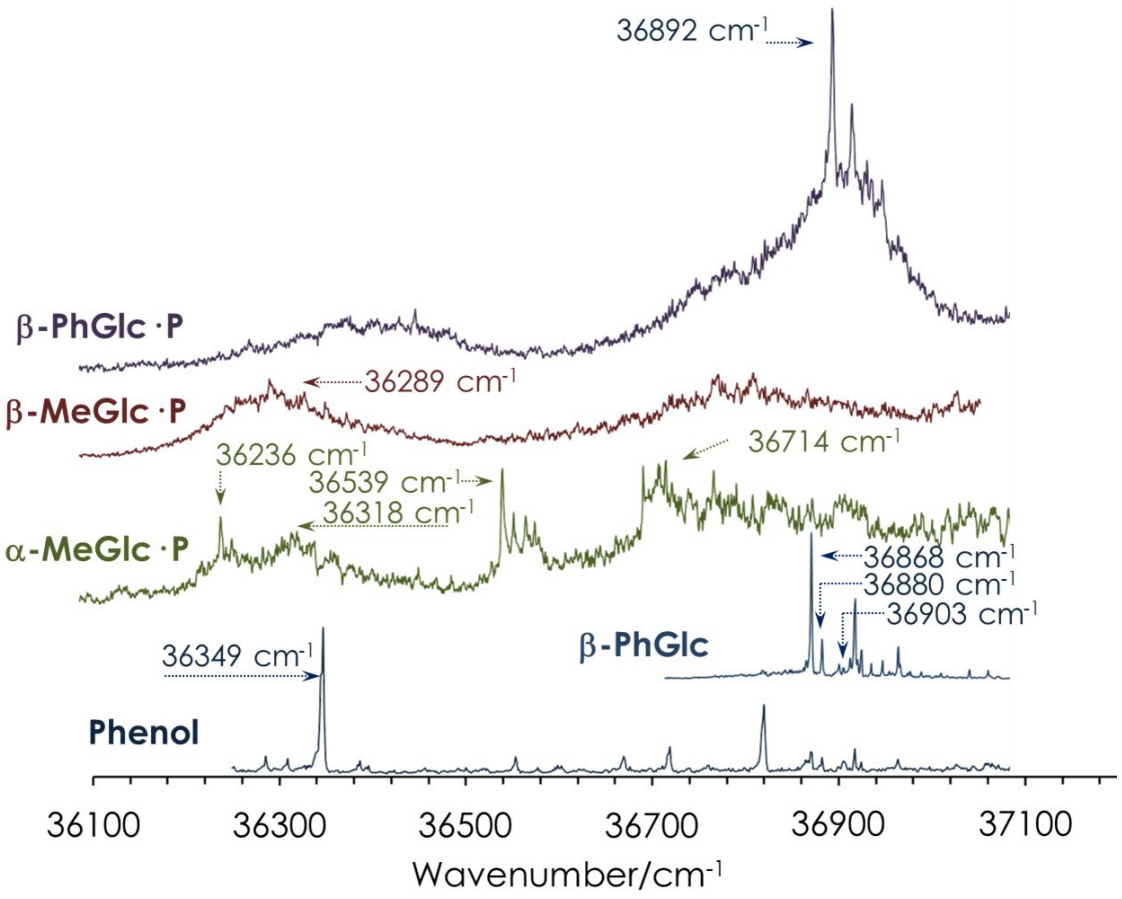


Figure 1. REMPI spectra of phenol, β-PhGlc, β-MeGlc·P, α-MeGlc·P and β-PhGlc·P. The arrows indicate the wavelengths probed in the IR/UV experiments.

Next, we will present and analyze the IR/UV spectra of each system.

α-MeGlcthyl-D-glucopyranose···phenol

Two different IR/UV spectra were obtained for this system, demonstrating the existence of at least two isomers (Figure 2). The spectrum obtained probing either at 36539 cm^{-1} or at 36714 cm^{-1} (upper trace in Figure 2) presents a group of three well-defined transitions between 3500 and 3650 cm^{-1}, a stronger and broader transition at ~ 3380 cm^{-1} and a congested region between 2850 and 3200 cm^{-1} corresponding to the C-H stretches. Comparison with the calculated structures shows an excellent agreement between the experimental trace and the predicted spectrum for the two most stable conformers (Figure 3 and S10 of the supporting information),

in which phenol is hydrogen bonded to the oxygen atom of the hydroxymethyl substituent in C6, and at the same time, there is an interaction between the methyl group in C1 and the aromatic ring. The difference between both conformers is a slight rotation of the aromatic ring respect to the sugar unit, and very likely the potential energy barrier between both structures is too small to be able to trap population in the less stable one.

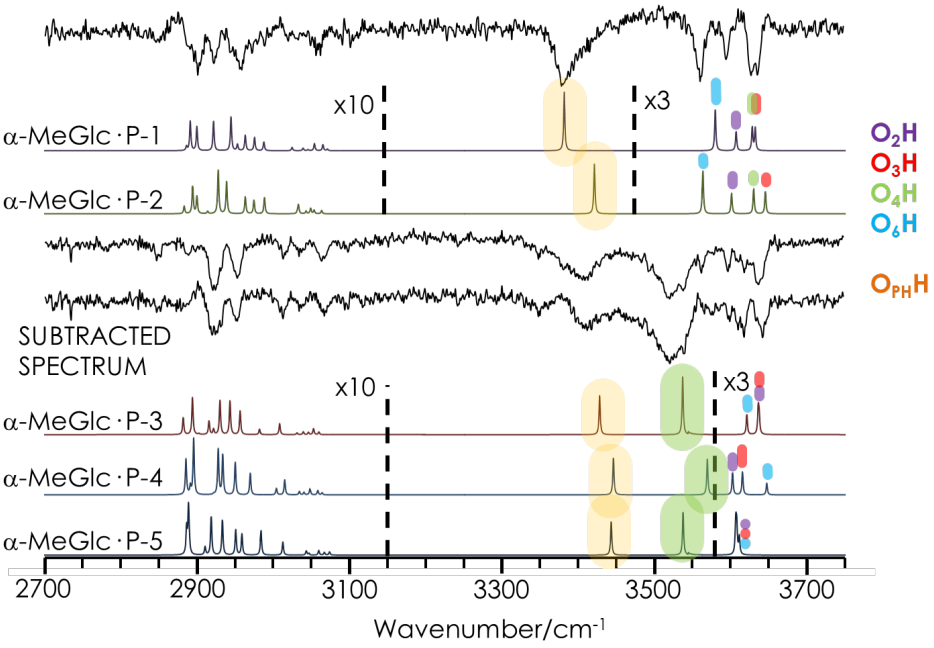


Figure 2. Comparison between experimental IR/UV spectra and simulated ones for methyl-α-D-Glucopyranose···Phenol. The same spectrum was obtained with the UV laser at 36539 and 36714 cm^{-1}, while the lower trace was obtained probing at 36318 cm^{-1}. The resulting spectrum from subtracting both traces is also shown for comparison. The complete set of optimized structures with their relative stability may be found in the electronic supplemental information. A scaling factor of 0.9385 in the OH region and of 0.9525 in the CH stretching region were used. Those factors were obtained as the best fit between the experimental spectra of β-PhGlc and phenol and the corresponding simulated spectra, obtained at the M06-2X/6-3111++G(d,p) level.

Thus, the doublet at 3630 cm^{-1} corresponds to the stretches of O_3H and O_4H, which are immerse in very similar environments. The next peak to the red is due to the stretching vibration of O_2H; its shift indicates formation of a stronger hydrogen bond with O_1. Likewise, the band at 3560 cm^{-1} corresponds to the stretch of O_6H, and its shift seem to indicate a strong inductive effect due to the formation of the hydrogen bond with phenol, whose OH stretch appears more to the red in the form of a broad absorption, consequence of the high anharmonicity induced by the formation of the strong hydrogen bond.

Regarding the other experimental isomer, the spectrum is less clean in part because it was not possible to isolate it from that of isomer 1, but still, the S/N ratio is good enough to identify several discrete transitions around 3615 cm^{-1}, and two broad bands at 3520 and 3415 cm^{-1} respectively. The signature in the CH stretching region is also significantly different from that of the first conformer and confirms that the spectrum is due to a different species. Comparison with the predicted spectra for isomers α-MeGlc·P-3, -4 and -5 shows an excellent agreement with the predicted spectrum of α-MeGlc·P-3, in which the phenol is inserted in the hydrogen bond network, between O_4 and O_6. Thus, the band at 3640 cm^{-1} is due to the superposition of the stretching vibrations of O_2H and O_3H, which, as in the other isomer, are surrounded by a similar neighborhood. The band at 3615 cm^{-1} corresponds to the stretch of O_6H and the broader absorption at 3550 cm^{-1} is originated by the O_4H stretch, which is now hydrogen bonded to the phenol OH moiety. Finally, the phenolic OH group appears as the red-most band at 3415 cm^{-1}.

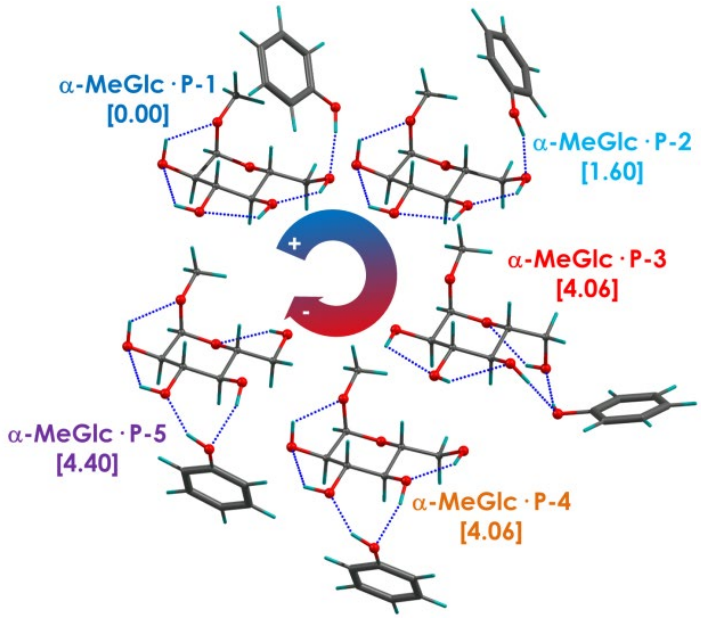


Figure 3. Five lowest energy structures of methyl-α-D-Glucopyranose···Phenol calculated at M062x/6-311++G(d,p) level. Relative stability in kJ/mol

β-MeGlcthyl-D-glucopyranose···phenol

The IR/UV spectrum of β-MeGlc···phenol, Figure 4, looks strikingly different from those of the clusters containing the alpha anomer. It is composed of a band with a shoulder at ~ 3640 cm^{-1} that very likely hides the contribution of several transitions, and two broad absorptions at ~ 3530 and 3420 cm^{-1} that presumably correspond to OH groups taking part in stronger hydrogen bonds. Finally, the signature in the CH stretching region is also different from those of the

clusters of the alpha anomer. Other regions of the REMPI spectrum were also probed but no spectrum that could be assigned to an additional conformer could be obtained (Figure S6).

Comparison with the predicted spectra for the seven most stable structures shows that the position of the broad bands is not reproduced by the two most stable isomers. However, the spectra of β-MeGlc·P-3 and β-MeGlc·P-4 are in good agreement with the experimental trace. In those species, Figure 5 and S11 of the supporting information, the phenol molecule inserts between O4 and O6, in a similar way as in α-MeGlc·P-3, while the aromatic ring lies away from the sugar unit. Isomer β-MeGlc·P-4 is identical, but with a different orientation of the aromatic ring, which now is interacting with the hydrophobic side of the sugar. This second structure presents a shift in the O_4H stretch from 3556 cm^{-1} in β-MeGlc·P-3 to 3538 cm^{-1}, and therefore it does not fit the experimental trace as well as the other isomer.

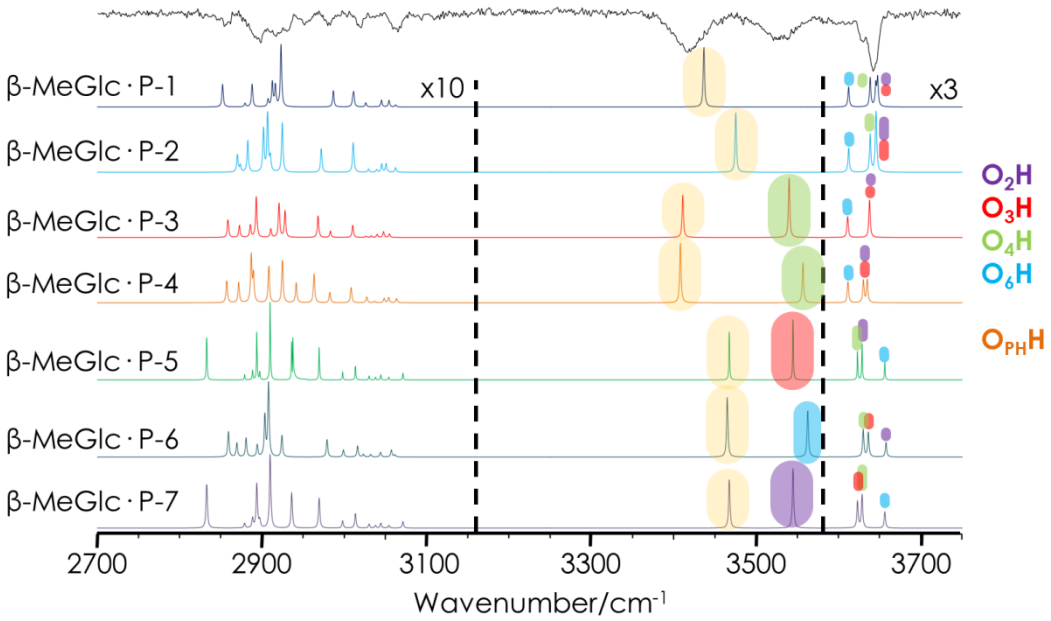


Figure 4. Comparison between experimental IR/UV spectra and simulated ones for methyl-β-D-Glucopyranose···Phenol

If one accepts as valid the assignment to isomer β-MeGlc·P-3, then the band at 3648 cm^{-1} in the experimental trace hides the contribution of three transitions, corresponding to the stretches of O_2H, O_3H (which are coupled) and O_6H, while the two broad bands correspond to the stretches of O_4H and the phenolic OH, which are the moieties involved in the formation of the strongest hydrogen bonds.

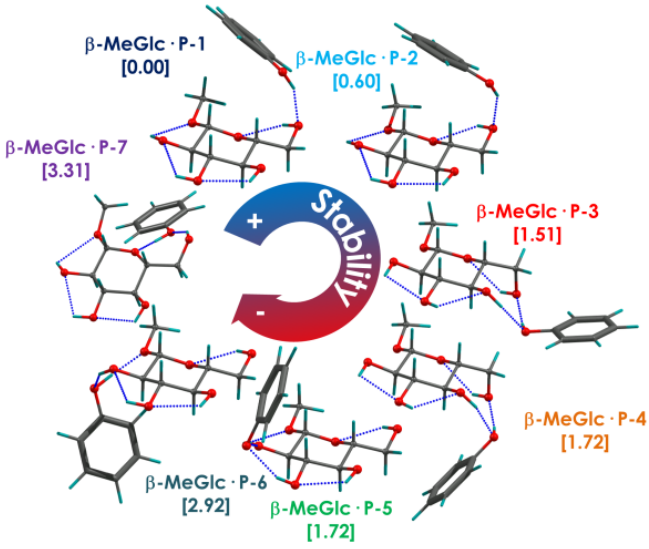


Figure 5. Seven lowest energy structures of methyl-β-D-Glucopyranose···Phenol calculated at M06-2X/6-311++G(d,p) level. Relative stability in kJ/mol

β-PhGlcenyl-D-glucopyranose···phenol

The voluminous substituent in β-PhGlc compared to β-MeGlc does not introduce noticeable changes in the sugar's conformational landscape, as it can be seen by the comparison between the structures in Figures S1 and S3 of the supporting information. However, the former presents an attractive site for phenol to form a stacking interaction. Furthermore, it has been demonstrated in solution that sugars are prone to establish stacking interactions with aromatic molecules.[31]

The spectrum of β-PhGlc·P, Figure 6, resembles that of β-MeGlc·P, but for the absence of the band at 3530 cm^{-1}. Thus, the spectrum presents a band with a shoulder at 3640 cm^{-1}, which very likely hides the contribution from several vibronic transitions, a second, narrower band at 3595 cm^{-1} and a broad absorption at 3390 cm^{-1}, pointing to the formation of a strong hydrogen bond. The signature in the CH stretching region is particularly weak for this complex and not very informative. It is worthy to note that no similar cluster with β-PhGlc was studied in ref [28]. To aid in the assignment, the 3620-3670 cm^{-1} region was also scanned at higher resolution using a dye laser instead of the OPO system.

Comparison with the spectra predicted for the calculated structures shows an excellent agreement with the spectrum of the global minimum (Figure 6), in which phenol is hydrogen bonded to hydroxymethyl moiety (Figure 7, and S12 of the supporting information), while at the same time presents C-H···π interactions with the aromatic ring. This structure is very similar to that of α-MeGlc·P-1 and to the global minimum of β-MeGlc·P, which however, was not

detected. Assuming that the spectrum corresponds to the global minimum, the band with a shoulder at the blue-end of the spectrum in Figure 6 contains the contribution from the stretching vibrations of O_2H, O_3H and O_4H, which are forming a hydrogen bond network, while the band at 3595 cm^{-1} is the stretch vibration of O_6H, and the broad band is, as in the case of the previous complexes, due to the phenolic hydroxyl group.

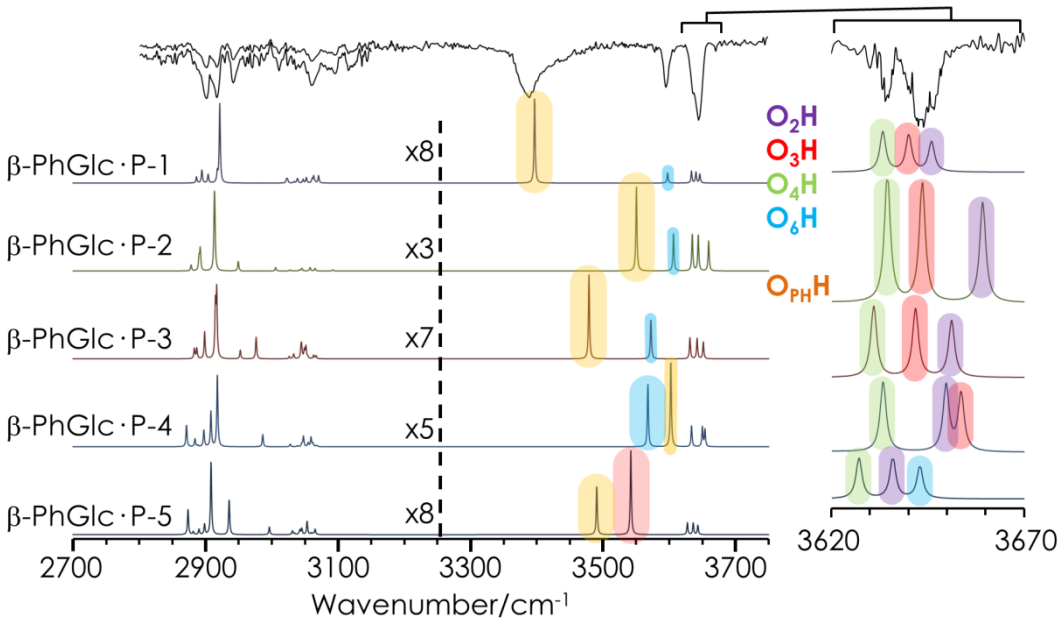


Figure 6. Comparison between experimental IR/UV spectra and simulated ones for phenyl-α-D-Glucopyranose···Phenol

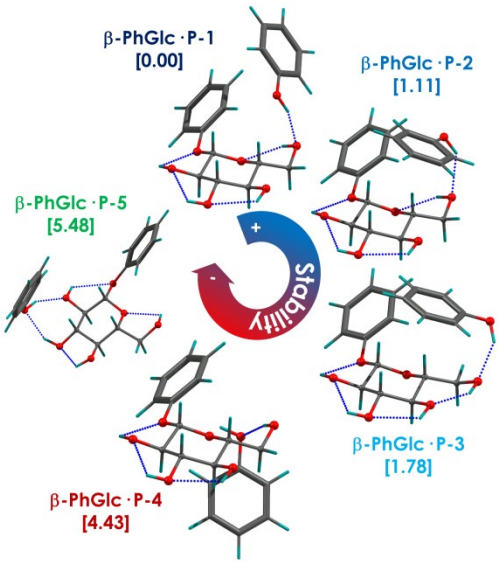


Figure 7. Five lowest energy structures of phenyl-β-D-Glucopyranose···Phenol calculated at M06-2X/6-311++G(d,p) level. Relative stability in kJ/mol

DICUSSION

Influence of the entropy on the stability of the aggregates

The assignments proposed in Figs. 2, 4 and 6 for the spectra of sugar-PhGlcenol aggregates points to calculated structures that are above the global minimum for some of the systems. Certainly, the best fit for the experimental spectrum of β-MeGlc·P is a family of structures that appear at 1.50 kJ/mol above the global minimum (β-MeGlc·P-3). It is also worthy to note that while a single isomer (or a single family of isomers) was detected for β-MeGlc·P, two isomers were detected for the other two systems studied. The number of species present in a beam depends on the number of possible conformers for the system, but also on the barriers of the isomerization paths connecting such minima and on the kinetics and dynamics of the cooling process. To understand the process, it is sometimes useful to examine not only the relative enthalpies of the calculated species, but also their relative entropy and the change in Gibbs Free Energy (ΔG) in the interval of temperatures from the desorption process to the final temperature in the beam. In the case of molecular aggregates the Gibbs relative energy, ΔG_R and the Gibbs binding free energy, ΔG_B may be even more informative, as it will be discussed below.

Figure 8a shows the variation in ΔG_R, obtained using the equations 1-5. The red bar in Figure 9a indicates the temperature limit for the stability of the organic compounds, while the gray bar indicates the temperature of decomposition of the pyranose ring and therefore it could not be surpassed by the monomers that survived the desorption process. Thus all the species detected should be reasonably stable at that temperature. The molecules start aggregating once they reach the temperature at which the ΔG_B becomes negative, and therefore the aggregate is stable (pink bars in Figure 8). Two isomers from two different families were detected for α-MeGlc·P. One of them, the global minimum, is also the most stable structure in Figure 9a below the aggregation temperature, while the other representative structure from that family (α-MeGlc·P-2) is 1-2 kJ/mol less stable in the interval of temperatures below the aggregation barrier. The structures from the other two families (α-MeGlc·P-3,-4 and -5) are also very close in stability and they seem to be favored by the entropy. From an energetic point of view, all five conformers should be formed and detected in the experiment. Therefore, either their spectra are too similar and each family is detected as a single isomer, or the barriers for interconversion between each family of structures are small enough to be overcome during the cooling process, and consequently, only the most stable structures of the two families were detected.

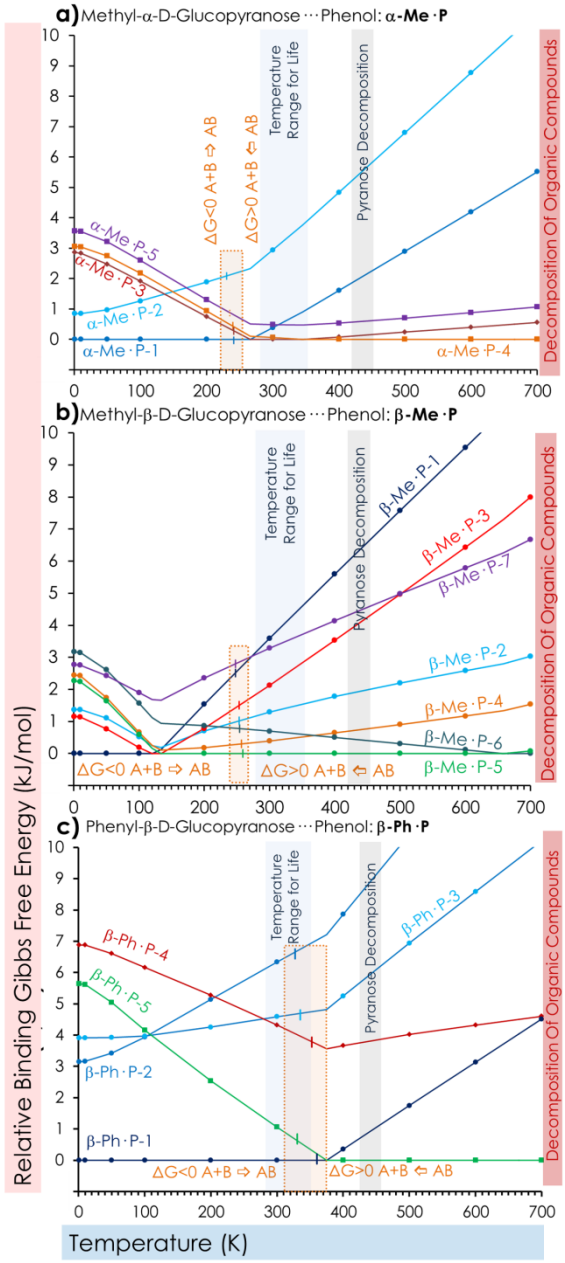


Figure 9. Gibbs relative free energy of the conformations in Figures 3, 5 and 7 for α-MeGlc·P (a), β-MeGlc·P (b) and β-PhGlc·P (c). The red bar indicates the temperature at which most of the organic compounds decompose, while the gray bar indicates the temperature of decomposition of pyranose. The pink bar indicates where ΔG becomes positive and therefore, the cluster is no longer stable.

The energetic landscape is significantly more complex for β-MeGlc·P (Figure 8b). Two very similar structures of the same family were proposed for the assignment of the singly detected experimental isomer: β-MeGlc·P-3 and -4. Analysis of Figure 8b shows five structures in less

than ~ 0.5 kJ/mol in the temperature window of our experiment. Such small energy difference is well within the calculation errors and therefore the energetics order may not be correctly predicted. Nevertheless, the landscape in Figure 9b points to a potential energy surface with a topography that allows the complex to reach the singly-detected isomer during the cooling process.

The simplest energetic panorama was found for β-PhGlc···P: the detected isomer was assigned to the calculated global minimum, which becomes also the most stable structure below the aggregation temperature (Figure 8c). It is also worthy to note that the extra stabilization energy due to the interaction with the phenyl ring results in a significantly higher temperature of aggregation.

All the above demonstrates that although the enthalpy of the system plays an important role, one has to take into account the Gibbs free energy to understand the energetics of the molecular aggregates containing sugar units.

Conformational preferences

The substituted Glcs whose clustering preferences have been analyzed in this work present four OH groups, which in the isolated molecule form a cooperative hydrogen bond network. At first glance, they may appear as an excellent insertion point to attach to for molecules with hydrophilic groups. However, the partner molecule needs first to spend some energy in breaking the network, which under the jet conditions is a hard and discouraging task. Accordingly, phenol attaches preferentially to the hydroxymethyl group in all the structures detected in this work, and apart from the less stable isomer of α-MeGlc·P, in a way that allows phenol's aromatic ring to interact with the anomeric substituent. The exocyclic hydroxymethyl substituent is also known to add extra strength to the sugar-ligand bonds in polisaccarides,[32] as it is the most flexible part of the sugar and therefore it is loose enough to accommodate its position to form stronger intermolecular hydrogen bonds. Thus, such group may be regarded as a hydrophilic arm that the molecule uses to interact with the surrounding medium.

An interesting fact is that the α/β position of the anomeric substituent has a larger influence in the cluster's structure than the nature of the substituent. Actually, the unusual preference of the axial form over the equatorial arrangements is known as the *anomeric effect* and its physical origin still generates a controversy nowadays. Certainly, the position of phenol in α-MeGlc·P is almost identical to that in β-PhGlc·P, but it is very different from the one adopted in β-MeGlc·P single detected isomer. As explained above, the global minimum of β-MeGlc·P is very similar to

that of the other two systems, but it becomes less stable at the temperatures in the beam with the Gibbs free energy is taken into account.

The carbohydrate-amino acid residue interactions in the gas phase are governed by hydrogen bonding, while the weak CH-π interactions give the "final push" to the structure to reach its final shape. This behavior contrasts with what is observed in physiological environment, where the hydroxyls groups of the carbohydrate are busy interacting with the solvent and therefore, CH-π interactions with the other ligand (ej. proteins) are predominant. The role of flexible exocyclic hydroxymethyl group is capital and favors the addition of polar molecules through an hydrogen bond in position (4,6).

CONCLUSIONS

The aggregational preferences of three sugars: α/β- methyl-D-glucopyranose and β- phenyl-D-glucopyranose with phenol were explored using a combination of mass-resolved excitation spectroscopy and quantum mechanical calculations. Interaction with the flexible hydroxymethyl substituent was observed in all cases, probably because the $-CH_2-$ group gives it an extra conformational freedom that the sugar uses to accommodate the phenol in an optimal position. These assignments were achieved thanks in part to an improved computational protocol that introduced the use of the Gibbs free energy of the aggregation process. The results from such analysis highlight the complicated conformational landscape of this kind of systems.

Acknowledgements

References

1 A. Lehninger, D. Nelson and M. Cox, *Lehninger Principles of Biochemistry*, W. H. Freeman, 2008.

2 H. Ghazarian, B. Idoni and S. B. Oppenheimer, *Acta Histochem.,* 2011, **113**, 236-247 (DOI:http://dx.doi.org/10.1016/j.acthis.2010.02.004).

3 E. C. Stanca-Kaposta, D. P. Gamblin, E. J. Cocinero, J. Frey, R. T. Kroemer, A. J. Fairbanks, B. G. Davis and J. P. Simons, *J. Am. Chem. Soc.,* 2008, **130**, 10691-10696.

4 H. Gabius, S. André, J. Jiménez-Barbero, A. Romero and D. Solís, *Trends Biochem. Sci.,* 2011, **36**, 298-313 (DOI:http://dx.doi.org/10.1016/j.tibs.2011.01.005).

5 S. Hakomori and Y. Igarashi, *J. Biochem.,* 1995, **118**, 1091-1103.

6 M. Fernández-Alonso del Carmen, F. J. Cañada, J. Jiménez-Barbero and G. Cuevas, *J. Am. Chem. Soc.,* 2005, **127**, 7379-7386 (DOI:10.1021/ja051020+).

7 T. B. Grindley, in , ed. ed. B. Fraser-Reid, K. Tatsuta and J. Thiem, Springer Berlin Heidelberg, 2008, pp.3-55.

8 B. Fiege, C. Rademacher, J. Cartmell, P. I. Kitov, F. Parra and T. Peters, *Angewandte Chemie International Edition,* 2012, **51**, 928-932 (DOI:10.1002/anie.201105719).

9 M. Fernández-Alonso del Carmen, D. Díaz, M. Á Berbis, F. Marcelo, J. Cañada and J. Jiménez-Barbero, *Curr. Protein Peptide Sci.,* 2012, **13**, 816-830 (DOI:10.2174/138920312804871175).

10 C. Oertelt, B. Lindner, M. Skurnik and O. Holst, *European Journal of Biochemistry,* 2001, **268**, 554-564.

11 E. J. Cocinero and P. Çarçabal, in *in Gas-Phase IR Spectroscopy and Structure of Biological Molecules,* ed. ed. A. M. Rijs and J. Oomens, Springer, New York, 2015, pp.299-334.

12 E. J. Cocinero, P. Carcabal, T. D. Vaden, J. P. Simons and B. G. Davis, *Nature,* 2011, **469**, 76-U1400.

13 E. J. Cocinero, P. C[redacted] ºarc[redacted] ºabal, T. D. Vaden, B. G. Davis and J. P. Simons, *J. Am. Chem. Soc.,* 2011, **133**, 4548-4557.

14 Z. Su, E. J. Cocinero, E. C. Stanca-Kaposta, B. G. Davis and J. P. Simons, *Chem.Phys.Lett.,* 2009, **471**, 17-21.

15 J. P. Simons, B. G. Davis, E. J. Cocinero, D. P. Gamblin and E. C. Stanca-Kaposta, *Tetrahedron-Asymmetry,* 2009, **20**, 718-722.

16 J. P. Simons, E. C. Stanca-Kaposta, E. J. Cocinero, B. Liu, B. G. Davis, D. P. Gamblin and R. T. Kroemer, *Phys. Scripta,* 2008, **78**.

17 E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J. Grabow, J. A. Fernández and F. Castaño, *Angewandte Chemie International Edition,* 2012, **51**, 3119-3124 (DOI:10.1002/anie.201107973).

18 E. J. Cocinero, A. Lesarri, P. Écija, A. Cimas, B. G. Davis, F. J. Basterretxea, J. A. Fernández and F. Castaño, *J. Am. Chem. Soc.,* 2013, **135**, 2845-2852 (DOI:10.1021/ja312393m).

19 E. Stanca-Kaposta, P. Çarçabal, E. J. Cocinero, P. Hurtado and J. P. Simons, *J Phys Chem B,* 2013, **117**, 8135-8142 (DOI:10.1021/jp404527s).

20 I. León, E. Cocinero, J. Millán, S. Jaeqx, A. Rijs, A. Lesarri, F. Castaño and J. A. Fernández, *Phys. Chem. Chem. Phys.,* 2012, **14**, 4398.

21 M. Frisch, 2009, .

22 E. Aguado, I. Leon, E. J. Cocinero, A. Lesarri, J. A. Fernandez and F. Castano, *Phys. Chem. Chem. Phys.,* 2009, **11**, 11608-11616.

23 I. Leon, R. Montero, A. Longarte and J. A. Fernandez, *Phys. Chem. Chem. Phys.,* 2015, **17**, 2241-2245 (DOI:10.1039/C4CP03667K).

24 I. León, J. Millán, E. J. Cocinero, A. Lesarri and J. A. Fernández, *Angewandte Chemie International Edition,* 2014, **53**, 12480-12483 (DOI:10.1002/anie.201405652).

25 I. León, I. Usabiaga, J. Millán, E. J. Cocinero, A. Lesarri and J. A. Fernández, *Phys. Chem. Chem. Phys.,* 2014, (DOI:10.1039/C4CP01702A).

26 K. K. Irikura, in , ed. nonymous American Chemical Society, 1998, pp.402-418.

27 S. F. Boys and F. Bernardi, *Mol. Phys.,* 1970, **19**, 553-566.

28 F. O. Talbot and J. P. Simons, *Phys. Chem. Chem. Phys.,* 2002, **4**, 3562-3565 (DOI:10.1039/B204132D).

29 J. L. Alonso, M. A. Lozoya, I. Pena, J. C. Lopez, C. Cabezas, S. Mata and S. Blanco, *Chem. Sci.,* 2014, **5**, 515-522 (DOI:10.1039/C3SC52559G).

30 D. S. McClure, *Can. J. Chem.,* 1958, **36**, 59-71.

31 E. Jimenez-Moreno, G. Jimenez-Oses, A. M. Gomez, A. G. Santana, F. Corzana, A. Bastida, J. Jimenez-Barbero and J. L. Asensio, *Chem. Sci.,* 2015, **6**, 6076-6085 (DOI:10.1039/C5SC02108A).

32 C. S. Barry, E. J. Cocinero, P. Çarçabal, D. P. Gamblin, E. Stanca-Kaposta, S. M. Remmert, M. C. Fernández-Alonso, S. Rudic, J. P. Simons and B. G. Davis, *J. Am. Chem. Soc.,* 2013, **135**, 16895-16903 (DOI:10.1021/ja4056678).