

# Isotopic Hydration of Cellobiose: Vibrational Spectroscopy and Dynamical Simulations

Madeleine Pincu,<sup>†</sup> Emilio J. Cocinero,<sup>‡</sup> Nitzan Mayorkas,<sup>§</sup> Brina Brauer,<sup>||</sup> Benjamin G. Davis,<sup>⊥</sup> R. Benny Gerber,<sup>\*,†,||</sup> and John P. Simons<sup>\*,||</sup>

<sup>†</sup>Department of Chemistry, University of California, Irvine, California 92697, United States

<sup>‡</sup>Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco, (UPV – EHU), Apartado 644, E-48940, Bilbao, Spain

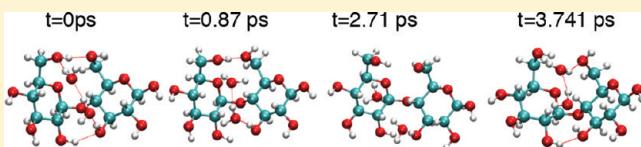
<sup>§</sup>Department of Physics, Ben Gurion University of the Negev, Beer Sheva 84105, Israel

<sup>||</sup>Institute of Chemistry and The Fritz Haber Research Center, The Hebrew University, Jerusalem 91904, Israel

<sup>⊥</sup>Department of Chemistry, University of Oxford, Chemical Research Laboratory, Mansfield Road, Oxford OX1 3TA, U.K.

<sup>\*</sup>Department of Chemistry, University of Oxford, Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford, OX1 3QZ, U.K.

**ABSTRACT:** The conformation and structural dynamics of cellobiose, one of the fundamental building blocks in nature, its C4' epimer, lactose, and their microhydrated complexes, isolated in the gas phase, have been explored through a combination of experiment and theory. Their structures at low temperature have been determined through double resonance, IR-UV vibrational spectroscopy conducted under molecular beam conditions, substituting D<sub>2</sub>O for H<sub>2</sub>O to separate isotopically, the carbohydrate (OH) bands from the hydration (OD) bands. Car–Parrinello (CP2K) simulations, employing dispersion corrected density functional potentials and conducted “on-the-fly” from ~20 to ~300 K, have been used to explore the consequences of raising the temperature. Comparisons between the experimental data, anharmonic vibrational self-consistent field calculations based upon ab initio potentials, and the CP2K simulations have established the role of anharmonicity; the reliability of classical molecular dynamics predictions of the vibrational spectra of carbohydrates and the accuracy of the dispersion corrected (BLYP-D) force fields employed; the structural consequences of increasing hydration; and the dynamical consequences of increasing temperature. The isolated and hydrated cellobiose and lactose units both present remarkably rigid structures: their glycosidic linkages adopt a “cis” (anti- $\phi$  and syn- $\psi$ ) conformation bound by inter-ring hydrogen bonds. This conformation is maintained when the temperature is increased to ~300 K and it continues to be maintained when the cellobiose (or lactose) unit is hydrated by one or two explicitly bound water molecules. Despite individual fluctuations in the intra- and intermolecular hydrogen bonding pattern and some local structural motions, the water molecules remain locally bound and the isolated carbohydrates remain trapped within the cis potential well. The Car–Parrinello dynamical simulations do not suggest any accessible pathway to the trans conformations that are formed in aqueous solution and are widespread in nature.



## INTRODUCTION

The factors which dictate the conformational preferences of cellobiose, Glcp- $\beta$ 1,4-D-Glc $p$ , the basic repeat structural unit of cellulose, remain something of an enigma. Ab initio<sup>1</sup> and density functional theory (DFT)<sup>2,3</sup> calculations of its intrinsic conformation, isolated in the gas phase, predict a cis configuration about the glycosidic linkage, i.e. one in which the two dihedral angles,  $\phi$  and  $\psi$ , defined in Figure 1, adopt anti- $\phi$  and syn- $\psi$  orientations and the two neighboring hydroxymethyl groups are linked through an inter-ring hydrogen bond. These predictions have recently been confirmed through an infrared spectroscopic investigation of phenyl  $\beta$ -cellobioside, isolated at low temperature in a molecular beam environment.<sup>3</sup> In aqueous solution at ~300 K, however, NMR measurements of scalar coupling and inter-residue nuclear Overhauser effects (NOEs), together with molecular dynamics simulations, indicate an average trans (syn- $\phi$ , syn- $\psi$ ) structure

about the glycosidic linkage in cellobiose<sup>4</sup> and also methyl  $\alpha$  and  $\beta$ -cellobioside;<sup>5–12</sup> the trans structure is supported by inter-ring hydrogen bonding between OH3 and OS', see Figure 1. Cellobiose<sup>13</sup> and methyl  $\beta$ -cellobioside<sup>14</sup> also adopt a similar, hydrogen-bonded trans conformation in the crystalline state and in the cellulose polymer the trans orientation of the  $\beta$ 1,4 linked cellobiose subunits creates an alternating structure,<sup>15</sup> supported by an extended chain of OH3-O5' hydrogen bonds.

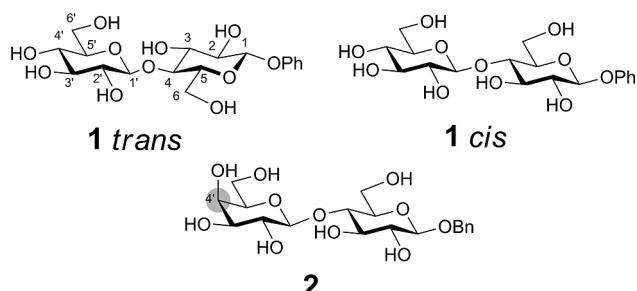
Is the change from the intrinsic cis conformation of cellobiose at low temperatures in the gas phase to the trans conformation

**Special Issue:** David W. Pratt Festschrift

**Received:** December 21, 2010

**Revised:** March 18, 2011

**Published:** June 02, 2011



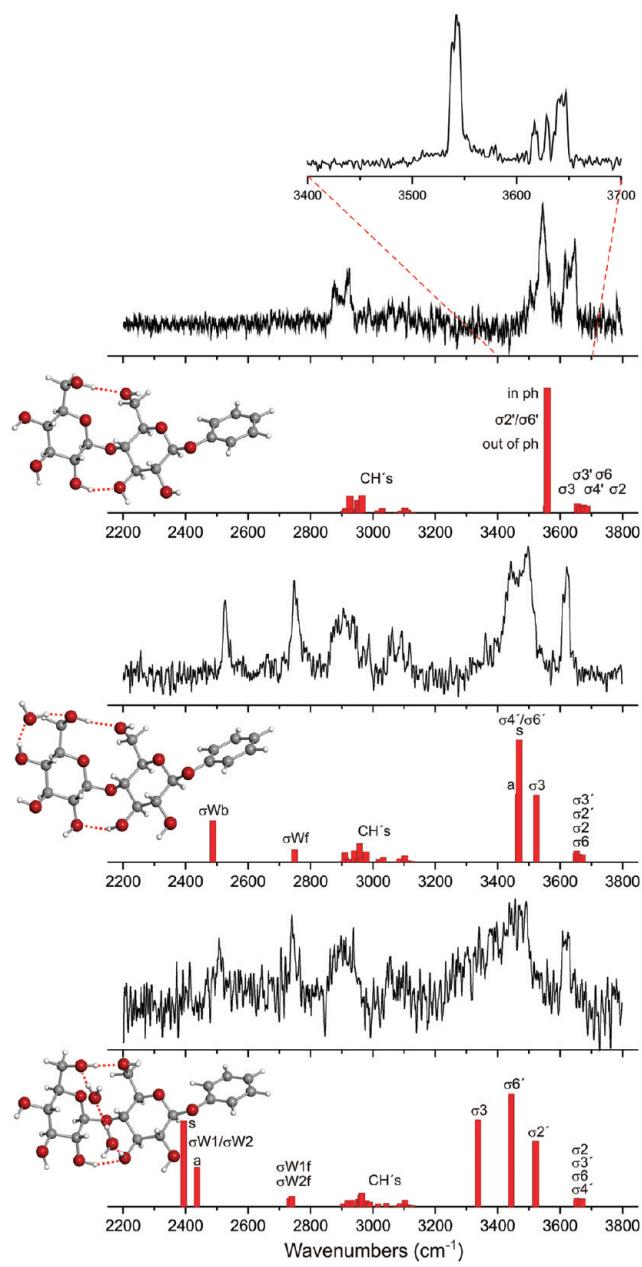
**Figure 1.** Structural representations of phenyl  $\beta$ -celllobioside, **1**, and benzyl  $\beta$ -lactoside, **2**, shown in its cis conformation; lactose is the C4' epimer of cellobiose. (Glycosidic dihedral angles are defined as  $\phi$ : H1'-C1'-O-C4 and  $\psi$ : C1'-O-C4-H4.)

adopted at  $\sim 300$  K in condensed phases associated with an aqueous environment, explicit solvation, the raised temperature, or perhaps all three? Is it promoted by a crystal field? Is the trans orientation of the cellobiose subunits in cellulose actively created during its biosynthesis when the glucose units are successively added at the nonreducing end of the growing  $\beta$ 1,4 linked chain,<sup>16</sup> or is it accessed by relaxation of cis-oriented cellobiose units at the growing terminus, perhaps through explicit hydration or other intermolecular interactions?

The effect of hydration, which provides a leitmotif throughout the present work, continues to be addressed theoretically through increasingly sophisticated molecular dynamics (MD) simulations conducted in aqueous solution, and in vacuo. The most recent and advanced examples of this approach<sup>8,11,12</sup> (see refs 11 and 12 for comprehensive reviews of the earlier literature) present the results of a series of 50 ns MD simulations of cellobiose, based upon a modified GROMOS force field (45A4),<sup>8,11</sup> and of methyl  $\beta$  cellobioside, based on the CHARMM carbohydrate<sup>11</sup> or a “carbohydrate solution” force-field (CSFF).<sup>12</sup> In both cases, the simulations, validated through comparisons with experimental NMR data, were in accord with the trans glycosidic conformations found in the condensed phase. They each identified a persistent intramolecular hydrogen bond between OH3 and O5', and in aqueous methyl cellobioside they also predicted persistent inter-ring water bridges.<sup>12</sup>

Explicit hydration in the gas phase has been investigated through measurements of the vibrational spectrum of singly hydrated phenyl  $\beta$ -celllobioside and also its C-4' epimer, benzyl  $\beta$ -lactoside (see Figure 1), isolated in a molecular beam at low temperature.<sup>3</sup> Structural assignments were based on comparisons between the experimental OH vibrational spectrum and the results of DFT and ab initio calculations. The water molecule was found to insert between OH-4' and the neighboring oxygen, O6, the binding site favored in glucose (but, perhaps importantly, blocked in the  $\beta$ 1,4 linked cellulose polymer itself). This insertion promoted a structural reorganization of the disaccharide into a configuration resembling that of its unhydrated C-4' epimer, but nevertheless, the cis conformation was retained.<sup>3</sup>

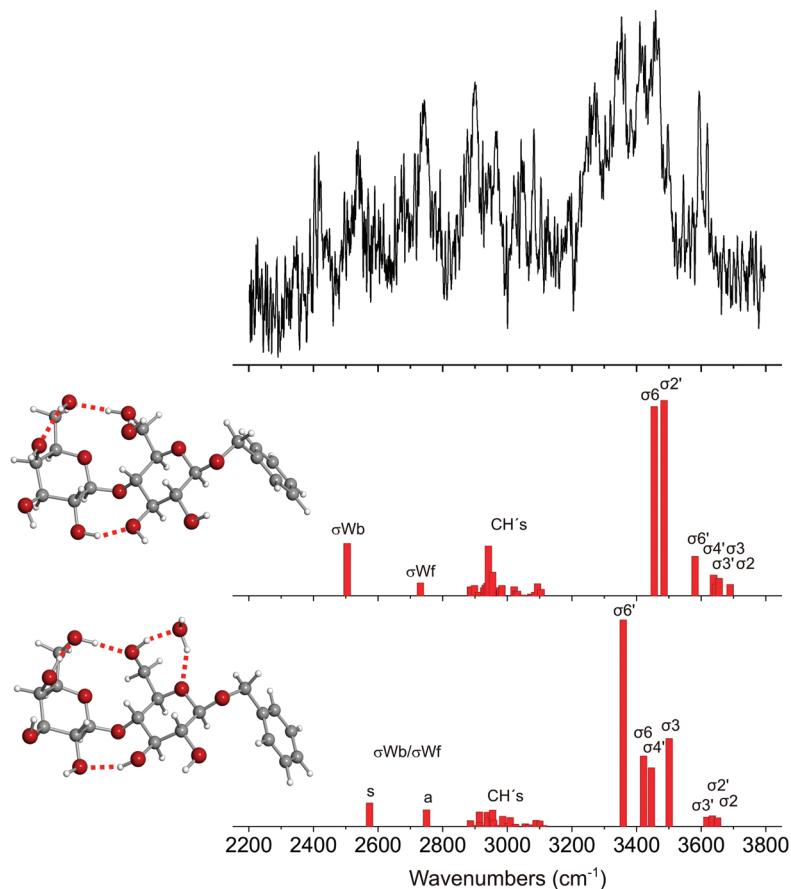
The MD simulations (which depend on the accuracy of the force fields employed and the reliability of classical mechanics) and the NMR measurements focus on average structures and thermodynamic properties and have been conducted in aqueous solution at a single temperature, 300 K. The vibrational spectroscopic measurements coupled with DFT and ab initio calculations (which rely on empirical scaling factors to bring the predicted harmonic frequencies into closer correspondence with those measured experimentally) focus on the assignment of



**Figure 2.** IRID and DFT computed IR spectra and the corresponding hydrogen-bonded molecular structures of phenyl  $\beta$ -celllobioside (top), phenyl  $\beta$ -celllobioside ·  $\text{D}_2\text{O}$  (middle), and phenyl  $\beta$ -celllobioside ·  $(\text{D}_2\text{O})_2$  (bottom).

static, minimum energy conformations and were conducted in the gas phase at a temperature  $\sim 10$  K.

The present work extends the focus by relating Car–Parrinello (CP2K) simulations,<sup>17</sup> conducted over a range of temperatures from  $\sim 20$  to  $\sim 300$  K, to new experimental spectroscopic measurements of microhydrated phenyl  $\beta$ -celllobioside and benzyl  $\beta$ -lactoside. These have enabled assessment of the reliability of classical MD predictions of the vibrational spectra of carbohydrates and the accuracy of the force fields they employ; the role of anharmonicity; the structural consequences of increasing hydration; and the dynamical consequences of increasing temperature. The strategies, which follow those



**Figure 3.** IRID and DFT computed IR spectra and corresponding hydrogen-bonded molecular structures of the two lowest energy structures of benzyl  $\beta$ -lactoside  $\cdot$  D<sub>2</sub>O; the labels  $\sigma$ Wf and  $\sigma$ Wb identify the vibrational bands associated with the “free” and hydrogen-bonded OD groups. The DFT calculations predict relative energies that differ by only 0.6 kJ mol<sup>-1</sup>.

successfully employed in a recent Car–Parrinello study of the spectroscopy of  $\alpha$  and  $\beta$ -D-glucose and sucrose,<sup>18</sup> are first evaluated through comparison with the new spectroscopic measurements and then used to predict the temperature dependence of their vibrational spectra and the structural dynamics they reflect. The experimental approach has been significantly enhanced by substituting D<sub>2</sub>O for H<sub>2</sub>O; this isotopically separates the carbohydrate (OH) bands from the water (OD) bands and simplifies analysis of the vibrational spectra. The Car–Parrinello simulations, conducted “on-the-fly”, employ dispersion corrected DFT (BLYP-D)<sup>19,20</sup> potentials. Comparisons are also made between the computed vibrational spectrum of cellobiose, the spectra predicted by quantum mechanical anharmonic vibrational self-consistent field (VSCF) calculations<sup>18,21–23</sup> based on high level potentials (at 0 K), and the experimental spectrum of phenyl  $\beta$ -cellobioside, recorded at low temperature in the gas phase.

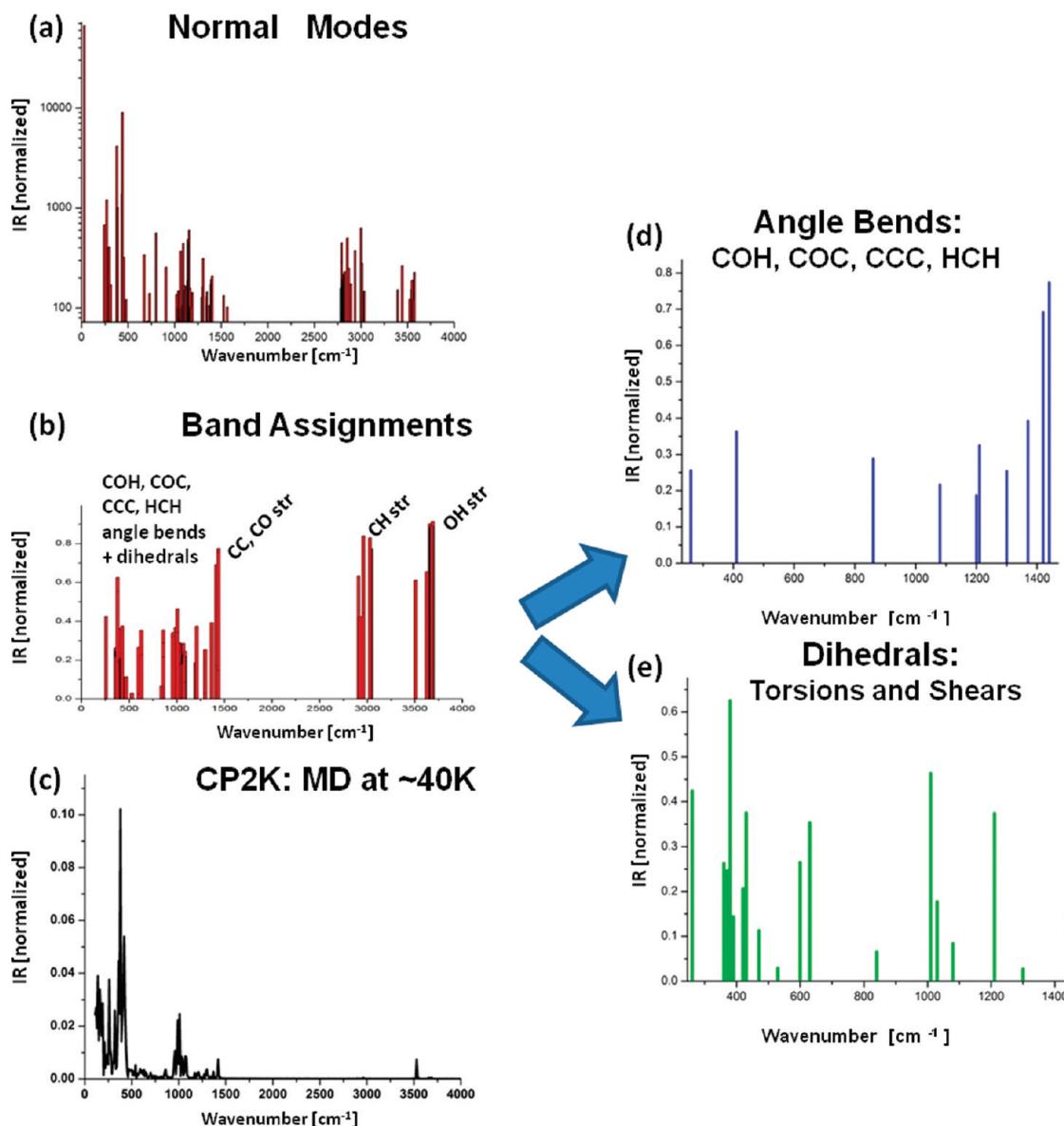
Although the results do not resolve the cis–trans enigma identified earlier in this section, a number of important properties of cellobiose have been established, for example, identification of the preferred sites for the hydration of  $\beta$ -cellobiose by one or two water molecules. More generally, the comparisons between theory and experiment have helped to validate the use of the Car–Parrinello simulations and in the process have provided a much needed and sometimes unexpected view of the spectroscopy, structure, hydration, and vibrational dynamics of one of the most important and abundant carbohydrate motifs in nature.

## METHODS AND CALCULATIONS

**VSCF Calculations.** Vibrational self-consistent field calculations, using recent variants and algorithms of the method, have proved a very effective tool for computing the vibrational spectra of large molecules and clusters at low temperatures.<sup>18,21,24,25</sup>

The simplest level of VSCF is based on assuming separability of the full vibrational wave functions in the normal modes into a product of single-mode wave functions that are anharmonic. A variant of improved accuracy includes the nonseparable effects, neglected at the simplest VSCF level, by second-order perturbation theory. The method, referred to as VSCF-PT2, and also as CC-VSCF, has emerged as an effective and reliable tool for large-molecule spectroscopic calculations. The main merits of VSCF-PT2, as relevant to the present calculations, are the following:

- (1) Good accuracy, as found in a range of applications, including to small biological molecules and their complexes with water;<sup>18,21,24,25</sup> we note specifically, recent applications to glucose and sucrose.<sup>18</sup>
- (2) VSCF-PT2 is computationally feasible also for moderately large biomolecules. This is made possible by a recent algebraic recasting of the algorithm that greatly improves the scaling of the computational effort of VSCF-PT2 with the number of vibrational modes.<sup>22</sup> This improvement was already employed for glucose, sucrose,<sup>18</sup> and other systems and is also exploited in the present application to cellobiose.



**Figure 4.** (a) VSCF normal mode spectrum for cellobiose obtained with a hybrid potential (PM3-MP2); (b) combined band assignments (bonds, angles, dihedrals) calculated from 5 ps MD/CP2K trajectories obtained with a dispersion corrected DFT potential (BLYP-d2) at  $\sim 40$  K; and (c) full vibrational spectrum (for  $\omega > 150$  cm<sup>-1</sup>). Principal contributions from (d) angle bending and (e) torsional/shearing modes.

(3) VSCF-PT2 can be directly used for ab initio potentials, generated as points for a multidimensional grid in coordinate space.<sup>23</sup>

In the VSCF-PT2 calculations here, an approximation of the full potential surface is employed. It includes terms that depend on single normal modes, together with interactions between pairs of normal modes only<sup>23</sup> and in several tests it was found superior to an approximate representation of the potential as a quartic force field.<sup>23,26</sup>

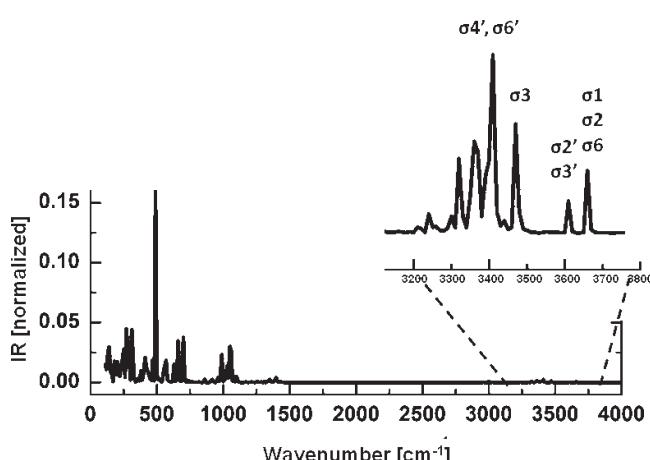
Finally, calculations of the large grid of potential surface points for a molecule the size of cellobiose require a computationally fast, yet sufficiently accurate potential. Following ref 21, we employ here an improved variant of the fast but insufficiently accurate semiempirical potential PM3. The improved PM3 potential surface is constructed by scaling of the standard PM3, using harmonic frequencies from the reliable MP2 ab initio

method. The scaling is such that the improved PM3 potential yields the same frequencies as MP2 when the harmonic approximation is made. Improved PM3 potentials were found much more accurate than standard PM3 ones in spectroscopic applications, notably to biological molecules,<sup>18,21,24,25</sup> and based on this record the VSCF-PT2 method with the above potentials seems to provide the most accurate approach feasible, at present, for molecules the size of cellobiose. (The challenge of first-principles anharmonic calculations with sufficiently accurate potential surfaces is such that for this purpose cellobiose should be regarded as a large molecule. The difficulty is not with the vibrational method itself but with the need to compute in the application a large grid of potential surface points. The calculations for cellobiose using MP2 improved by harmonic input from MP2 are thus quite demanding.)

**"On-the-Fly" Dynamics Calculations.** Vibrational spectra can be computed from system dipole moment correlation

**Table 1.** Experimental IRID Data for the OH Vibrational Bands of Phenyl  $\beta$ -Cellobioside (Assigned by Comparison with DFT/B3LYP Predictions at 0 K) and the Corresponding Theoretical Data for the OH Vibrational Modes in Cellobiose, Obtained from MD Trajectories Based upon a DFT (BLYP-d2) Potential

phenyl $\beta$ -cellobioside		$\beta$ -cellobiose	
experimental: IRID spectrum		calculated: $T \sim 40$ K	
OH vibrational mode	vibrational wavenumber/ $\text{cm}^{-1}$	vibrational wavenumber/ $\text{cm}^{-1}$	relative intensity within OH band
$\sigma_3'$	~3630	3660	0.90
$\sigma_2'$	~3545	3510	0.61
$\sigma_4'$	~3640–50	3680	0.90
$\sigma_6'$	~3545	3530	0.66
$\sigma_1$	N/A	3690	0.66
$\sigma_2$	~3645	3690	0.91
$\sigma_3$	~3620	3670	0.88
$\sigma_6$	~3645	3680	0.89



**Figure 5.** Vibrational spectrum and OH band assignments of lactose calculated from 5 ps MD/CP2K trajectories obtained with a DFT (BLYP-d3) potential at  $\sim 40$  K.

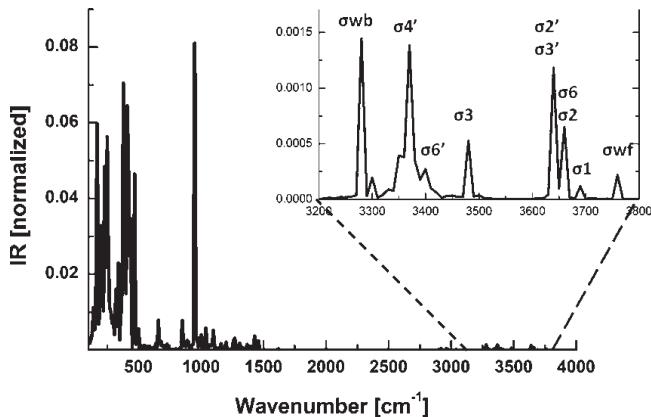
functions obtained from classical trajectories of MD simulations.<sup>27,28</sup> The calculations reported here were based on the Car–Parrinello implementation of the DFT, specifically within the CP2K/QUICK-STEP (CP2K/QS) dynamics simulation package.<sup>29,30</sup> Ab initio potentials were computed at the BLYP level of the Kohn–Sham formulation of DFT with the double- $\zeta$  valence polarization (DZVP) basis set and the Goedecker, Teter, and Hutter (GTH) type pseudopotential.<sup>31</sup> The potentials included Grimme’s dispersion correction at the D2 level,<sup>19</sup> and the (most recently available) D3 level.<sup>20</sup> Earlier implementations of the CP2K/QS package did not include dispersion corrections, which are extremely important in larger systems, especially those that contain water and involve hydrogen-bonding. (Vacha et al.<sup>32</sup> show that dispersion accounts for  $\leq 50\%$  of the total energy of pyridine adsorbed on a water surface.)

Unlike the methods employed for systems of 100+ atoms, where accurate ab initio methods cannot be employed, this quasi classical method is efficient and scalable. (A cellobiose plus  $2\text{H}_2\text{O}$  complex (51 atoms) trajectory of  $\sim 6$  ps duration at 0.5 fs time

**Table 2.** Experimental IRID Data for the OH Vibrational Bands of Benzyl  $\beta$ -Lactoside (Assigned by Comparison with DFT/B3LYP Predictions at 0 K) and the Corresponding Theoretical Data for the OH Vibrational Modes in Lactose, Obtained from MD Trajectories Based upon a DFT (BLYP-d3) Potential

benzyl $\beta$ -lactoside		$\beta$ -lactose	
experimental: IRID spectrum		calculated: $T \sim 40$ K	
OH vibrational mode	vibrational wavenumber/ $\text{cm}^{-1}$	vibrational wavenumber/ $\text{cm}^{-1}$	relative intensity within OH band
$\sigma_3'$	~3600	3620	0.55
$\sigma_2'$	~3585	3610	0.85
$\sigma_4'$	~3390/3430 <sup>a</sup>	3410	0.83
$\sigma_6'$	~3390/3430 <sup>a</sup>	3410 <sup>b</sup>	0.45 <sup>b</sup>
$\sigma_1$	N/A	3660	0.80
$\sigma_2$	~3615	3660	0.86
$\sigma_3$	~3490	3470	0.45
$\sigma_6$	~3620	3660 <sup>b</sup>	0.44 <sup>b</sup>

<sup>a</sup>Coupled vibrational modes. <sup>b</sup>Multiple mode structure.



**Figure 6.** Vibrational spectrum and OH band assignments of cellobiose ·  $\text{H}_2\text{O}$  calculated from 5 ps MD/CP2K trajectories obtained at  $\sim 40$  K; the labels  $\sigma_{\text{wf}}$  and  $\sigma_{\text{wb}}$  identify the free and hydrogen bonded OH vibrational modes in the  $\text{H}_2\text{O}$  molecule.

step, running on an IBM iDataPlex system consisting of nodes, each with two quad-core processors, takes  $\sim 90$  h (total) if it is run on a single node.) Most importantly, when used to simulate vibrational spectra, the calculations reflect the finite temperature at which the simulation was run; spectral broadening as well as contributions from all conformers explored in the course of the trajectory (provided the simulation is sufficiently long) can be directly compared to experimental data at corresponding, or similar temperatures.<sup>27,33,34</sup> In addition, the simulated spectra inherently contain anharmonic effects, such as vibrational energy flow between modes and into collective motions, which are not treated by current VSCF calculations. These benefits make CP2K/QS very attractive and the method has already been applied to calculations of peptide spectra at room temperature,<sup>33</sup> to the study of water clusters (demonstrating their surface acidity),<sup>34</sup> and to saccharides, comparing experimental and computed vibrational spectra.<sup>18</sup> In each case, its predictions have been found to have reasonable accuracy despite its use of BLYP potentials (B3LYP potentials, considered to be superior,<sup>35</sup> are

**Table 3. Comparison between the Vibrational Wavenumbers of Singly Hydrated, Phenyl  $\beta$ -Celllobioside·H<sub>2</sub>O(D<sub>2</sub>O), Measured Experimentally in a Molecular Beam at  $T \sim 10$  K and the Corresponding Wavenumbers of Celllobiose·H<sub>2</sub>O-(D<sub>2</sub>O) Calculated from MD/CP2K Trajectories**

OH vibrational mode	phenyl $\beta$ -celllobioside·H <sub>2</sub> O (·D <sub>2</sub> O)		$\beta$ -celllobiose·H <sub>2</sub> O (·D <sub>2</sub> O)	
	vibrational wavenumber/ $\text{cm}^{-1}$ experimental: IRID	vibrational wavenumber/ $\text{cm}^{-1}$ calculated: $T \sim 40$ K	vibrational wavenumber/ $\text{cm}^{-1}$ experimental: IRID	vibrational wavenumber/ $\text{cm}^{-1}$ calculated: $T \sim 40$ K
$\sigma_2'$	3610–30	(3620)	3660	(3650)
$\sigma_3'$	3610–30	(3620)	3640	(3640)
$\sigma_4'$	~3440	(3430)	~3370 <sup>a</sup>	(3400) <sup>a</sup>
$\sigma_6'$	~3490	(3460)	~3400 <sup>a</sup>	~(3410) <sup>a</sup>
$\sigma_1$	N/A		3690	(3690)
$\sigma_2$	3610–30	(3620)	3660	(3660)
$\sigma_3$	~3500	(3490)	3480	(3510)
$\sigma_6$	3610–30	(3620)	3660	(3660)
$\sigma_{\text{wb}}$	~3420	(2520)	~3280 <sup>a</sup>	~(2440) <sup>a</sup>
$\sigma_{\text{wf}}$	3730	(2750)	3760	(2730)

<sup>a</sup>Calculated mode structure spread over several peaks.

unavailable in CP2K/QS) and the absence of dispersion corrections. The recent improvements made by Grimme,<sup>19,20</sup> which have enabled dispersion corrections to be incorporated into the DFT code, now allow an assessment of their influence on the vibrational spectroscopy.

The Born–Oppenheimer trajectory calculations started with initial minimum energy structures of phenylated cellobiose and lactose and their complexes with one or two H<sub>2</sub>O or D<sub>2</sub>O molecules. The phenyl group was replaced with an H atom to reconstitute the original cellobiose and lactose molecules. The resulting complexes were reoptimized with CP2K/DFT (BLYP-D). Ab initio MD trajectories of 5–6 ps duration were run on constant energy ensembles that consisted of cellobiose with one, or two molecules of H<sub>2</sub>O and D<sub>2</sub>O, and lactose with one H<sub>2</sub>O or D<sub>2</sub>O. The duration of the simulation determines the lowest wavenumber for which the IR spectrum can be accurately predicted, estimated to be  $\sim(210 - 250)$   $\text{cm}^{-1}$ . Simulations were carried out at average temperatures of  $\sim 20$ ,  $\sim 40$ , and  $\sim 300$  K and calculated at time step intervals of 0.5 fs. System dipole moments were computed at each time step over a unit cell of side dimensions 24 Å. Optimization was carried out to ensure that the size of the unit cell did not affect the energy of the optimized conformer. The one-electron orbitals were expanded in a plane wave basis set with a kinetic energy cutoff of 300 Ry, using default values for the Martyna–Tuckerman<sup>36</sup> Poisson solver. The SCF convergence criterion was  $10^{-6}$  Hartree/atom, while the accuracy for the framework's energy computations was set at  $10^{-12}$  Hartree/atom. All computations included the Grimme dispersion correction, using BLYP-D3<sup>20</sup> wherever possible (NERSC installation dependent). IR spectral bands were assigned to individual vibrational modes by matching main peaks of the Fourier Transform of the time dependence of selected bond lengths, angles, or dihedrals, along trajectories of motion.

## EXPERIMENTAL SPECTROSCOPY

The analysis and interpretation of the experimental data began with an extensive search of the many possible complex structures that might be accessed. These were generated and grouped into

**Table 4. Comparison between the Vibrational Wavenumbers of Singly Hydrated, Benzyl  $\beta$ -Lactoside·H<sub>2</sub>O and Benzyl  $\beta$ -Lactoside·D<sub>2</sub>O, Measured Experimentally at  $T \sim 10$  K and the Corresponding Wavenumbers of Lactose·H<sub>2</sub>O and Lactose·D<sub>2</sub>O Calculated from MD Trajectories Using the BLYP-d3 Potential**

OH vibrational mode	benzyl $\beta$ -lactoside·H <sub>2</sub> O (·D <sub>2</sub> O)		$\beta$ -lactose·H <sub>2</sub> O (·D <sub>2</sub> O)	
	vibrational wavenumber/ $\text{cm}^{-1}$ experimental: IRID	vibrational wavenumber/ $\text{cm}^{-1}$ calculated: $T \sim 40$ K	vibrational wavenumber/ $\text{cm}^{-1}$ experimental: IRID	vibrational wavenumber/ $\text{cm}^{-1}$ calculated: $T \sim 40$ K
$\sigma_2'$	~3640	~(3630)	3620	(3610)
$\sigma_3'$	~3620	~(3620)	3610	(3610)
$\sigma_4'$	~3450	~(3450)	3360	(3360)
$\sigma_6'$	~3350	~(3400)	3260	~(3330) <sup>a</sup>
$\sigma_1$	N/A		3660	(3660)
$\sigma_2$	~3650	~(3640)	3660	(3660)
$\sigma_3$	~3500	~(3500)	3460	~(3460) <sup>a</sup>
$\sigma_6$	~3420	~(3420)	3330	(3330)
$\sigma_{\text{wb}}$	~3560	~(2580)	3540	(2550)
$\sigma_{\text{wf}}$	~3760	~(2750)	3760	(2730, 2550) <sup>a</sup>

<sup>a</sup>Calculated mode structure spread over several peaks.

structural families using a combination of the large scale low mode (which uses low frequency modes to construct conformational changes) and Monte Carlo multiple minimization procedures<sup>37</sup> as implemented in the MacroModel software (MacroModel v. 8.5, Schrödinger, LLC21). Subsequent geometry optimization of the lower energy members of each family (typically  $\sim 50$  structures with relative energies  $\leq 15$  kJ mol<sup>-1</sup>), using density functional theory, the B3LYP functional, a 6-311++G(d,p) basis set, and the Gaussian03 suite of programs,<sup>38</sup> led to a new set of structures, zero point corrected relative energies, and (harmonic) vibrational spectra that could be compared with experiment. More accurate energies were then calculated for the optimized DFT structures at the MP2/6-311++G(d,p) level of theory. Structural assignments were based primarily on the level of correspondence between the experimental and computed wavenumbers of the carbohydrate (OH) and water (OD) vibrational bands (scaled by the recommended<sup>39</sup> “anharmonicity” factor, 0.9533, to bring them into better accord with experiment) and second by their calculated relative energies. In all cases, the minimum energy structures provided the best agreement between experiment and theory.

Phenyl  $\beta$ -celllobioside was synthesized following methods described previously<sup>3</sup> and its hydrated (D<sub>2</sub>O) complexes were generated in the gas phase using a combination of pulsed laser ablation and molecular beam procedures. Ground powdered samples of the carbohydrate were thoroughly mixed with graphite powder ( $\sim 20\%$  graphite/80% sample, w/w), deposited as a thin uniform surface layer on a graphite substrate, and placed in a vacuum chamber close to and just below the exit of a pulsed, cylindrical nozzle expansion valve (0.8 mm diameter). The carbohydrates, desorbed by laser evaporation from the surface, were entrained and cooled in an expanding argon jet ( $\sim 4$  bar backing pressure) seeded with D<sub>2</sub>O before passing into the detection chamber through a 2 mm diameter skimmer to create a collimated molecular beam. This was crossed by pulsed tunable UV and IR laser beams in the extraction region of a linear,

Wiley—McLaren time-of-flight mass spectrometer (R.M. Jordan). Mass-selected resonant two photon ionization spectra of the hydrated complexes were recorded using a frequency-tripled pulsed Nd:YAG-pumped dye laser (Sirah).

Their vibrational spectra were subsequently recorded in the OH and OD (and CH) stretch regions, through infrared ion dip (IRID) spectroscopy,<sup>40,41</sup> a double resonance IR-UV technique that depletes the ground state population of the target molecule, or molecular complex whenever the IR radiation is “in tune” with one of its vibrational bands. Monitoring the resulting dip in the mass-selected, resonant two photon ionization (R2PI) signal provides a means of recording its vibrational spectrum. In the present system, the necessary UV chromophore was provided by the (structurally benign) phenyl “tag”. The UV laser was tuned onto a selected resonant two photon ionization (R2PI) band and the IR radiation was tuned over the range 2200–3800 cm<sup>-1</sup> and pulsed at 5 Hz; the UV laser, delayed by ~150 ns, ran at 10 Hz to allow subtraction of the background signals. The IR radiation (linewidth 2–3 cm<sup>-1</sup>, ~5–10 mJ/pulse) was provided by the idler output of an OPO/OPA laser system (LaserVision), pumped by a pulsed Nd/YAG laser (Continuum, Surelite II). Several spectra, typically  $\geq 5$ , were recorded and averaged to achieve acceptable signal/noise levels.

## ■ RESULTS AND DISCUSSION

**Vibrational Spectroscopy at Low Temperature: An Overview. Experimental Measurements in the Gas Phase.** The IRID spectra of phenyl  $\beta$ -cellobioside and its hydrated complexes, phenyl  $\beta$ -cellobioside·(D<sub>2</sub>O)<sub>1,2</sub>, recorded in a molecular beam environment, are shown in Figure 2. The spectra remained essentially the same when the probe UV laser was tuned to different regions of their respective R2PI spectra, reflecting the population, in each case, of a single, predominant structure. Figure 2 also shows their calculated (DFT//MP2) minimum energy structures and the corresponding vibrational spectra; the IRID and the calculated spectra are all in good accord.

The cis (anti- $\phi$ , syn- $\psi$ ) conformation of the bare molecule is supported by two inter-ring hydrogen bonds, OH<sub>6'</sub>→O<sub>6</sub> and OH<sub>2'</sub>→O<sub>3</sub>. This bonding is retained in the monohydrate, phenyl  $\beta$ -cellobioside·D<sub>2</sub>O and its well-resolved vibrational spectrum confirms the earlier structural assignment based on the IRID spectrum of the corresponding H<sub>2</sub>O complex.<sup>3</sup> The reversed orientation of the peripheral OH groups from counter clockwise to clockwise, provides a “water binding pocket” at the 4',6' site where the bound water molecule can insert to complete the cooperative hydrogen-bonded chain, OH<sub>2</sub>→OH<sub>3</sub>→OH<sub>2'</sub>→OH<sub>3'</sub>→OH<sub>4'</sub>→W→OH<sub>6'</sub>→OH<sub>6</sub> (recalling the behavior of the singly hydrated monosaccharide, phenyl  $\beta$ -D-glucopyranoside·H<sub>2</sub>O).<sup>42</sup>

It might be thought that the addition of a second water molecule would follow a similar rule. In doubly hydrated phenyl  $\beta$ -D-glucopyranoside, the two water molecules add sequentially with the first occupying the 4,6 site as before and the second located in the vacancy on the other side of the hydroxymethyl group at the 6,5 site.<sup>43</sup> In phenyl  $\beta$ -cellobioside, the corresponding 6',5' site would be blocked by the inter-ring hydrogen bond, OH<sub>6'</sub>→OH<sub>6</sub>, though the 6,5 site on the neighboring ring could provide an alternative. The DFT calculations do indeed predict two such structures, but at energies 4.3 and 4.9 kJ mol<sup>-1</sup> above that of the global minimum structure and the vibrational spectrum calculated for the latter structure is in very good accord

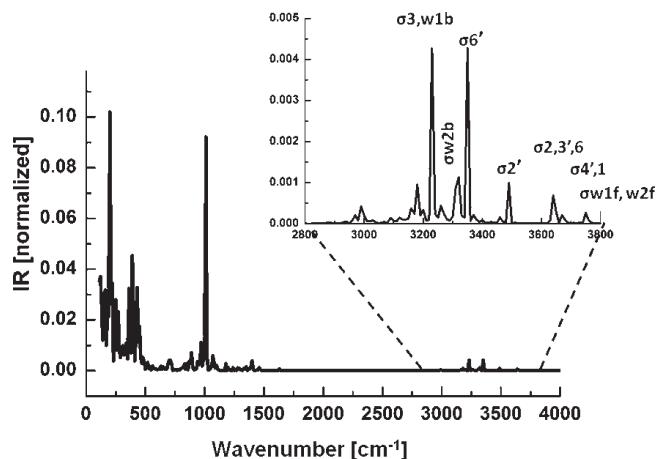


Figure 7. Vibrational spectrum and OH band assignments of cellobiose·(H<sub>2</sub>O)<sub>2</sub> calculated from 5 ps MD/CP2K trajectories at ~40 K.

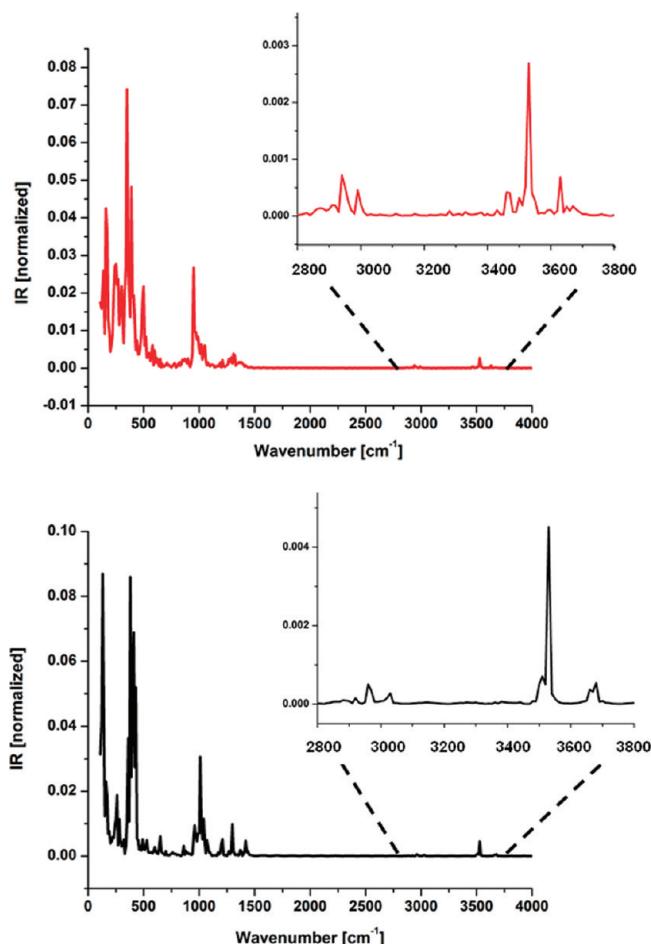
Table 5. Comparison between the Vibrational Wavenumbers of Doubly Hydrated, Phenyl  $\beta$ -Cellobioside·(D<sub>2</sub>O)<sub>2</sub> Measured Experimentally in a Molecular Beam at T ~ 10 K and the Corresponding Wavenumbers of Cellobiose·(D<sub>2</sub>O)<sub>2</sub> and Cellobiose·(H<sub>2</sub>O)<sub>2</sub> Calculated from MD Trajectories

OH vibrational mode	vibrational wavenumber/cm <sup>-1</sup> experimental: IRID	vibrational wavenumber/cm <sup>-1</sup> calculated: T ~ 40 K	phenyl $\beta$ -cellobioside·(D <sub>2</sub> O) <sub>2</sub>	$\beta$ -cellobiose·(D <sub>2</sub> O) <sub>2</sub> {·(H <sub>2</sub> O) <sub>2</sub> }
$\sigma 2'$	~3450	3490	{3490}	
$\sigma 3'$	~3620	3640	{3640}	
$\sigma 4'$	~3620	3680	{3680}	
$\sigma 6'$	~3400	3350	{3350}	
$\sigma 1$	N/A	3670	{3670}	
$\sigma 2$	~3620	3650	{3630}	
$\sigma 3$	~3300	3220	{3230}	
$\sigma 6$	~3620	3650	{3650}	
$\sigma w1b, \sigma w2b$	~2500	2320, 2390 <sup>a</sup>	{3230, 3310}	
$\sigma w1f, \sigma w2f$	~2720	2730, 2700 <sup>a</sup>	{3760, 3720}	

<sup>a</sup> Each pair of bands has additional weaker peaks that include each other's main peak.

with the experimental IRID spectrum; see Figure 2. Instead of binding two separate molecules, the cellobiose unit accommodates a water dimer, which links O<sub>6'</sub> (acting as the hydrogen bond acceptor) and OH<sub>3</sub> (acting as the donor) to provide a diagonal bridge across the two rings. The glycosidic conformation of the cellobiose unit remains cis, however, and its framework structure is very similar to that of the bare molecule, again supported by two inter-ring hydrogen bonds, OH<sub>6'</sub>→O<sub>6</sub> and OH<sub>2'</sub>→OH<sub>3</sub>.

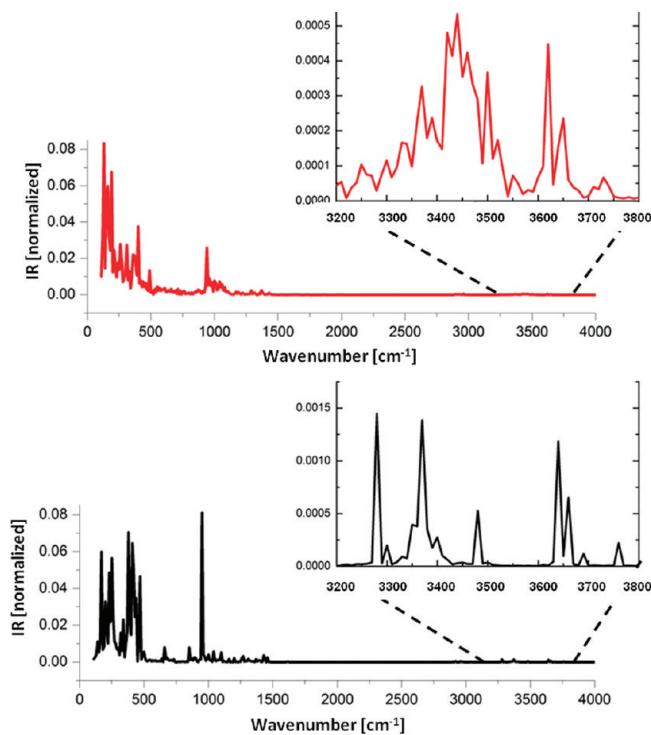
The advantage of using D<sub>2</sub>O rather than H<sub>2</sub>O, to separate the hydrate (OD) region of the IRID spectrum from the crowded OH region associated with the carbohydrate, is again well demonstrated by the IRID spectrum of phenyl  $\beta$ -lactoside·D<sub>2</sub>O shown in Figure 3. The complexity of the congested spectrum associated with the corresponding H<sub>2</sub>O complex reported earlier<sup>3</sup> appeared to be incompatible with its assignment to a single structure. The new spectrum confirms the earlier tentative



**Figure 8.** Predicted temperature dependence of the vibrational spectrum of cellobiose, calculated from MD trajectories (2.5 ps) obtained at  $\sim 300$  (red) and  $\sim 22$  K (black). The spectra on the right-hand side at higher wavenumber focus on the OH and CH modes.

conclusion that it is generated by the two lowest energy structures, shown in Figure 3, which are both populated in the cold molecular beam. The two bands at  $\sim 2420$  and  $\sim 2530\text{ cm}^{-1}$  are associated with the bound OD modes, labeled  $\sigma_{\text{Wb}}$  in Figure 3. In the former case, the OD is bound to O $6'$  on the exocyclic hydroxymethyl group, which is then bound, via OH acting as a donor, to O $6'$  through an inter-ring hydrogen bond. In the latter case, the (lactose) OH groups adopt a reversed, clockwise orientation, and OH $6$  becomes a hydrogen bond donor to D $_2$ O and there is a (weaker) bond linking OD to O $5$ .

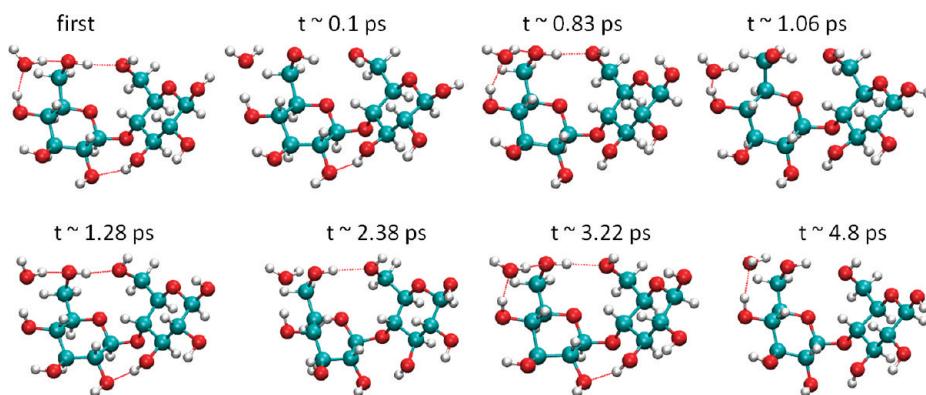
**VSCF Calculations and “On-the-Fly” Dynamical Simulations: Cellobiose and Lactose.** The computed vibrational spectra shown in Figures 2 and 3 are all based upon harmonic frequencies, scaled empirically to correct for anharmonicity (and any other effects) and bring them into better correspondence with the experimental spectra. A more rigorous approach takes proper account of anharmonicity by running VSCF calculations based upon anharmonic potentials, for example, the semiempirical PM3 potential or better, a Hartree–Fock potential, “upgraded” by reference to the normal modes predicted by high level MP2 calculations. Because the VSCF method predicts anharmonic vibrational spectra at  $T = 0$  K, the accuracy, or otherwise, of the force fields it employs can be validated through comparison with the corresponding experimental data obtained at low



**Figure 9.** Predicted temperature dependence of the vibrational spectrum of cellobiose·H<sub>2</sub>O, calculated from MD trajectories obtained at  $\sim 300$  (red) and  $\sim 40$  K (black). The expanded spectra at higher wavenumber focus on the OH modes only.

temperatures in the gas phase.<sup>18</sup> Alternatively, the vibrational spectra can be calculated through CP2K simulations (conducted at a suitably low temperature) based on DFT potentials, including dispersion<sup>19,20</sup> or when the systems are too large, on semiempirical potentials. Fourier transformation of the system dipole moment correlation functions obtained along the classical MD trajectories generates vibrational spectra associated with the given force field. Again, their accuracy can be assessed through comparison with experiment but, importantly in the present context, they also provide a way of predicting the spectroscopic and structural consequences of raising the temperature, potentially allowing this contribution to the cellobiose unit conformation in cellulose to be probed, at least theoretically.

Figure 4 compares the results of an MD simulation (at  $T \sim 40$  K) of the vibrational spectrum of cellobiose obtained through Car–Parrinello calculations using a DFT (BLYP-D) potential, and the corresponding predictions of a hybrid VSCF ( $T = 0$  K) calculation, based upon a PM3 potential and scaled against a high level MP2 calculation for  $\beta$ -glucose. Reassuringly, the MD band analysis accords well with the VSCF normal mode computations. There is also a good correlation and indeed quite close quantitative agreement between the OH vibrational wavenumbers (for  $\beta$ -cellobiose) computed at  $\sim 40$  K and the experimental wavenumbers of the corresponding bands in phenyl  $\beta$ -cellobioside recorded at  $\sim 10$  K, see Table 1. These results support an earlier conclusion: the associated chromophoric tags do not significantly perturb either the structure of the carbohydrate or its vibrational spectrum.<sup>18</sup> More importantly, the results serve to validate the accuracy of the dispersion corrected DFT (BLYP-D) force field and the reliability of classical dynamics simulations performed at the low temperature.

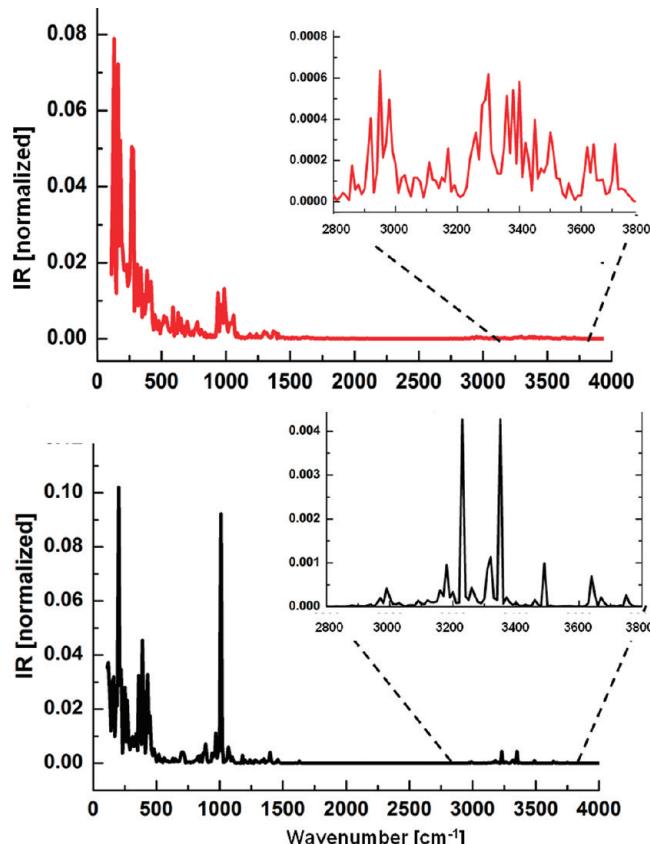


**Figure 10.** Snapshots of MD trajectories for cellobiose·H<sub>2</sub>O, calculated at  $T \sim 300$  K. Note the rapid fluctuations in both intermolecular and intramolecular hydrogen bonding, which is reflected in the increasing complexity of the associated vibrational bands and their shift to higher wavenumber; see Table 3.

Figure 5 presents the results of a similar MD simulation of the vibrational spectrum of  $\beta$ -lactose at  $\sim 40$  K, based on a DFT (BLYP-D) potential and using the structural assignment of the IRID spectrum of benzyl  $\beta$ -lactoside reported earlier.<sup>3,44</sup> Table 2 provides a comparison between these results and those obtained through experiment. Band analysis of the OH bond stretching reveals the OH<sub>6'</sub> vibration to be “split” with a main peak at  $\sim 3410\text{ cm}^{-1}$  and a secondary one at  $\sim 3370\text{ cm}^{-1}$ , enhanced by a secondary mode of OH<sub>4'</sub> at the same wave number; a third but much weaker mode is present in OH<sub>6'</sub> at  $\sim 3320\text{ cm}^{-1}$ , but it cannot explain the intensity observed in the computed IR spectrum. Again, despite the absence of the benzyl tag, the computed vibrational frequencies and those assigned to the experimental spectrum are in quite good accord.

**Car–Parrinello Calculations for Hydrated Cellobiose and Lactose: Theory and Experiment.** Previous Car–Parrinello simulations<sup>18</sup> using DFT potentials have not included dispersion. In the present work, the introduction of the dispersion-corrected functional, BLYP-D,<sup>19,20</sup> marks an important step forward particularly for the hydrated systems where both hydrogen-bonded and dispersion interactions will contribute. Figure 6 presents the computed vibrational spectrum of singly hydrated cellobiose at  $\sim 40$  K, based on the structure assigned experimentally to phenyl  $\beta$ -cellobioside·H<sub>2</sub>O. The spectra of the OH (and also the OD) bands determined from the computed trajectories are listed in Table 3 where they can be compared with the corresponding experimental data. When the calculated modes are clearly defined by a single intense peak, for example,  $\sigma_2'$ ,  $\sigma_3'$ ,  $\sigma_2$ , and  $\sigma_{\text{wf}}$ , the comparisons are in good accord with experiment. They are somewhat poorer for the strongly hydrogen-bonded modes,  $\sigma_4'$ ,  $\sigma_6'$  and  $\sigma_{\text{wb}}$ , where the simulations generate multiple peaks in the computed spectrum. Within these limitations however, they are able to replicate successfully, the overall spectral sequence shown in Figure 2, namely the group of “spectator modes”,  $\sigma_2, 6, 2', 3'$ , lying between 3600 and 3650  $\text{cm}^{-1}$ ; the strongly hydrogen-bonded modes,  $\sigma_4', 6'$ , shifted to lower wavenumber by  $\sim 150$ – $200\text{ cm}^{-1}$ ; the less strongly hydrogen bonded mode,  $\sigma_3$ , which lies between these two groups; and the two bands,  $\sigma_{\text{wb}}, \text{wf}$ , associated with the bound and free D<sub>2</sub>O (or H<sub>2</sub>O) stretch modes.

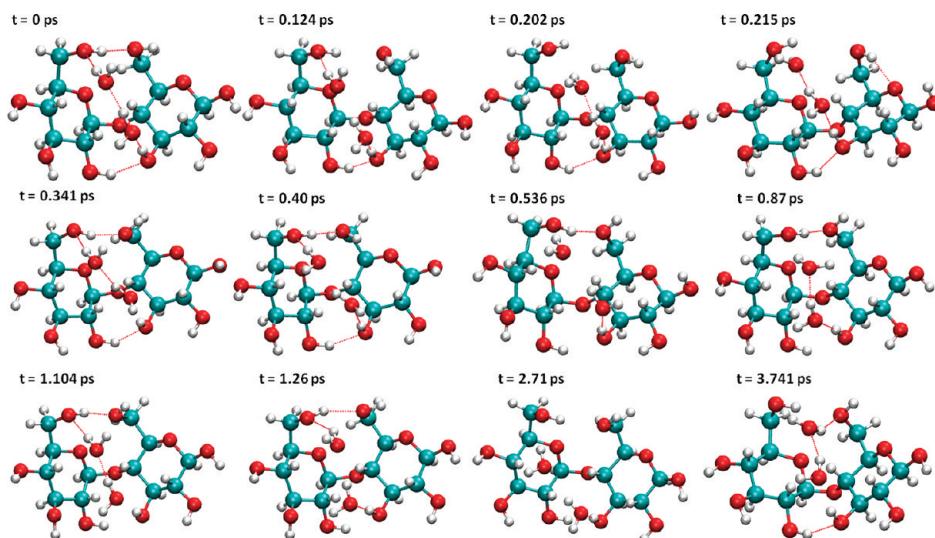
Similar considerations apply to the simulations conducted for singly hydrated  $\beta$ -lactose at  $T \sim 40$  K. The comparisons between the OH or OD frequencies determined from the MD trajectories and those of benzyl  $\beta$ -lactoside·H<sub>2</sub>O and benzyl  $\beta$ -lactoside·D<sub>2</sub>O



**Figure 11.** Predicted temperature dependence of the vibrational spectrum of cellobiose·(H<sub>2</sub>O)<sub>2</sub>, calculated from MD trajectories obtained at  $\sim 300$  (red) and  $\sim 40$  K (black). The expanded spectra at higher wavenumber focus on the OH modes only.

based upon their experimental IRID spectra, are summarized in Table 4. The correspondence between the assignments based on the alternative sets of data, which in one case is purely theoretical and in the other is based on experiment, is largely sustained across the full OH/OD vibrational spectrum.

The OH spectrum generated by a Car–Parrinello simulation of the dihydrate, cellobiose·(H<sub>2</sub>O)<sub>2</sub> at  $T \sim 40$  K, is shown in Figure 7, together with the vibrational assignments determined



**Figure 12.** Snapshots of MD trajectories for cellobiose· $(\text{H}_2\text{O})_2$ , calculated at  $T \sim 300$  K.

from an analysis of the MD trajectories. The structure of the dihydrate corresponded to that of the cellobiose unit in phenyl  $\beta$ -cellobioside· $(\text{D}_2\text{O})_2$ , determined experimentally (see Figure 2), that is, with the two water molecules bound as a dimer, bridging diagonally across the pyranose rings, between  $\text{O}6'$  and  $\text{OH}3$ . MD calculations were also conducted for cellobiose· $(\text{D}_2\text{O})_2$  and both sets of data are summarized in Table 5. There is a high degree of accord between the experimental assignments and frequencies for the OH and OD modes in phenyl  $\beta$ -cellobioside· $(\text{D}_2\text{O})_2$  and the corresponding MD predictions for doubly hydrated cellobiose, as there was for the monohydrates of cellobiose and lactose. In all three cases, the comparisons between experiment and theory have been able to validate the MD calculations and the accuracy of the BLYP-D potential and also confirm the benign nature of the phenyl or benzyl “tag”.

## ■ VIBRATIONAL SPECTRA AND STRUCTURAL DYNAMICS AT ELEVATED TEMPERATURES

The inter-ring hydrogen bonding, which supports the intrinsic cis conformation of cellobiose at low temperature in the gas phase, creates an inherently rigid structure. If hydrogen bonding were reduced or more labile at more elevated temperatures, however, this structure might relax into other forms. MD simulations to explore this possibility were conducted for cellobiose and its singly and doubly hydrated complexes based in each case upon the structures determined experimentally at low temperature,  $T \sim 10$  K. The results are shown in Figures 8 to 12. In cellobiose, the rise in temperature did excite some additional OH (and to a lesser extent CH) motion (see Figure 8), which led to brief loss and reconstitution of intramolecular OH–O hydrogen bonds, but nonetheless, the sharp spectral features were retained. In the gas phase, cellobiose preserves its rigid cis conformational structure at temperatures  $\leq 300$  K; this remained the case even when calculations were run at 600 K, although there was some torsional motion of the glycosidic bond. This indicates perhaps that the cis to trans transition is governed by the barrier between the minima rather than their energy differences.

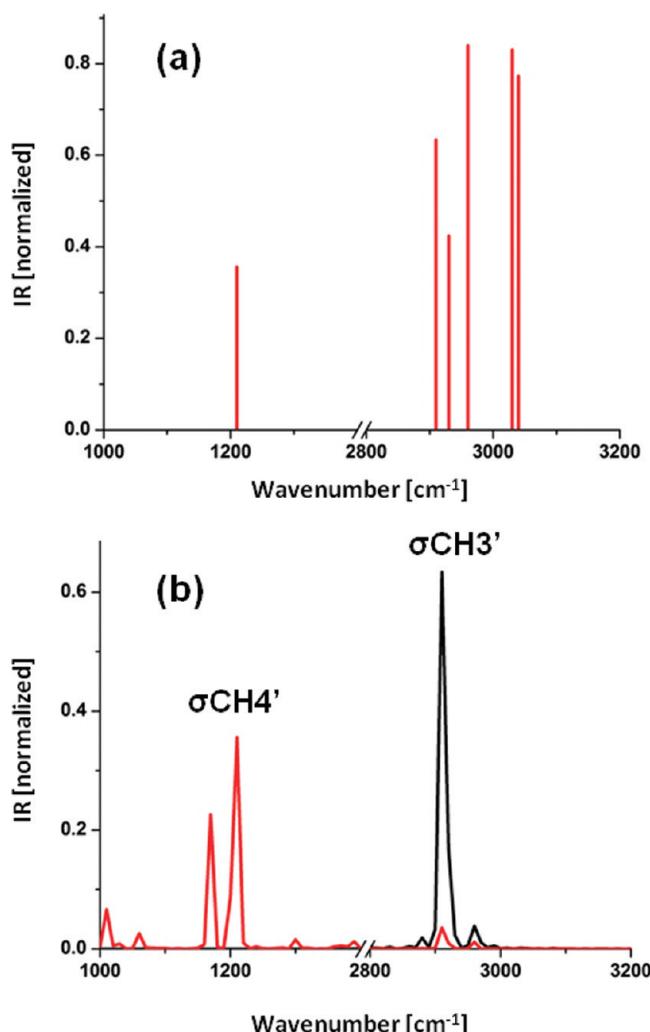
The same is broadly true for the monohydrate, but the OH motions, particularly those involving intermolecular interactions with the water molecule, were much more strongly excited at

300 K; see Figure 9. At 40 K, little changed over a period  $\sim 5$  ps. The water molecule remained in its minimum energy location, forming a stable bridge between  $\text{OH}4'$  and  $\text{O}6'$  and, with the exception of  $\text{OH}6'$  and  $\text{OH}3$  (involved in the inter-ring hydrogen-bonding), the other (“spectator”) OH groups are also stable. The  $\text{OH}6' \rightarrow \text{O}6$  and  $\text{OH}3 \rightarrow \text{O}2'$  hydrogen bonds occasionally cleave and reform during the observation period, reflected in the signatures of their associated vibrational modes,  $\sigma_{6'}$  and  $\sigma_3$ , which present a band spectrum characterized by two or more peaks of comparable intensity. At 300 K, the sharp spectator bands in the OH spectrum lying above  $3600 \text{ cm}^{-1}$  show little change but the bands lying below  $3500 \text{ cm}^{-1}$ , associated with the hydrogen-bonded OH groups, merge into a complex broad feature peaking between  $3400$  and  $3450 \text{ cm}^{-1}$ . The structural changes that these reflect are presented graphically in the series of snapshots taken along the MD trajectories, shown in Figure 10. They include changes in rotamer configuration and the loss/reconstitution of both internal hydrogen bonds and those involving the water molecule, but throughout these events the global cis glycosidic conformation does not change.

The Car–Parrinello simulations conducted for cellobiose· $(\text{H}_2\text{O})_2$  at  $T \sim 300$  K predict very similar behavior; see Figures 11 and 12. The structure remains very rigid at  $T \sim 40$  K; at the elevated temperature, despite the evident increase in the complexity of the OH (and CH) vibrational spectra, the structure of the dihydrate is still broadly retained. Various hydrogen bonds can be seen to cleave and reform in the series of snapshots shown in Figure 12, but although the water molecules begin to move and rotate the  $(\text{H}_2\text{O})_2$  bridge is maintained throughout the entire trajectory and so too, as before, is the cis configuration about the glycosidic linkage.

## ■ CH MODE COUPLING

In phenyl  $\beta$ -cellobiose, the CH vibrational bands associated with the cellobiose unit occupy a range of  $\sim 100 \text{ cm}^{-1}$  centered around  $2900 \text{ cm}^{-1}$  (the CH bands at higher wavenumber, between  $3000$  and  $3100 \text{ cm}^{-1}$ , are associated with the phenyl group); see Figure 4. The most intense CH bands predicted by the MD simulations at  $\sim 40$  K for cellobiose are shown in Figure 13. Interestingly, while most of them show single sharp peaks, one, associated with the  $\text{CH}4'$  mode, has its largest contribution at a



**Figure 13.** CH bands in cellobiose predicted from CP2K trajectories at  $T \sim 40$  K using a DFT BLYP-d3 potential. (a) A selection of the most intense bands, (b) a comparison between  $\sigma\text{CH}_3'$  and the strongly displaced mode,  $\sigma\text{CH}_4'$ .

much lower wavenumber,  $\sim 1200 \text{ cm}^{-1}$ , a region dominated by the heavy atom CC and CO bond stretches, and by angle bends and torsions (see Figure 4). This is shown in Figure 13b, which compares the band structures calculated for  $\sigma\text{CH}_3'$  and the displaced band,  $\sigma\text{CH}_4'$ . The strong displacement indicates strong mode coupling and perhaps a particular ability for CH groups in saccharides to absorb energy from neighboring, or collective modes.

## ■ CONCLUDING REMARKS

The reliability and accuracy of Car–Parrinello simulations of the vibrational spectra of cellobiose, one of the fundamental building blocks in nature, and its C4' epimer, lactose, have been demonstrated through comparisons with experimental measurements conducted at low temperature in the gas phase. They reinforce the results of similar comparisons for glucose that were based on simulations employing a DFT (BLYP) potential and necessarily, also included the role of anharmonicity.<sup>18</sup> The structural consequences of hydrating the cellobiose (and lactose) unit, which have been established experimentally, have also been confirmed by further simulations using dispersion corrected BLYP-D potentials and this success has enabled an exploratory

of their structural dynamics through simulations conducted at increasing temperature. The spectroscopic and structural measurements were greatly aided by substituting D<sub>2</sub>O for H<sub>2</sub>O, which separated the water (OD) vibrational bands from the disaccharide (OH) bands and helped to disentangle their vibrational spectra.

The water molecules in the dihydrate do not add sequentially at different sites but are bound as a dimer, which forms a diagonal bridge linking the OH3 site on one pyranose ring to the O6' site on the other. The unexpected formation of this new hydration motif contrasts with the successive binding observed in the microhydration of the monosaccharide unit, glucopyranose,<sup>43</sup> and assumed in earlier DFT calculations of microhydrated cellobiose.<sup>2</sup> (Coincidentally, a recent Car–Parrinello simulation<sup>12</sup> of methyl  $\beta$ -cellobioside in aqueous solution also indicated a high probability of finding bridging water molecules located between OH3 and O6' but bridging across a trans glycosidic configuration.)

The isolated and hydrated cellobiose and lactose units both present remarkably rigid structures; their glycosidic linkages adopt a cis (anti- $\phi$  and syn- $\psi$ ) conformation bound by interring hydrogen bonds and this conformation is maintained when the temperature is increased to  $\sim 300$  K. It continues to be maintained when the cellobiose (or lactose) unit is hydrated by one or two explicitly bound water molecules. At 300 K, despite individual fluctuations in the intra- and intermolecular hydrogen bonding pattern and some local structural motions, the water molecules remain locally bound and the isolated carbohydrates remain trapped within the “cis” potential well. These MD simulations do not suggest any accessible pathway to any trans conformers, yet they are widespread in nature.

Could the preference for a trans conformation in an aqueous solution be associated, at least in part, with its polarity? DFT/MP2 calculations for the water-bridged cis dihydrate, phenyl  $\beta$ -cellobioside·(H<sub>2</sub>O)<sub>2</sub>, predict a low dipole moment, 1.6 D, but the lowest energy doubly hydrated trans structure, (relative energy  $\sim 16 \text{ kJ mol}^{-1}$ ) which now has the two water molecules located separately, inserted at the 2',6 and 6,5 sites, has a dipole moment of 6.3 D. Assuming a mean-field approximation and using Onsager's PCM model, its relative electrostatic stabilization energy would be  $\sim 8 \text{ kJ mol}^{-1}$ , a change in the right direction; its relative free energy at 300 K is estimated to fall still further to  $\sim 1 \text{ kJ mol}^{-1}$  but this is still a long way short of the results of MD calculations<sup>11</sup> of the relative free energies in aqueous solution (as opposed to the gas phase). These place the cis conformation  $\sim 16 \text{ kJ mol}^{-1}$  below the trans at 300 K and predict a low barrier,  $\Delta G^\ddagger \sim 5 \text{ kJ mol}^{-1}$ , for the cis  $\rightarrow$  trans conversion.<sup>11</sup> Perhaps yet further hydration in the gas phase might move the system beyond the tipping point but for the moment, the issue remains an open one. The retention, of the cis conformation in the isolated  $\beta$ -cellobiose motif, even when it binds two water molecules, deepens the mystery of its assuming a trans conformation when incorporated as the basic building block in  $\beta$ -cellulose, implying a need for the intervention, perhaps, of the biological catalyst promoting cellulose biosynthesis.

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*(J.P.S.) Tel: +44 1865 275400. Fax: +44 1865 275410. E-mail: john.simons@chem.ox.ac.uk.

\*(R.B.G.) Tel: +972 2 658 5732. Fax: +972 2 651 3742. E-mail: benny@fhs.huji.ac.il.

## ■ ACKNOWLEDGMENT

We appreciate the generous support provided by the U.S. Department of Energy Office of Science, Grant DE-FG02-09ER64762; the NERSC computational facilities and the expert assistance of their support staff (R.B.G.); EPSRC for an LSI Platform grant (B.G.D.); and the Laser Support Facility of the STFC (J.P.S.). We also thank the Leverhulme Trust for the award of an Emeritus Fellowship (J.P.S.) and the Royal Society for a Wolfson Research Merit Award (B.G.D.). E.J.C. thanks the Spanish ministry of science and innovation (MICINN) for funds (CTQ2009-14364, 2010/CSD2007-00013) and also the MICINN for a “Ramón y Cajal” contract. Last but not least we thank David Gamblin, who synthesized the phenyl $\beta$ -cellobioside samples and Lukasz Cwiklik and Pavel Jungwirth for their help in implementing the Grimme, 2006 (D2) correction.

## ■ REFERENCES

- (1) (a) French, A. D.; Johnson, G. P. *Can. J. Chem.* **2006**, *84*, 603.  
(b) French, A. D.; Johnson, G. P. *Mol. Simul.* **2008**, *34*, 365.
- (2) (a) Strati, G. L.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2002**, *337*, 1833. (b) Bosma, W. B.; Appell, M.; Willett, J. L.; Momany, F. A. *J. Mol. Struct., THEOCHEM* **2006**, *776*, 1.
- (3) Cocinero, E. J.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *J. Am. Chem. Soc.* **2009**, *131*, 11117.
- (4) Lipkind, G. M.; Shashkov, J. S.; Kochetkov, N. K. *Carbohydr. Res.* **1985**, *141*, 191.
- (5) Asensio, J. L.; Jiménez-Barbero, J. *Biopolymers* **1995**, *35*, 55.
- (6) Cheetham, N. W. H.; Dasgupta, P.; Ball, G. E. *Carbohydr. Res.* **2003**, *338*, 955.
- (7) Larsson, E. A.; Staaf, M.; Söderman, P.; Höög, C.; Widmalm, G. *J. Phys. Chem. A* **2004**, *108*, 3932.
- (8) (a) Pereira, C. S.; Kony, D.; Baron, R.; Müller, M.; van Gunsteren, W. F. *Biophys. J.* **2006**, *90*, 4337. (b) **2007**, *93*, 706.
- (9) Olsson, U.; Serianni, A. S.; Stenutz, R. *J. Phys. Chem. B* **2008**, *112*, 4447.
- (10) Zhang, W.; Zhao, H.; Carmichael, I.; Serianni, A. S. *Carbohydr. Res.* **2009**, *344*, 1582.
- (11) (a) Perić-Hassler, L.; Hansen, H. S.; Baron, R.; Hüenberger, P. H. *Carbohydr. Res.* **2010**, *345*, 1781. (b) Hatcher, H.; Säwén, E.; Widmalm, G.; MacKerell, A. D. *J. Phys. Chem. B* **2011**, *115* (3), 597–608.
- (12) Christensen, N. J.; Hansen, P. I.; Larsen, F. H.; Folkerman, T.; Motawia, M. S.; Engelsen, S. B. *Carbohydr. Res.* **2010**, *345*, 474.
- (13) Chu, S. S. C.; Jeffrey, G. A. *Acta Crystallogr.* **1968**, *B24*, 830.
- (14) Ham, J. T.; Williams, D. G. *Acta Crystallogr., Sect B* **1970**, *26*, 1373.
- (15) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300.
- (16) Saxema, I. M.; Brown, R. M., Jr. *Ann. Bot.* **2005**, *96*, 9.
- (17) Marx, D.; Hutter, J. Ab initio molecular dynamics: Theory and Implementation. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, 2000; NIC Series, Vol. 1, pp 301–449.
- (18) Brauer, B.; Pincu, M.; Buch, V.; Bar, I.; Simons, J. P.; Gerber, R. B. *J. Phys. Chem. A* [Online early access]. DOI: 10.1021/jp110043k.
- (19) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.
- (20) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
- (21) Brauer, B.; Chaban, G. M.; Gerber, R. B. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2543.
- (22) Pele, L.; Brauer, B.; Gerber, R. B. *Theor. Chem. Acc.* **2007**, *117*, 69.
- (23) Chaban, G. M.; Jung, J. O.; Gerber, R. B. *J. Chem. Phys.* **1999**, *111*, 1823.
- (24) Adesokan, A.; Fredj, E.; Brown, E. C.; Gerber, R. B. *Mol. Phys.* **2005**, *103*, 1505.
- (25) Adesokan, A. A.; Gerber, R. B. *J. Phys. Chem. A* **2009**, *113*, 1905.
- (26) Jung, J. O.; Gerber, R. B. *J. Chem. Phys.* **1996**, *105*, 10682.
- (27) Gaigeot, M.-P.; Sprik, M. *J. Phys. Chem. B* **2003**, *107*, 10344.
- (28) Gaigeot, M.-P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3336.
- (29) de Vondele, V.; Krack, M.; Mohammed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103.
- (30) CP2K (QUICKSTEP) website, <http://cp2k.berlios.de>. Accessed May 1, 2011.
- (31) Goedecker, S.; Teter, M.; Hutter, M. *Phys. Rev. B: Condens. Matter* **1996**, *54*, 1703.
- (32) Vacha, R.; Cwiklik, L.; Rezac, J.; Hobza, P.; Jungwirth, P.; Valsaraj, K.; Bahr, S.; Kemper, V. *J. Phys. Chem. A* **2008**, *112*, 4942.
- (33) Gaigeot, M.-P.; Martinez, M.; Vuilleumier, R. *Mol. Phys.* **2007**, *105*, 2857.
- (34) Buch, V.; Milet, A.; Vacha, R.; Jungwirth, P.; Devlin, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7342.
- (35) Chaban, G. M.; Gerber, R. B. *Theor. Chem. Acc.* **2008**, *120*, 381.
- (36) Martyna, G. J.; Tuckerman, M. E. *J. Chem. Phys.* **1999**, *110*, 2810.
- (37) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrikson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.
- (38) Frisch, M. J.; et al. *Gaussian 03*, Revision B.03; Gaussian, Inc.: Pittsburgh, PA, 2003.
- (39) Bouteiller, Y.; Gillet, J. C.; Grégoire, G.; Schermann, J.-P. *J. Phys. Chem. A* **2008**, *112*, 11656.
- (40) Zwier, T. S. *J. Phys. Chem. A* **2006**, *110*, 4133.
- (41) Robertson, E. G.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1.
- (42) Cocinero, E. J.; Stanca-Kaposta, E. C.; Scanlan, E. M.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *Chem.—Eur. J.* **2008**, *14*, 8947.
- (43) Cocinero, E. C.; Stanca-Kaposta, E. C.; Dethlefsen, M.; Liu, B.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *Chem.—Eur. J.* **2009**, *15*, 13427.
- (44) 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.