

## Martini Coarse-Grained Force Field: Extension to Carbohydrates

Cesar A. López,<sup>†</sup> Andrzej J. Rzepiela,<sup>†</sup> Alex H. de Vries,<sup>†</sup> Lubbert Dijkhuizen,<sup>‡</sup>  
Philippe H. Hünenberger,<sup>§</sup> and Siewert J. Marrink\*,<sup>†</sup>

Groningen Biomolecular Sciences and Biotechnology Institute & Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), Centre for Carbohydrate Bioprocessing, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands, and Laboratorium für Physikalische Chemie, ETH Zürich, CH-8093 Zürich, Switzerland

Received June 21, 2009

**Abstract:** We present an extension of the Martini coarse-grained force field to carbohydrates. The parametrization follows the same philosophy as was used previously for lipids and proteins, focusing on the reproduction of partitioning free energies of small compounds between polar and nonpolar phases. The carbohydrate building blocks considered are the monosaccharides glucose and fructose and the disaccharides sucrose, trehalose, maltose, cellobiose, nigerose, laminarabiose, kojibiose, and sophorose. Bonded parameters for these saccharides are optimized by comparison to conformations sampled with an atomistic force field, in particular with respect to the representation of the most populated rotameric state for the glycosidic bond. Application of the new coarse-grained carbohydrate model to the oligosaccharides amylose and Curdlan shows a preservation of the main structural properties with 3 orders of magnitude more efficient sampling than the atomistic counterpart. Finally, we investigate the cryo- and anhydro-protective effect of glucose and trehalose on a lipid bilayer and find a strong decrease of the melting temperature, in good agreement with both experimental findings and atomistic simulation studies.

### 1. Introduction

Carbohydrates (saccharides), the most abundant product of photosynthesis, play an important role in the energetic metabolism of living species and the signaling and immunological responses and are a fundamental component of the external cell wall of many organisms.<sup>1</sup> In addition, saccharides are present in a variety of emerging classes of biomimetic materials.<sup>2</sup> Furthermore, due to their cryo- and anhydro-protective properties, many sugars have been shown to be effective stabilizers of biological components, such as proteins and membranes, in the low-temperature or dehydrated states.<sup>3–5</sup> This class of compounds encompasses a huge variety of possible monomeric units (differing in

stereochemistry and functionalization) that can be connected in chains presenting a virtually infinite number of possible residue sequences, linkage types, and degrees of branching. Despite their importance, the experimental characterization of the structural and dynamical properties of oligosaccharides in general has proven rather problematic. Unlike proteins, nucleic acids, and lipids, which tend to predominantly adopt a well-defined (native) conformation under the conditions where they are biologically functional, carbohydrates are typically associated with a high extent of conformational heterogeneity. As a result of this structural diversity and conformational heterogeneity, carbohydrates arguably represent the most challenging class of biomolecules in terms of experimental characterization and elucidation of structure–function relationships.<sup>6–8</sup> Static structures of carbohydrates may often be obtained from (X-ray) crystallography (of crystals or fibers), but it is always uncertain whether these

\* Corresponding author e-mail: s.j.marrink@rug.nl.

<sup>†</sup> Zernike Institute for Advanced Materials.

<sup>‡</sup> Microbiology Department, Haren.

<sup>§</sup> ETH Zürich.

molecules adopt similar conformations in solution as in the solid state. On the other hand, nuclear magnetic resonance (NMR) spectroscopy provides information about carbohydrates in solution but only in the form of averages over all the populated conformational states present in solution (i.e., over all the molecules or molecular segments in the sample as well as over the time scale of the NMR experiment). Many other experimental techniques (e.g., electron microscopy, light or neutron diffraction, circular dichroism, infrared spectroscopy, or rheology) provide useful but even more indirect information about carbohydrate conformations.

Molecular dynamics (MD) simulations can, in principle, provide the link between structure and physical properties that are more readily measured (i.e., radius of gyration, gel transition temperature; mostly concerning long polymers<sup>1,9–13</sup>). Many force fields have been extensively parametrized for carbohydrates<sup>14–19</sup> and have been used to provide details of the structure and dynamics at an all-atom (AA) level, for example, to explore the ring puckering of glucose,<sup>20–22</sup> conformational changes in disaccharides and trisaccharides,<sup>14,23–25</sup> and stability of oligosaccharides like amylose and Curdlan.<sup>9,26–28</sup> However, such studies are necessarily limited to small system sizes with a limited sampling of the potentially very large conformational space. Simulations of longer oligosaccharides or the association of these in colloidal systems are very challenging at the AA level.

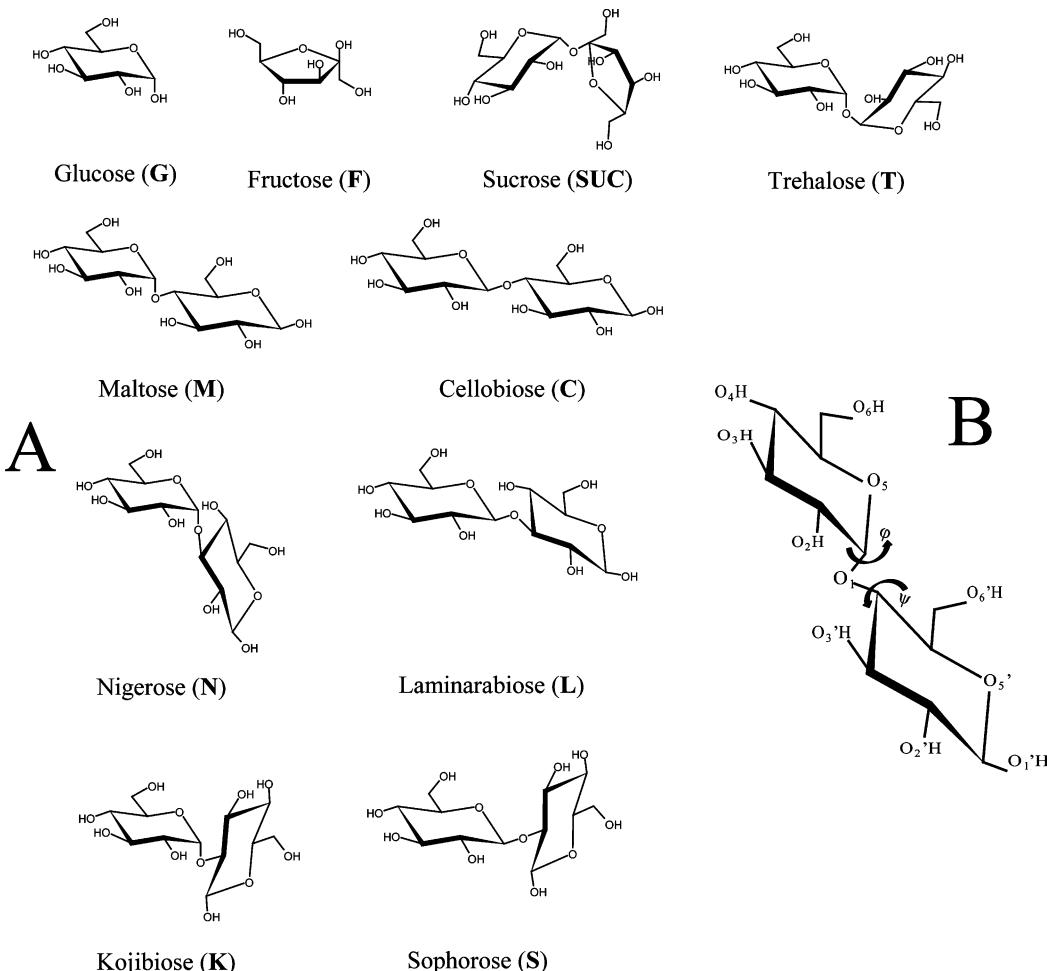
An alternative to the AA approach is the use of coarse-grained (CG) force fields, which provide a useful methodology to study large systems on a long time scale at reasonable computational cost. CG models can capture the most fundamental physical and chemical properties after averaging out some of the atomistic information, both spatially and temporally. A large diversity of CG approaches for biomolecular systems is available. They range from qualitative, solvent-free models to models including chemical specificity.<sup>29</sup> Most of the effort has been canalized into the development of models for the simulation of proteins and lipids. The design of reliable coarse-grained models for carbohydrates is hindered by their high structural diversity and the limited amount of experimental data available. Pioneering efforts in the context of glucose-based carbohydrates have been undertaken by Liu et al.<sup>30</sup> and Molinero et al.,<sup>31</sup> in which the glucose ring is represented by three particle types and its behavior in water is optimized with respect to simulations at the AA level. Bonded interaction potentials are obtained by Boltzmann inversion of the distributions of the bonds, angles, and torsion from atomistic simulations of glucose. Amylose chains have been used as a test model in both cases, revealing excellent agreement with experimental data. Another possible approach is the one adopted by Bathe et al.,<sup>32</sup> in which explicit atom models of isolated disaccharides are used to generate pretabulated potentials of mean force for the glycosidic torsions. In addition, electrostatic and steric interactions between nonadjacent residues were included, making use of virtual sites. The model was optimized for application to glycosaminoglycans. In spite of some promising results, these approaches cannot be easily extended to other systems without a full reparameterization or be used in combination with other (bio)molecules. A more

general force field for CG simulations has been developed by one of us,<sup>33–35</sup> coined the Martini force field. It is based on a four-to-one mapping scheme, implying that on average four heavy atoms and associated hydrogens are represented as a single CG site. The Martini model has been parametrized extensively by using a chemical building block principle. Its key feature is the reproduction of thermodynamic data, especially the partitioning of the building blocks between polar and nonpolar phases. It has been successfully applied to a range of lipid and protein systems.<sup>36–38</sup>

In this work, we extend the Martini force field to include carbohydrates. We base our parametrization on the conformational sampling of small carbohydrates with our CG model in comparison to AA simulations. In addition, octanol/water partitioning free energies are calculated to select the appropriate CG particle types. The set of carbohydrates used in the parametrization is illustrated in Figure 1 and comprises the monosaccharides glucose (G) and fructose (F) and the disaccharides sucrose (SUC), trehalose (T), maltose (M), cellobiose (C), kojibiose (K), sophorose (S), laminarabiose (L), and nigerose (N). The group of disaccharides includes the most important sugar–sugar linkages (1–1, 1–2, 1–3, 1–4) except for the 1–6 linkage, which proved difficult to model at the CG level. To test the transferability of the parameters to oligosaccharides, simulations of two different oligomers (amylose and Curdlan) are presented and compared with their AA counterpart. Finally, the compatibility of the carbohydrate parameters with the lipid parameters in the Martini force field is tested by looking at the stability of the liquid-crystalline phase of a dipalmitoyl-phosphatidylcholine (DPPC) bilayer in the presence of glucose and trehalose solutions.

At this point it is important to stress the inherent limitations of our CG carbohydrate model. First, the level of coarsening does not allow distinction between different ring conformations; the model represents the most populated chair  $^4\text{C}_1$  puckering state. Second, no distinction is made between different anomers at the level of a reducing residue. Consequently,  $\alpha$ - and  $\beta$ -anomers are represented by the same topology. Note, however, that for the glycosidic linkage between sugar units, the  $\alpha$ - and  $\beta$ -linkages are distinguished at the CG level through the use of different angle and dihedral interaction potentials. Third, in its current state, the model can only represent a single conformation (denoted ‘syn’) for the glycosidic linkage. The more flexible 1–6 linkage is not considered at present. Finally, the model is aimed at simulations of saccharides in solution, not in a crystal state. For a more elaborate discussion of the scope and limitations of the model, we refer to the last section of the manuscript.

The rest of this article is organized as follows. The methods section is devoted to giving a detailed account of the parametrization procedure. It is followed by the results section, reporting the results obtained for mono-, di-, and oligosaccharides. A conclusive section, with limitations and outlook, ends this article.



**Figure 1.** (A) Saccharides considered in this work: glucose (Glc $\alpha$  (G)), fructose (Fru $\beta$  (F)), maltose (M; Glc $\alpha$ (1–4)Glc $\beta$ ), cellobiose (C; Glc $\beta$ (1–4)Glc $\beta$ ), kojibiose (K; Glc $\alpha$ (1–2)Glc $\beta$ ), sophorose (S; Glc $\beta$ (1–2)Glc $\beta$ ), nigerose (N; Glc $\alpha$ (1–3)Glc $\beta$ ), laminarabiose (L; Glc $\beta$ (1–3)Glc $\beta$ ), sucrose (SUC; Glc $\alpha$ (1–2)Fru $\beta$ ), and trehalose (T; Glc $\alpha$ (1–1) $\alpha$ Glc). (B) The definitions of the dihedral angles  $\varphi$  ( $O_5$ –C $_1$ –O $_1$ –C $_{n'}$ ) and  $\psi$  (C $_1$ –O $_1$ –C $_{n'}$ –C $_{n-1}$ ) are illustrated for maltose.

## 2. Computational Methods

**2.1. Model.** The Martini CG model is used for the basic parametrization of the carbohydrate force field, which is therefore fully compatible with the Martini lipid<sup>35</sup> and protein<sup>34</sup> models. In this section we provide a brief overview of the basic parametrization procedure followed for carbohydrates: definition of the mapping and parametrization of nonbonded and bonded interactions. More details about the basic Martini model can be found in the original articles.<sup>34,35</sup>

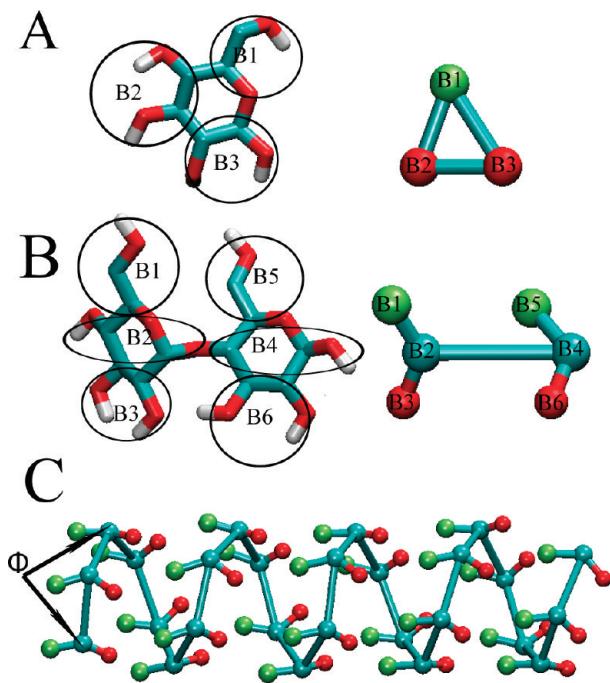
**2.1.1. Mapping of CG Sugars.** According to the mapping procedure for the Martini force field, on average four heavy particles are represented by one CG site. For a single sugar ring, consisting of 12 atoms (hydrogen atoms not counted), three particles are therefore required. This level of resolution preserves the geometrical shape of the rings (Figure 2A) and allows for a distinction between different types of monosaccharides through variations in the bond lengths, angles, and CG particle types. Disaccharides are modeled as two three-bead units connected by a single bond, which mimics the glycosidic linkage (Figure 2B). This geometry allows for the definition (and subsequent parametrization) of the glycosidic dihedral angles  $\varphi$  and  $\psi$  which determine the relative orientation of the two sugar residues and the flexibility of

the linkage. The set of fine-grained particles represented by the CG beads is chosen to be different for a monosaccharide and for the individual residues in a disaccharide. This somewhat nonobvious choice confers to the model the ability to represent the typical polar/apolar character of the disaccharides with the apolar part corresponding to the central part along the glycosidic linkage. Oligosaccharides are constructed by connecting disaccharide residues through additional bonds (Figure 2C).

**2.1.2. Parameterization of Nonbonded Interactions.** Nonbonded interactions are described by a Lennard-Jones (LJ) 12–6 potential energy function

$$U_{\text{LJ}}(r) = 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r} \right)^{12} - \left( \frac{\sigma_{ij}}{r} \right)^6 \right] \quad (1)$$

with  $\sigma_{ij}$  representing the distance at zero energy (collision diameter) between two particles  $i$  and  $j$  and  $\varepsilon_{ij}$  the strength of their interaction. The other nonbonded component of the Martini force field, the Coulomb interaction between charged particles, is not relevant for the saccharides considered in this work. The Martini model considers two different particle sizes: normal types and ring particle types, which differ in the  $\sigma_{ii}$  value of 0.47 and 0.43 nm, respectively. The



**Figure 2.** Coarse-grained mapping for mono-, di-, and oligosaccharides. (A) Monosaccharides, like glucose, are represented by three beads B1–B3. (B) Disaccharides, such as trehalose, are composed of six beads B1–B6, with the bond between B2 and B4 representing the glycosidic bond. (C) Oligosaccharides are based on the disaccharide topology but using an extra angle potential ( $\Phi$ ) for three consecutive backbone beads. Different colors are used to indicate different levels of polarity of the CG beads, see Table 3.

strength of the pairwise particle–particle interaction is determined by the value of the LJ parameter  $\varepsilon_{ij}$ . Larger values (i.e., stronger attraction) mimic polar interactions, whereas smaller values (weaker attraction) are used to mimic the hydrophobic effect. In the full interaction matrix, four main types of interaction sites are differentiated: polar (P), nonpolar (N), apolar (C), and charged (Q). The special class of ring-type particles is further denoted by the letter “S” and has a reduced value of  $\varepsilon_{ii}$ . Within a main type, subtypes are distinguished either by a letter denoting the hydrogen-bonding capabilities (d = donor, a = acceptor, da = both, 0 = none) or by a number indicating the degree of polarity (from 1 = low polarity to 5 = high polarity). Each of these particle types is representative of a specific chemical building block, i.e., inferred from a class of small compounds with similar chemical properties. The Martini force field has been parametrized extensively to reproduce the correct partitioning free energies of small molecules between a diversity of polar and apolar solvents. The full interaction matrix  $\varepsilon_{ij}$  can be found in the original publication.<sup>35</sup>

For the parametrization of the saccharides, the chemical nature of the underlying fine-grained structure is used to select the most appropriate building block and corresponding particle types. The division of the saccharides into building blocks can be done in multiple ways, however, leaving some room for adjustment. Therefore, the partitioning free energy of monosaccharides and disaccharides between water and octanol has been computed to fine tune the appropriate particle-type selection for the nonbonded interactions. An-

other alternative concerns the use of the normal particle type versus the ring particle type. On first thought, the ring particle type might seem more appropriate to model the sugar rings. However, the class of ring particle types has been parametrized based on unsubstituted ring compounds such as cyclohexane and benzene for which a four-to-one mapping is inadequate. In contrast, in the case of carbohydrate rings, the standard four-to-one mapping scheme applies (cf. Figure 2A). The properties of monosaccharide solutions over a large concentration range were used to test the two possible models. Note that the partitioning free energies do not depend on this choice, which only affects sugar–sugar interactions.

**2.1.3. Parameterization of Bonded Interactions.** Three types of bonded interactions are considered for the carbohydrates. CG particles chemically connected are described by a harmonic potential  $V_{\text{bond}}(R)$

$$V_{\text{bond}}(R) = \frac{1}{2}K_{\text{bond}}(R - R_{\text{bond}})^2 \quad (2)$$

with equilibrium distance  $R_{\text{bond}}$  and force constant  $K_{\text{bond}}$ . LJ interactions between bonded neighbors are excluded. Since the degrees of freedom are reduced at the coarse-grained level, it is necessary to preserve the topology of differently linked sugars by using both angle and dihedral potentials. A cosine-harmonic potential  $V_{\text{angle}}(\theta)$  is used for the angles

$$V_{\text{angle}}(\theta) = \frac{1}{2}K_{\text{angle}}[\cos(\theta) - \cos(\theta_0)]^2 \quad (3)$$

where  $K_{\text{angle}}$  and  $\theta_0$  are the force constant and equilibrium angle, respectively. For the dihedrals, a proper dihedral potential  $V_{\text{pd}}(\phi)$  is used with a multiplicity of 1

$$V_{\text{pd}}(\phi) = K_{\text{pd}}[1 + \cos(\phi - \phi_{\text{pd}})] \quad (4)$$

In this case,  $\phi$  denotes the angle between planes containing the beads  $i, j, k$  and  $j, k, l$ , respectively, with equilibrium angle  $\phi_{\text{pd}}$  and force constant  $K_{\text{pd}}$ .

The set of bonded parameters featured in eqs 2–4 has been parametrized by comparison to simulations of sugars at the AA level. To this end, the AA trajectories were converted to pseudo-CG trajectories using the center of mass of the appropriate fine-grained particles<sup>39</sup>

$$\mathbf{r}_i^{\text{CG}} = \frac{\sum_{j=1}^p \mathbf{r}_j \cdot m_j}{\sum_{j=1}^p m_j} \quad (5)$$

The vector  $\mathbf{r}_i^{\text{CG}}$  describes the position of the pseudo-CG bead,  $p$  is the number of atoms mapped to a given coarse bead,  $m_j$  is the mass of the atom  $j$ , and  $\mathbf{r}_j$  is its coordinates. The mapping between the AA and CG representation is shown in Figure 2. From the AA trajectory the target distribution functions were obtained for the various bonds, angles, and dihedrals considered. In a couple of iterative steps, the CG parameters were adjusted manually to obtain as close a match as possible between the pseudo-CG and real CG distributions.

**2.2. Simulation Details.** **2.2.1. System Setup.** The following saccharides were modeled (cf. Figure 1). Monosac-

charides: glucose (G;  $\alpha$ -D-glucopyranose) and fructose (F;  $\beta$ -D-fructofuranose). Disaccharides: maltose (M; Glc $\alpha$ (1-4)-Glc $\beta$ ), cellobiose (C; Glc $\beta$ (1-4)Glc $\beta$ ), kojibiose (K; Glc $\alpha$ (1-2)Glc $\beta$ ), sophorose (S; Glc $\beta$ (1-2)Glc $\beta$ ), nigerose (N; Glc $\alpha$ (1-3)Glc $\beta$ ), laminarabiose (L; Glc $\beta$ (1-3)Glc $\beta$ ), sucrose (SUC; Glc $\alpha$ (1-2)Fru $\beta$ ), trehalose (T; Glc $\alpha$ (1-1) $\alpha$ Glc). Oligosaccharides: maltoheptaose (7 D-glucopyranose monomers in  $\alpha$  1-4 linkage), amylose (26 D-glucopyranose monomers in  $\alpha$  1-4 linkage), laminaraheptabiose (7 D-glucopyranose monomers in  $\beta$  1-3 linkage), and Curdlan (3 chains of 26 D-glucopyranose monomers in  $\beta$  1-3 linkage).

For each of these sugars both AA and CG simulations were performed. The monosaccharide and disaccharide systems, used for the parametrization, consisted of a single sugar molecule either in pure water or in water-saturated octanol. The water-saturated octanol consists of a 0.255 water/octanol molar fraction.<sup>40</sup> The oligosaccharides were simulated either in water or in nonane. An additional set of CG simulations was performed in which the concentration of glucose was increased systematically up to supersaturated solutions (60 wt %). A pure glucose system was also simulated. The cryo- and anhydro-protection effect of sugars was investigated by simulating a pre-equilibrated DPPC bilayer consisting of 64 lipids per leaflet either in pure water or in saccharide solutions. Two sugars were used for this investigation: glucose at 4 M (664 sugars and 2166 CG water particles) and trehalose at 2 M (332 sugars and 2166 CG water particles) concentration. Moreover, a control system was used in which the membrane was completely hydrated (2166 CG waters). All simulations were performed using the Gromacs package, version 3.3.1.<sup>41</sup> Table 1 provides the complete list of the systems simulated in this work, including details about compositions and total simulation times. The details of the simulation parameters are given below, including a description of the method by which the partition coefficients were computed.

**2.2.2. Coarse-Grained Simulation Parameters.** In the simulations at the coarse-grained level, we followed the standard simulation protocol used in the Martini parametrization.<sup>35</sup> The nonbonded interactions are cut off at a distance  $r_{\text{cut}}$  of 1.2 nm. To reduce generation of unwanted noise, the standard shift function of Gromacs<sup>41</sup> is used in which both the energy and the force smoothly vanish at the cutoff distance. The LJ potential is shifted from  $r = 0.9$  nm to the cutoff distance. The time step used to integrate the equations of motion is 20 fs. Note that here and throughout the entire manuscript actual simulation time is reported (i.e., no scaling of the time axis has been applied to provide an effective time scale). The temperature is maintained at 310 K by weak coupling of the solvent and solute separately to a Berendsen heat bath<sup>42</sup> with a relaxation time of 1 ps. The pressure is maintained at 1.0 bar by weak coupling to a pressure bath via isotropic coordinate scaling with a relaxation time of 5 ps. Simulations of bilayers are performed at three different temperatures (270, 325, and 475 K) with a semi-isotropic coupling of the lateral and perpendicular box dimensions to a pressure of 1.0 bar. The topology and parameters for water and octanol are taken from the Martini force field<sup>35</sup> data set.

In the case of nonane, the molecule is represented using two C1 particles similar to octane.

**2.2.3. All-Atom Simulation Parameters.** The AA simulations of hexopyranoses were performed using the latest Gromos force field parameters set for carbohydrates<sup>19</sup> (note that although the Gromos force field is a united-atom force field, it will be referred to as AA). In the case of furanose (i.e., in fructose and sucrose), parameters were adapted from the hexopyranose force field set (see Supporting Information). Verification of the furanose parameters was done by comparison to results from previous simulations<sup>43,44</sup> (Figures S1 and S2, Supporting Information) obtained with different force fields. For each simulation, the solute was placed in the center of a periodic cubic box with minimum wall–solute distances of 2 nm. The SPC water model<sup>45</sup> was used to solvate the system. For the simulations in nonane and octanol, the procedure to set up the systems was similar to that in the aqueous environment. The parameters for aliphatic hydrocarbons were taken from the Gromos 53a6 force field.<sup>46</sup> A steepest descent algorithm<sup>41</sup> was used to relax the internal interactions in vacuum. After that, the box was filled with the respective solvent and the minimization procedure was repeated. In all cases, a 2 fs time step was used to integrate Newton's equations of motion. The LINCS algorithm<sup>47</sup> was applied to constrain all bond lengths with a relative geometric tolerance of  $10^{-4}$ . Before production time, the systems were pre-equilibrated by slow heating up to 310 K. The temperature was maintained at 310 K by weak coupling of the solvent and solute separately to a Berendsen heat bath<sup>42</sup> with a relaxation time of 0.1 ps. Pressure coupling was maintained at 1.0 bar using an NPT ensemble by weak coupling via isotropic coordinate scaling with a relaxation time of 1 ps. Nonbonded interactions were handled using a twin-range cutoff<sup>48</sup> scheme. Within a short-range cutoff of 0.9 nm, the interactions were evaluated every time step based on a pair list recalculated every 5 time steps. The intermediate-range interactions up to a long-range cutoff radius of 1.4 nm were evaluated simultaneously with each pair list update and assumed constant in between. To account for electrostatic interactions beyond the long-range cutoff radius, a reaction field approach<sup>49</sup> was used with a relative dielectric permittivity of 66 for water, 2 for nonane, and 10.3 for octanol. Analysis of the dihedral distributions of the various disaccharides in an aqueous environment showed a good agreement with the ones obtained by Pereira et al.<sup>25</sup> using the same force field. Mapping of the AA trajectories to pseudo-CG trajectories was performed at a frequency of once per 40 ps. Table 1 provides a complete overview of the system composition and total simulation time for each of the simulations performed.

**2.2.4. Partitioning Free Energies.** In order to compute octanol/water partition coefficients  $P_{\text{OW}}$  directly, the free energies of solvation of the sugar compounds were calculated in both aqueous and organic phases. Given the appropriate free energies of solvation, computation of the partition coefficient is straightforward. The difference between the solvation free energy in the aqueous ( $\Delta G^{\text{W}}$ ) and organic phase ( $\Delta G^{\text{O}}$ ) is the partitioning free energy ( $\Delta \Delta G_{\text{OW}}$ ) of the carbohydrate between water-saturated octanol solution and water:

$$\Delta\Delta G_{\text{OW}} = -2.3RT \log P_{\text{OW}} \quad (6)$$

$\Delta G^{\text{W}}$  and  $\Delta G^{\text{O}}$  were calculated as the free energy difference  $\Delta F$  of the solute in vacuum (state A) and in the condensed phase (state B) using the thermodynamic integration (TI) procedure:<sup>50</sup>

$$\Delta F_{\text{BA}} = F_{\text{B}} - F_{\text{A}} = \int_{\lambda_A}^{\lambda_B} d\lambda \left\langle \frac{\partial U_{\text{uv}}(\lambda)}{\partial \lambda} \right\rangle_\lambda \quad (7)$$

Here  $U_{\text{uv}}(\lambda)$  denotes the potential energy function describing the total solute–solvent interaction, the average  $\langle \dots \rangle$  is taken over the MD trajectory, and  $\lambda$  is a coupling parameter that regulates the strength of  $U_{\text{uv}}$  and varies linearly from zero ( $\lambda_A$ )

**Table 1.** Summary of the Simulations Performed in This Work

	composition				temp. (K)	time (ns)		
	water	carbohydrate	nonane	octanol				
<b>(A) all-atom (AA)</b>								
<i>mapping procedure</i>								
monosaccharides	876	1			310	200		
disaccharides	876	1			310	200		
<i>log P oct/water</i>								
monosaccharides in vacuum		1			310	$25 \times 6^a$		
monosaccharides in water	1400	1			310	$25 \times 6$		
monosaccharides in octanol	66	1	199		310	$25 \times 6$		
disaccharides in vacuum		1			310	$25 \times 6$		
disaccharides in water	1400	1			310	$25 \times 6$		
disaccharides in octanol	66	1	199		310	$25 \times 6$		
<i>oligosaccharides</i>								
maltoheptaose in water	7106	1			310	40		
maltoheptaose in nonane		1	400		310	40		
amylose in water	21 223	1			310	40		
amylose in nonane		1	1825		310	40		
laminaraheptabiose in water	7106	1			310	40		
laminaraheptabiose in nonane		1	400		310	40		
curdlan in water	21 223	1			310	40		
curdlan in nonane		1	1077		310	40		
<b>(B) coarse-grained (CG)</b>								
<i>mapping procedure</i>								
monosaccharides	616	1			310	200		
disaccharides	616	1			310	200		
<i>log P oct/water</i>								
monosaccharides in vacuum		1			310	$25 \times 10$		
monosaccharides in water	1000	1			310	$25 \times 10$		
monosaccharides in octanol	43	1	519		310	$25 \times 10$		
disaccharides in vacuum		1			310	$25 \times 10$		
disaccharides in water	1000	1			310	$25 \times 10$		
disaccharides in octanol	43	1	519		310	$25 \times 10$		
<i>density profile</i>								
glucose in solution	0–100% w/w	0–100% w/w			310	$50 \times 50^b$		
<i>oligosaccharides</i>								
maltoheptaose in water	905	1			310	40		
maltoheptaose in nonane		1	500		310	40		
amylose in water	5137	1			310	40		
amylose in nonane		1	2000		310	40		
laminaraheptabiose in water	905	1			310	40		
laminaraheptabiose in nonane		1	500		310	40		
curdlan in water	5137	1			310	40		
curdlan in nonane		1	2000		310	40		
<i>cryo-protection effect</i>								
128 DPPC + 4 M glucose	2166	664			270/325/475	500		
128 DPPC + 2 M trehalose	2166	332			270/325/475	500		
128 DPPC	2166				270/325/475	500		

<sup>a</sup> Multiple simulations at different  $\lambda$  points were reported; see methods. <sup>b</sup> Fifty different sugar concentrations were used, simulated for 50 ns at each concentration level.

= 0) to full ( $\lambda_B = 1$ ) interaction. All bonded interactions were interpolated linearly; on the other hand, to remove the singularities in the potentials for the nonbonded interactions a soft-core approach was used.<sup>51</sup> Calculations were performed at 25 intermediate  $\lambda$  values until a smooth curve for the free energy derivative was obtained, which was then integrated numerically (by trapezoidal integration). For each individual  $\lambda$  point, at least 10 (CG) or 6 ns (AA) was used for the analysis. Additional  $\lambda$  points, especially in high-curvature regions, were required for the disaccharides at the AA level. Simulations in vacuum were performed using a stochastic dynamics approach with the same number of  $\lambda$  points as used for the water and octanol systems. To estimate the error in the free energy calculation, each  $\lambda$  set



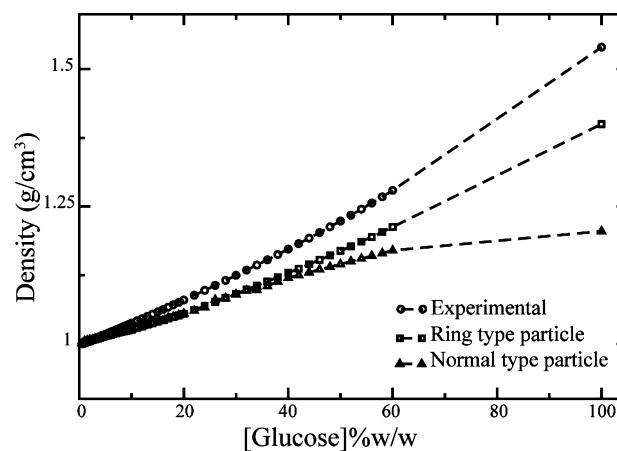
**Table 3.** Particle-Type Selection for CG Carbohydrates<sup>a</sup>

molecule	B1	B2	B3	B4	B5	B6
glucose (G)	P1	P4	P4			
fructose (F)	P1	P3	P4			
sucrose (SUC)	P1	P2	P4	P1	P1	P4
maltose (M)	P1	P2	P4	P2	P1	P4
cellobiose (C)	P1	P2	P4	P2	P1	P4
kojibiose (K)	P1	P2	P4	P2	P4	P1
sophorose (S)	P1	P2	P4	P2	P4	P1
nigerose (N)	P1	P2	P4	P2	P4	P1
laminarabiose (L)	P1	P2	P4	P2	P4	P1
trehalose (T)	P1	P2	P4	P2	P1	P4

<sup>a</sup> A complete overview of non-bonded interactions parameters can be found in the original Martini force field article.<sup>35</sup>

less polar P1 particle type represents propanol. Since the mapping of the AA structure to the CG model is not unique, we tested various options trying to optimize the performance of the model with respect to reproducing the experimental density and partitioning data (see below). We considered both normal particle types and the special class of ring "S" particles. The latter choice seems more logical given the ring structure of the sugars; however, these ring parameters were originally derived for use in a 2-to-1 or 3-to-1 mapping scheme, whereas the current mapping of the sugars is 4-to-1, consistent with the standard mapping. Although the use of ring particles improves the density profile for our model (see the density section below), this particle type decreases the sugar–sugar self-interaction to such an extent that even pure sugar remains liquid at room temperature. We decided not to change the strength of the self-interaction because it plays a crucial role in the packing and recognition between different sugar groups. We eventually settled on the topology as shown in Table 3. The inhomogeneous distribution of the polarity around the sugar rings is well reflected by the combination of the two more polar particle types (P3,P4) with one less polar particle type (P1). The latter maps to the part of the sugar containing carbons in positions 5 and 6. As we will show below, using this topology, the experimental water/octanol partition coefficient as well as the density of sugar solutions is reasonably well approximated.

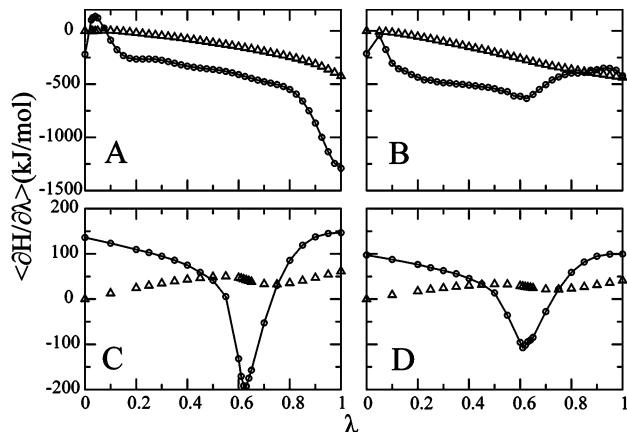
**3.1.2. Density.** To test if our CG model produces density of aqueous sugar solutions comparable with experimental data, a systematic set of simulations at different glucose concentrations was performed. Figure 3 shows the density of the solution as a function of glucose concentration. We compared our CG model to data obtained from the literature.<sup>52</sup> We find that the experimental densities are reproduced to within 10% in the condensed phase (e.g., to 60% w/w). At low concentrations (up to ~20% w/w), the agreement is even better. The underestimation of the density, especially at higher glucose concentrations, points to an effect due to sugar–sugar interactions, i.e., the packing of the CG glucose monomers is not as efficient as it is in reality. The situation can be improved somewhat by switching to the special class of ring particle types, which have a smaller effective size as set by the LJ parameter  $\sigma_{ii}$ . However, also the effective interaction strength, controlled by the value of  $\epsilon_{ii}$ , is reduced for this class of particles; as already mentioned before, this has the unwanted consequence that even pure sugar systems are found in a liquid state. With the normal particle types,



**Figure 3.** Density of aqueous glucose solutions as a function of concentration. Results from CG simulations at 310 K, for both ring-type particles (squares) and normal-type particles (triangles), are shown. For comparison, experimental data<sup>52</sup> obtained at  $T = 305$  K (circles) are included.

pure glucose forms a solid structure. Although we do not pretend to be able to reproduce the correct crystal packing with our CG model, at least the effective sugar–sugar interaction is strong enough to capture the transition from a solution toward the solid state upon increasing sugar concentration. Experimentally, the solubility of glucose at 300 K is around 47% w/w content. Using the normal particle types, visual inspection of the trajectories shows that our model becomes clearly more crystal-like around 50% w/w, with increasing aggregation of the solutes and phase separation of sugar rich clusters surrounded by layers of water.

**3.1.3. Partitioning.** In the Martini force field development, reproducing realistic partitioning behavior is of central importance. To select the optimal particle types for the carbohydrates, we therefore calculated the octanol/water partitioning free energies of our basic building blocks, glucose and fructose, in order to compare the values to either experimental results or to results obtained from AA simulations. We used the thermodynamic integration approach to calculate the free energies of solvation of the sugars in both water and water-saturated octanol. From the difference, the partitioning free energy is obtained as explained in the Methods section. Figure 4 depicts a comparison of the thermodynamic integration profile for glucose at the AA and CG levels of modeling. The derivative of the Hamiltonian  $H$  with respect to the integration parameter  $\lambda$ , as well as the running integrand, is plotted for successive  $\lambda$  points (since only the solute–solvent interaction is  $\lambda$  dependent, and the atomic masses are unchanged during the process,  $\partial H/\partial \lambda = \partial U_{uv}(\lambda)/\partial \lambda$  in eq 7). It is interesting to note that the  $\partial H/\partial \lambda$  profile looks very different for the AA versus the CG model, especially for the simulations in water. This is caused by the difference in particle size of the atomistic solvent molecules versus the CG water beads which unite four individual water molecules. The sharp drop in the profile at  $\lambda = 0.6$  in the case of the CG water molecules and for both CG and AA octanol reflects the transition between configurations with overlapping sugar/solvent molecules to the nonoverlapping, normal, situation at full interaction strength. For the relatively small AA water molecules, this transition



**Figure 4.** Thermodynamic integration profiles as a function of the integration parameter  $\lambda$  for glucose in water (A, C) and water-saturated octanol (B, D). Panels A and B are obtained with the AA model and panels C and D with the CG model. The solid line denotes the derivative of the free energy with respect to  $\lambda$ ; the triangles represent the integrated curve. The integration has been performed using a trapezoidal scheme. The integrated value at  $\lambda = 1$  corresponds to the free energy of hydration (A, C) or solvation in water-saturated octanol (B, D). The magnitude of the error at each  $\lambda$  point is smaller than the diameter of the circles, i.e., less than 10 kJ/mol for the AA and 1 kJ/mol for the CG system.

takes place at a much smaller value of  $\lambda$  (not clearly noticeable). The free energy, however, does not depend on the details of the integration path and corresponds to the integrated value at  $\lambda = 1$ .

The results of the free energy calculations are summarized in Table 4 and compared to available experimental values. The CG model reproduces the correct trend for free energies of solvation, although the actual values are systematically too low. This observation is in line with the results for different functional groups in the Martini force field.<sup>35</sup> As long as its application is aimed at studying the condensed phase and not at reproducing gas/fluid or solid/fluid coexistence regions, the most important thermodynamic property is the partitioning free energy. Importantly, the water/octanol partitioning of both monosaccharides can be accurately reproduced with the current parametrization of the CG model. A comparison to AA simulations and experimental data reveals a close agreement to within 2kT.

**3.2. Disaccharides.** **3.2.1. Topology.** Each of the disaccharides is modeled by two three-bead monosaccharide residues joined together by a single bond (see Figure 2). As discussed in the Method section, the monosaccharide units are represented differently from the individual monosaccharides. Analogous to the procedure followed for the monosaccharides, the parameters for the bonded interactions were obtained from a comparison to mapped AA trajectories. The full set of parameters is listed in Table 2. Bond lengths, defining the overall geometry of the molecules, vary somewhat depending on the sugar type, are found in the range 0.22–0.4 nm within each of the two sugar residues, and are slightly larger (up to 0.47 nm) for the B2–B4 bond representing the glycosidic linkage between them. Little variability was found for the stiffness of these bonds, so the force constant was set to 30 000 kJ mol<sup>-1</sup> nm<sup>-2</sup> for all of

them. Similar to the case of the monosaccharides, these bonds can be replaced by constraints in practice.

The conformation of disaccharides is mainly determined by the populations of rotamers around the glycosidic linkage. At the AA level, these rotameric conformations can be described by the glycosidic torsional angles  $\varphi$  ( $O_5-C_1-O_1-C_n'$ ) and  $\psi$  ( $C_1-O_1-C_n'-C_{n-1}'$ ) around a  $(1-n)$ -linkage<sup>25</sup> (with  $n = 2, 3, 4, 6$  and  $\varphi = \psi$  for trehalose). In the CG representation, the distinction between these dihedral angles is lost. Instead, the rotameric phase space available to the disaccharides needs to be represented by a set of dihedral angles (eq 4) and normal angles (eq 3) along the B2–B4 glycosidic bond. Upon transformation of the atomistic trajectories to effective CG trajectories, specific distributions for the angles and dihedrals were obtained, reflecting the conformational freedom of the disaccharides in water. As an example, Figures 5 and 6 show these distributions for every angle and dihedral of trehalose. The distributions obtained with the CG model are also shown, revealing that the atomistic configurations can be quite accurately mapped by our simplistic CG topology. The probability distributions associated with the  $\varphi$  and  $\psi$  glycosidic dihedral angles are essentially unimodal. Full rotation around these angles is observed at most once or twice for  $\psi$  (never for  $\varphi$ ) on the 50 ns time scale in the AA simulation studies performed by Pereira et al.,<sup>25</sup> corroborated with our AA simulations (data not shown). The exception is the distribution for 1–6 linked disaccharides (isomaltose and gentiobiose) for which a bimodal distribution is observed.<sup>25</sup> Since it is impossible to represent such a distribution with the dihedral approach used here (eq 4), the 1–6 linked sugars were omitted from the current study and left for future refinement.

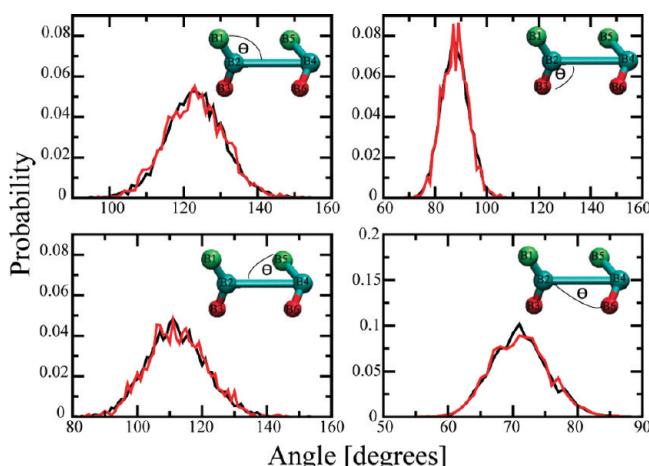
The assignment of particle types follows the assignment done for the monosaccharides, with the exception of the B2 and B4 beads involved in the glycosidic bond. The formation of a glycosidic bond between two glucose residues decreases the number of hydroxyl groups, which are largely responsible for the polarity of the molecule. To mimic this effect, a less polar particle type, P2, was used for the B2 and B4 beads (compared to P4 for the monosaccharides, see Table 3). For sucrose, the polarity was even further decreased for the correct reproduction of the partition coefficient, see below.

**3.2.2. Partitioning.** The water/octanol partitioning free energy, corresponding to the difference in solvation energy in water and octanol, is summarized in Table 4 for the different disaccharides considered in this study. Note that we only calculated the free energies for sucrose and maltose explicitly; the other sugars share the same particle assignment with maltose. Test runs showed that the geometrical fine details and overall conformational flexibility do not affect the free energies to within the error estimate of 1 kJ mol<sup>-1</sup>. For only two of the sugars (sucrose and trehalose) experimental data is available,<sup>53</sup> which is well matched by our CG model. Regarding the other disaccharides, comparison can be made to the atomistic model, for which we also computed the free energies of solvation and partitioning as listed in Table 4. There is a general good agreement between the results obtained at the atomistic and coarse-grained level. The atomistic model reveals little effect of the chemical

**Table 4.** Thermodynamic Parameters of Solvation and Partitioning Calculated for CG and AA Carbohydrates<sup>a</sup>

molecule	$\Delta G^W$ (AA) (kJ mol <sup>-1</sup> )	$\Delta G^O$ (AA) (kJ mol <sup>-1</sup> )	$\Delta\Delta G_{OW}$ (AA) (kJ mol <sup>-1</sup> )	$\log P_{OW(AA)}$	$\Delta G^W$ (CG) (kJ mol <sup>-1</sup> )	$\Delta G^O$ (CG) (kJ mol <sup>-1</sup> )	$\Delta\Delta G_{OW}$ (CG) (kJ mol <sup>-1</sup> )	$\log P_{OW(CG)}$	$\log P_{OW}$ (exp)
glucose (G)	-89	-74	15	-2.5	-60	-43	17	-2.9	-2.8
fructose (F)	-80	-69	11	-2.0	-60	-44	16	-2.7	
sucrose (SUC)	-107	-89	18	-3.0	-103	-83	20	-3.4	-3.3
maltose (M)	-121	-96	25	-4.2	-120	-96	24	-4.0	
cellobiose (C)	-114	-90	24	-4.0	-120	-96	24	-4.0	
kojibiose (K)	-121	-93	28	-4.7	-120	-96	24	-4.0	
sophorose (S)	-120	-88	32	-5.4	-120	-96	24	-4.0	
nigerose (N)	-119	-89	30	-5.0	-120	-96	24	-4.0	
laminarabiose (L)	-120	-91	29	-5.0	-120	-96	24	-4.0	
trehalose (T)	-120	-92	28	-5.0	-120	-96	24	-4.0	-3.78

<sup>a</sup> The partition coefficient of octanol–water  $\log P_{OW}$  is based on the difference between the independently calculated free energy of hydration ( $\Delta G_W$ ) and free energy of solvation in water-saturated octanol ( $\Delta G_O$ ), according to eq 6. Simulation data were obtained at 310 K, whereas the temperature of the experimental data ( $\log P_{OW(exp)}$ ) is 300 K.<sup>53</sup> The statistical accuracy of the free energies obtained from CG simulations is 1 kJ mol<sup>-1</sup>. The CG values for cellobiose, kojibiose, sophorose, nigerose, laminarabiose, and trehalose were set equal to the values for maltose (as the topologies are based on the same kind of coarse particle types).

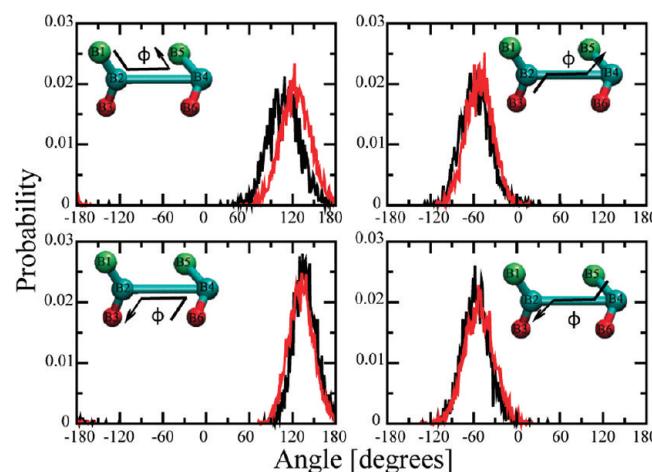


**Figure 5.** Angle distributions ( $\theta$ ) for trehalose obtained from AA simulations after mapping to CG particles (black line) and from CG simulations (red line). Four angles, indicated by the insets, were used to preserve the characteristic rotameric states of the disaccharide at the coarse-grained level.

details of the sugars on their partitioning, justifying our choice of the same particle types for all disaccharides except sucrose.

**3.3. Oligosaccharides.** Molecular systems in which carbohydrates are involved are not limited to monosaccharides or disaccharides but also include long polymeric sugar chains exhibiting a large variety in monosaccharide composition, linkage type, and degree of branching. For this reason, the following part of the article illustrates how the parameters derived for the simulation of mono- and disaccharides can be applied to study the structure and dynamics of oligosaccharides. As an example, two different oligosaccharides have been studied at both the CG and the AA levels. First, two amylose-type chains of different lengths are considered as an example of  $\alpha$  1–4 linked polymers. Second, a triple-helix structure representing Curdlan is considered as an example of  $\beta$  1–3 linked structure.

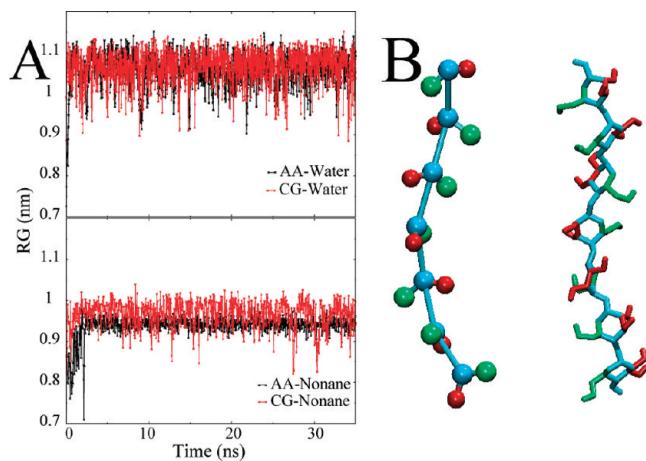
**3.3.1. Amylose.** Before considering the full amylose chain, we studied the short oligosaccharide maltoheptaose. Maltoheptaose consists of seven glucose monomers, connected by  $\alpha$  1–4 glycosidic linkages just like amylose. The behavior of maltoheptaose was studied both in aqueous and in



**Figure 6.** Dihedral angle distributions ( $\phi$ ) for trehalose obtained from AA simulations after mapping to CG particles (black line) and from CG simulations (red line). In addition to the angle potentials (cf. Figure 5), four dihedral angles, indicated by the insets, were used to preserve the characteristic rotameric states of the disaccharide at the coarse-grained level.

nonpolar environments (nonane). Results from the CG simulations are compared to results obtained with an AA using a similar setup consisting of a single molecule in excess solvent.

In our first attempt, we took the parameters derived for the disaccharides and simply extended the topology to model the heptamer. However, with this setup, the configurations sampled at the CG and AA levels did not overlap (data not shown). Therefore, an extra angle potential (eq 3) was used to reproduce the correct structural shape of this molecule in water as well as in nonane. Three consecutive backbone particles were subject to an angle potential, as illustrated in Figure 2C. We found that the optimal angle parameters are an equilibrium angle ( $\theta_0$ ) of 154° in water and 120° in nonane with a force constant ( $K_{angle}$ ) of 100 and 250 kJ mol<sup>-1</sup> respectively. Including these additional ‘three-sugar’ potentials, the CG representation nicely matches the structure observed in the AA simulation, as is illustrated by the snapshots shown in Figure 7B for the case of maltoheptaose in water. Figure 7A also provides a comparison between the



**Figure 7.** Structure and dynamics of the maltoheptaose oligomer. (A) Radius of gyration (RG) as a function of the simulation time for the AA (black lines) and CG (red lines) systems. (B) Snapshot obtained from the CG (left) and AA (right) simulation in water.

temporal evolution of the radius of gyration (RG) of maltoheptaose in both water and nonane. The average value for both the AA and the CG simulation in water is 1.05 nm, which is in agreement with the previous value obtained by Shimada et al.,<sup>13</sup> revealing a more extended structure compared to the structure obtained by small-angle X-ray scattering. In nonane, the structure is somewhat more compact with an average RG of 0.95 (AA) and 0.98 nm (CG). Judging from the fluctuations of the RG over time, the AA structure is slightly more rigid compared to the CG one.

Our results show that, with the addition of an angle term acting between three consecutive sugar moieties, the short oligosaccharide maltoheptaose can be quite accurately modeled. The question is, will this suffice to also model the longer oligosaccharide amylose, the principal component of starch? Like maltoheptaose, amylose is a linear oligosaccharide of 1–4  $\alpha$  linked glucose monomers. In aqueous solution it behaves as a flexible random coil with stretches of left-handed helical segments, which are more pronounced at low hydration levels.<sup>54</sup> In fact, AFM experiments have shown that after the assisted unfolding of the molecule, it tries to refold again to a helical conformation. However, this refolding was not complete<sup>55</sup> unless a less polar solvent (butanol) was used. In general, stable secondary conformations, known as A-, B-, and V-amylose,<sup>54</sup> are formed in either ionic solutions or less polar solvents. The A and B allomers consist of parallel left-handed double helices with six glucopyranosil units per turn, differing only in the number of helices packed in the unit cell. V-Amylose, cocrystallized with compounds such as iodine, DMSO,<sup>56</sup> alcohols, or fatty acids, reveals a strict structure of left-handed helix<sup>57</sup> with six to eight glucose residues per turn. Multiple helices form a central channel in which the additives are complexed. In fact, a large number of V-amylose crystalline structures have been obtained, depending on the exact crystallization conditions.<sup>54</sup>

On the basis of the parameter set for maltoheptaose, several CG simulations were performed for a 26-glucose amylose

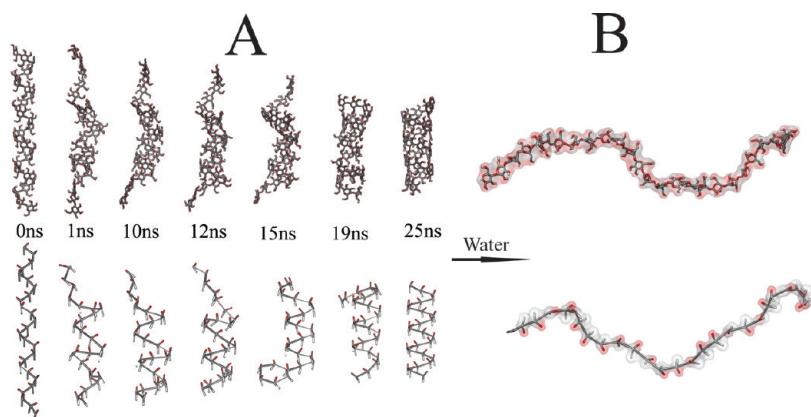
chain. For comparison, AA simulations were performed as well. Both water and nonane were used as solvents. Figure 8A shows a comparison of the structure formation of the amylose chain in nonane at the AA and CG level. The initial conformation was a regular helical structure characterized by the torsional angles  $\varphi$  ( $O_5-C_1-O_1-C_4'$ ) = 87.27°,  $\psi$  ( $C_1-O_1-C_4'-C_3'$ ) = 101°, and  $\omega'$  ( $O_5-C_5-C_6-O_6$ ) = -125.97°. After 25 ns simulation, both simulations (AA and CG) refold to the same helical form characteristic of V-amylose. This structure proved stable for the remainder of the simulation (40 ns in total). Figure 9 shows a close-up view of both the AA and the CG equilibrium structure, with the values for the pitch of the helix and diameter of the channel indicated. The atomistic and coarse-grained structures are nearly indistinguishable. Experimentally, the V-amylose conformation is observed by X-ray studies<sup>11,57</sup> in complex with nonpolar solvents, in agreement with our simulations. Moreover, the experimentally determined pitch value, counting 6 glucose residues per turn, gives an average value of 7.9 Å. In our simulations, the average value is 7.5 Å (for both AA and CG), in good agreement with the experiment, given the differences between the experimental and simulation conditions.

In water, the amylose chain remains largely unfolded during the simulation, also in agreement with the experimental observation.<sup>58</sup> Figure 8B shows typical snapshots of AA and CG amylose, revealing a somewhat extended, fully solvated structure. To quantify the degree of extension, the radius of gyration was calculated. Averaged over 40 ns, the radius of gyration of the amylose chain is found to be similar at both levels of resolution, namely, 3.2 and 3.0 nm ( $\pm 0.1$  nm) for the CG and AA system, respectively.

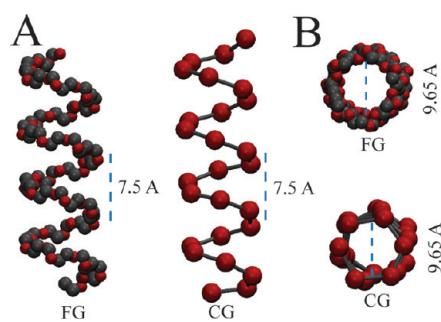
**3.3.2. Curdlan.** In recent years, there has been a great interest in the  $\beta$ -1,3-D-glucan series (Curdlan) because they show antitumor and anti-HIV viral activity in humans.<sup>56</sup> After a rigorous purification step, the structure of Curdlan has been identified as a right-handed triple-helical complex under aqueous conditions.<sup>59</sup> A number of additional crystal structures<sup>60,61</sup> also reveal a triple helix formed by three parallel independent chains and stabilized by both inter- and intramolecular hydrogen bonds.<sup>28</sup>

The CG model of Curdlan was based on the laminarabiose disaccharide. Similar to the case of extending maltose into oligomers, an extra angle potential was used for three consecutive backbone beads (cf. Figure 2C). The value of this angle was obtained from matching CG to AA conformations of a chain consisting of seven  $\beta$  1–3 glucose monomers in water, sampled over a 40 ns trajectory. The optimal parameters were found to be  $\theta_0$  = 136° with  $K_{\text{angle}}$  of 500 kJ mol<sup>-1</sup>. A simulation of the same molecule was performed in nonane in order to determine if this angle depends on the solvent. Contrary to the case of the  $\alpha$  1–4 linked sugars described above, here we found no significant dependency of the additional angle potential on solvent environment.

Next, three chains composed of 26  $\beta$  1–3 glucose monomers, representing Curdlan, were simulated in nonane. The additional angle potential was included. The starting structure was taken from Deslandes' CUR data,<sup>59</sup> determined by X-ray crystallography. Figure 10A shows the final



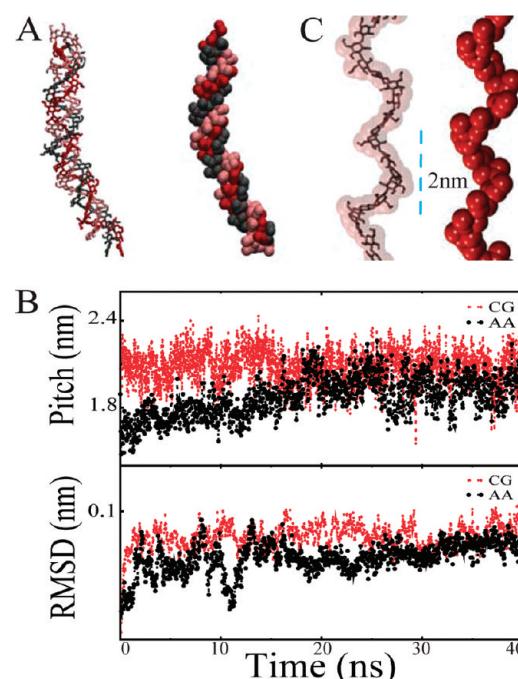
**Figure 8.** Structure of amylose in nonane and water. (A) Snapshots of a 26-mer of amylose, simulated in nonane, at AA (top drawings) and CG (bottom drawings) levels. In both cases, starting from an elongated helix, a transition to a stable V shape is observed during 25 ns of simulation. (B) After transferring this molecule to an aqueous solution, amylose evidences an increased flexibility with unfolding of the helical structure.



**Figure 9.** Structural characteristics of amylose in nonane. Only backbone particles (gray/red spheres) are shown. The pitch value (A) and helix diameter (B) of both AA and CG representations are identical, as indicated.

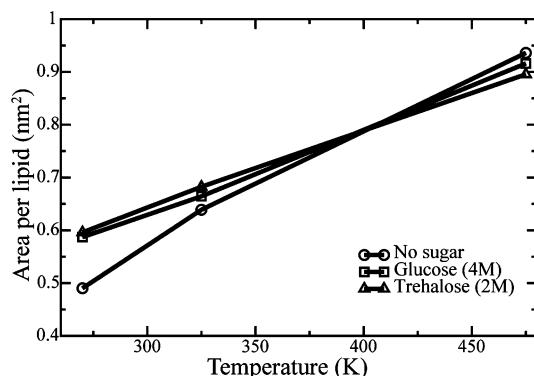
structures obtained after 40 ns both at the AA and CG levels of resolution. The characteristic triple-helical structure of Curdlan is well preserved. No significant difference was detected between the average CG and the AA structure. This is illustrated in Figure 10B, showing root mean squared deviation (rmsd) between the AA (mapped to CG coordinates) and the CG representation. Moreover, we calculated the value for the pitch of each turn, which is defined as the length between the six main-chain glucose units along the helix *c* axis.<sup>59</sup> The result is also shown in Figure 10B. The average pitch value for both AA and CG structure, 2.0 nm, is in good agreement with the experimental value of 1.8 nm,<sup>59</sup> despite the differences in environment (apolar solvent versus crystal). The pitch value obtained in the simulations also agrees well with previous simulation studies.<sup>28</sup> In addition, we tested the stability of the triple-helix configuration of Curdlan in water at both levels of resolution. We basically found the same results as observed with nonane as a solvent, indicating that our CG approach can effectively mimic the strength of the intra- and intermolecular hydrogen bonds, preserving Curdlan's triple-helical structure regardless of the solvent.

**3.4. Cryo- and Anhydro-Protection Effects.** The ability of sugars to act as cryo- and anhydro-protective agents has been well established.<sup>3,62</sup> Several organisms make use of this property of sugars; by increasing their intracellular sugar concentration they have been found to survive under low-



**Figure 10.** Comparison of the AA and CG structure of Curdlan in nonane. (A) Snapshots of the triple-helix structure, stable for both the AA (left) and the CG (right) systems (each helix represented by a different color type). (B) Temporal evolution of the helical pitch (top graph) and rmsd with respect to the starting structure (bottom graph) for both levels of resolution. (C) Indication of the pitch distance for a single helical strand at AA (left) and CG (right) resolution. For best comparison, only one helix at the AA level (stick representation with transparent spheres) and one at the CG representation (red balls) were considered.

temperature or low-hydration conditions over extended periods of time. The origin of the cryo- or anhydro-protective effect of membranes is usually explained by different mechanisms, namely, (i) replacement of the lipid–water hydrogen bonds by lipid–sugar hydrogen bonds, (ii) entrapment of lipid hydration water, and (iii) vitrification effects. There is an ongoing debate about which of these mechanisms dominates in the modulation of many bilayer properties.<sup>63</sup> One of these properties is the main phase transition temper-

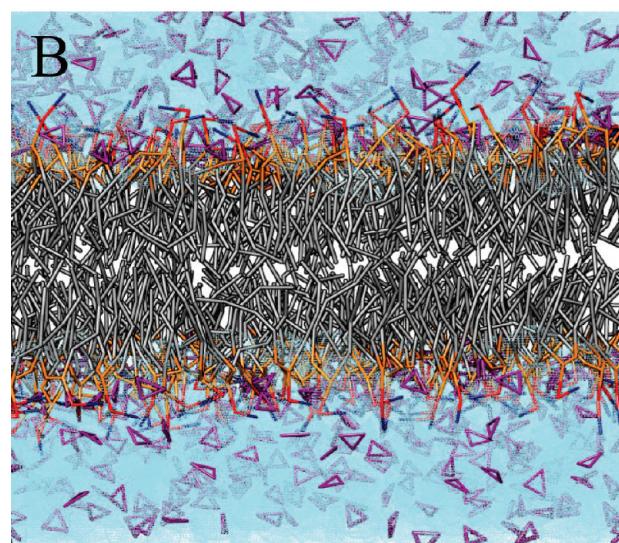
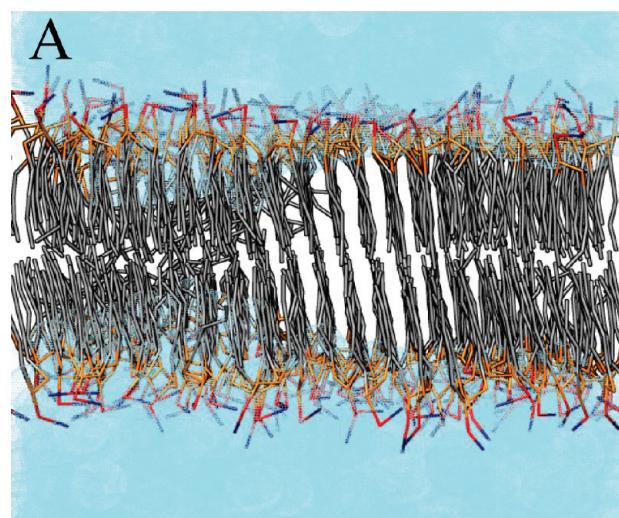


**Figure 11.** Average area per lipid of a DPPC membrane as a function of temperature. Circles: system with no sugars. Squares: 4 M glucose solution. Triangles: 2 M trehalose solution.

ature of the lipid bilayer, which is lowered in the presence of sugars. Thus, the biologically relevant liquid-crystalline phase is stabilized, protecting the cell membrane from freezing.

In order to test the capabilities of our model to reproduce this effect, we performed several 500 ns CG simulations of a DPPC bilayer in the presence of glucose and trehalose at 4 and 2 M, respectively, varying the temperature in the range 270–475 K. A pure DPPC bilayer without sugars was also simulated for comparison. The area per lipid was used to monitor the protective effect of the sugars. In Figure 11 the dependency of the average area per lipid on the temperature is shown for each of the three systems. A clear difference between the pure DPPC membrane, on the one hand, and the DPPC membrane in the presence of sugars, on the other hand, can be appreciated. Whereas the pure system adopts an area per lipid of 0.48 nm<sup>2</sup> at 270 K, characteristic of a gel phase, both trehalose and glucose manage to keep the area per lipid at a value similar to the value at 325 K, in a liquid-crystalline state. At elevated temperatures, the sugars appear to have less effect, although the thermal expansivity (i.e., the slope of the curves in Figure 11) is somewhat smaller in the presence of the sugars. A close view of the structural effect of the sugars on the bilayers at the low-temperature range is depicted in Figure 12. In the pure system, the DPPC tails at 270 K have adopted a straight conformation, and the system has transformed into a gel phase. Note that the Martini model does not reproduce the experimentally observed tilt of the lipid tails in the gel state.<sup>64</sup> However, the bilayer remains in a liquid-crystalline state when a large amount of glucose is added; this is clearly noticeable by the disordered acyl chains. The same phenomenon was also observed when trehalose was added (using one-half the concentration of glucose).

Although a real vitrification process cannot be directly observed at the CG level, our simulations point to a direct interaction between the sugars and the DPPC lipid head groups. Figure 13 shows the density profiles of the lipids, water, and sugars across the bilayer normal. The profile for glucose shows that the sugars are able to bind to the lipid/water interface and can penetrate the membrane up to the level of the carbonyl groups. Consequently, the amount of water in the interface is reduced. By intercalating between

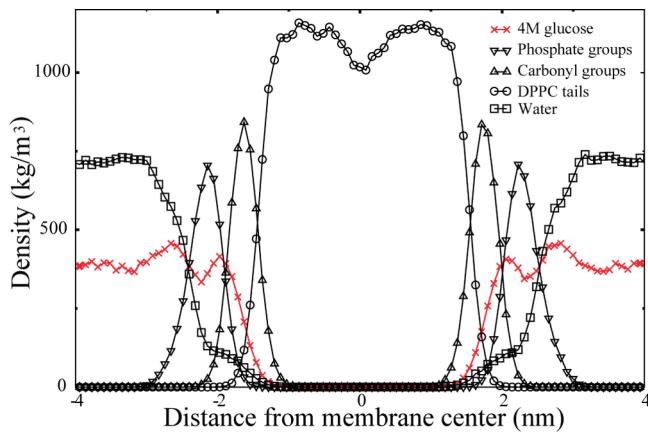


**Figure 12.** Snapshots of CG DPPC bilayers at  $T = 270$  K. (A) Gel phase formed in pure water. (B) Fluid phase stabilized in the presence of 4 M glucose. Lipid tails are depicted in gray, head groups are shown blue/orange for choline/phosphate moieties, and sugars are shown in purple. Water is represented by semitransparent blue dots.

the head groups, the sugar molecules are replacing part of the hydrating water and the freezing of the acyl chains is avoided. The same characteristic was also observed for trehalose. These results are most easily interpreted in the context of the water-replacement hypothesis, as has been proposed by Pereira et al.<sup>63,65</sup> and Sum et al.,<sup>66</sup> based on atomistic simulations.

#### 4. Limitations and Outlook

The potential range of applications of the carbohydrate Martini model is very broad. Along the lines of the parametrization as presented in this article, the model could be extended toward other (oligo)saccharides. There are, however, certain important limitations which should be kept in mind. An obvious limitation is that the limited resolution of our CG model prevents the distinction between  $\alpha$  and  $\beta$  anomers of reducing ends as well as the stereochemistry of



**Figure 13.** Mass density profiles across a CG DPPC bilayer in the presence of 4 M glucose at  $T = 270$  K. Crosses: glucose. Triangles down: phosphate group. Triangles up: carbonyl groups. Circles: lipid tails. Squares: water.

exocyclic OH groups. In aqueous solution, the sugar hydroxyl groups are directed outward to oxygen atoms of adjacent water molecules, thus destroying the intramolecular hydrogen bonding.<sup>67</sup> Furthermore, an important simplification of the model is that ring puckering (i.e., chair–chair or chair–boat transformations) in hexopyranoses is completely neglected. Only the  ${}^4\text{C}_1$  chair conformation is considered. This is not a real problem for the simulation of long polymers, since the  ${}^4\text{C}_1$  chair conformation is the dominant puckering state<sup>68,69</sup> and the conformations of oligosaccharides are mainly determined by the accessible rotameric states around the glycosidic linkage. There are, however, exceptions, notably idose and some sulfated sugars.<sup>22</sup> In the case of furanoses, a single conformation is less of a problem; five-membered rings are floppy, but the overall shape does not change substantially as the ring undergoes pseudo-rotation. Special attention is also required in systems with high sugar density, e.g., at low hydration conditions. We have shown an important deviation of the carbohydrate packing density under these conditions. This effect is most severe in the limit of a pure crystalline sugar phase but might also show up, for instance, in the condensing efficiency of long oligosaccharide chains in poor solvents.

The disaccharides used in this work exhibit primarily a single state for the glycosidic bond, which we showed to be easily represented using a dihedral potential at the CG level. Whether this is true for oligosaccharides in general is questionable. The glycosidic linkage is indeed in a single state (denoted syn) for all  $\alpha$ -linked disaccharides but not necessarily for  $\beta$ -linkages. There is evidence from NMR that the other conformations are populated in solution, and crystal structures of protein/sugar complexes indicate antistates.<sup>70–72</sup> The underlying AA force field may also exhibit these states, but revealing them might require more extensive (nonequilibrium) sampling. Multiple states are only sampled in the case of the 1–6 bond. As previously found for the rotation of the hydroxymethyl group in hexopyranoses,<sup>18</sup> there is a clear preference for the gg ( $\omega = 180^\circ$ ) and gt ( $\omega = -60^\circ$ ) rotamers which are characterized by nearly identical free energies. At the CG level, this behavior is not easily represented by a single well potential, and therefore, the 1–6

linkage has not been considered here. An alternative to overcome the above limitations is the use of tabulated two- or three-well angle or dihedral potentials, which is currently under investigation. Anyhow, we have to keep in mind that the underlying atomistic force field may also have its shortcomings, despite the fact that it has been tested thoroughly.<sup>16,19,22,25</sup> We also note that the current parametrization is restricted to D-D sugars; L-L sugars are easily represented by changing the sign of the dihedral reference value, but for the L-D sugars, reparametrization of the bonded parameters is required.

Another important issue is transferability of the CG parameters. We found that the disaccharide glycosidic bond parameters were easily extrapolated to oligomers, except for the use of an additional angle potential required to adopt the appropriate structure (as judged from AA simulations). This angle potential is likely to be nontransferable, i.e., it needs to be parametrized for different type of linkages. Besides, the optimal parameters were found to depend on the type of solvent, at least in the case of 1–4  $\alpha$  linked sugars. This is not a desirable situation. Especially for applications of oligosaccharides near interfaces, the current parametrization is expected to be problematic.

The parametrization presented in this article should be viewed only as a first step toward a comprehensive carbohydrate force field. Improvements are expected to take place hand in hand with the ongoing development of the Martini force field. The model is easily extendable to include polymers containing more than one type of sugar–sugar linkage or featuring branched sugars. Potential applications include a variety of oligosaccharides such as cellulose and dextran and mixed systems such as membranes containing glycolipids and the bacterial cell wall.

As a spin off of the parametrization of the CG force field, we found that fast folding of oligosaccharides can be observed in nonpolar solvents, even at an atomistic level of resolution. As an example, the folding of a 26 sugar residue amylose chain was presented (cf. Figure 9). Within 25 ns, the molecule changed its fold from an extended conformation toward the experimentally observed crystal structure, the V-shape. In simulations, crystallization conditions of sugars are seemingly effectively reproduced by the use of low-dielectric solvents. In contrast to proteins, which possess side chains of different polarity, sugars can be forced to fold toward the native crystal structure by replacing the hydration shell by nonpolar solvents.

In summary, in this article, an extension of the Martini force field parameters to carbohydrates has been presented. On the basis of atomistic simulations, a complete set of bonded parameters was extracted to model the dynamics and structure of several mono- and disaccharides at the CG level. Standard particle types of the Martini force field were used for the nonbonded interactions, assuring that the carbohydrate model is fully compatible with the other biomolecular components. Since most applications of the CG model are naturally in the condensed phase, the reproduction of the correct partitioning free energies between polar and nonpolar phases is essential. We demonstrated that our model predicts values for water/octanol partitioning in close agreement with

atomistic data and where available with experimental measurements. With an additional angle potential to account for the collective effect of three linked sugar residues, the model appears well suited for application to oligosaccharides. Both an amylose chain and a triple-helical Curdlan structure were modeled; despite the lack of explicit hydrogen bonding at the CG level, the conformation and dynamics were found to be in good agreement with simulations at the all-atom level. In addition, our CG model is able to reproduce semiquantitatively the modulating effect of sugars on lipid bilayers, in particular their cryo- and anhydro-protective effect.

Keeping in mind certain inherent limitations of the CG carbohydrate model, such as the inability to represent ring puckering or some of the complex rotameric states exhibited by certain sugar links, the model shows great promise for exploring the phase space of carbohydrate systems which are computationally too costly at full atomistic resolution. Moreover, the sugar parameters are fully compatible with the other parameters in the Martini force field, opening the way to explore a large variety of sugar-containing biomolecular systems at an unprecedented scale.

**Acknowledgment.** The authors would like to acknowledge Cristina Pereira for her kind help with the AA simulations.

**Supporting Information Available:** The complete list of parameters for the simulation of furanose rings (fructose and sucrose) at full atom resolution is provided as supporting information. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- Bush, C. A.; Martin-Pastor, M.; Imbert, A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 269–293.
- Ladimir, V.; Melia, E.; Haddleton, D. M. *Eur. Polym. J.* **2004**, *40*, 431–449.
- Dashnau, J. L.; Sharp, K. A.; Vanderkooi, J. M. *J. Phys. Chem. B* **2005**, *109*, 24152–24159.
- Crowe, J. H.; Oliver, A. E.; Tablin, F. *Integr. Comp. Biol.* **2002**, *42*, 497–503.
- Crowe, L. M. *Comp. Biochem. Physiol., A: Mol. Integr. Physiol.* **2002**, *131*, 505–513.
- Wormald, M. R.; Petrescu, A. J.; Pao, Y.-L.; Glithero, A.; Elliott, T.; Dwek, R. A. *Chem. Rev.* **2002**, *102*, 371–386.
- Duus, J.; Gotfredsen, C. H.; Bock, K. *Chem. Rev.* **2000**, *100*, 4589–4614.
- Yu, F.; Prestegard, J. *Biophys. J.* **2006**, *91*, 1952–1959.
- Momany, F. A.; Willett, J. L. *Biopolymers* **2002**, *63*, 99–110.
- Corzana, F.; Motawia, M. S.; Du Penhoat, C. H.; Perez, S.; Tschampel, S. M.; Woods, R. J.; Engelsen, S. B. *J. Comput. Chem.* **2004**, *25*, 573–86.
- Buléon, A.; Delage, M. M.; Brisson, J.; Chanzy, H. *Int. J. Biol. Macromol.* **1990**, *12*, 25–33.
- Li, H.; Rief, M.; Oesterhelt, F.; Gaub, H. E.; Zhang, X.; Shen, J. *Chem. Phys. Lett.* **1999**, *305*, 197–201.
- Shimada, J.; Kaneko, H.; Takada, T.; Kitamura, S.; Kajiwara, K. *J. Phys. Chem. B* **2000**, *104*, 2136–2147.
- Behrends, R.; Kaatze, U. *ChemPhysChem* **2005**, *6*, 1133–1145.
- Green, D. F. *J. Phys. Chem. B* **2008**, *112*, 5238–5249.
- Kräutler, V.; Müller, M.; Hünenberger, P. H. *Carbohydr. Res.* **2007**, *342*, 2097–2124.
- Guvenc, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *J. Comput. Chem.* **2008**, *29*, 2543–2564.
- Kony, D.; Damm, W.; Stoll, S.; van Gunsteren, W. F. *J. Comput. Chem.* **2002**, *23*, 1416–1429.
- Lins, R. D.; Hünenberger, P. H. *J. Comput. Chem.* **2005**, *26*, 1400–12.
- Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.
- Spivok, V.; Lipovová, P.; Skálová, T.; Vondráčková, E.; Dohnálek, J.; Hašek, J.; Králová, B. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 887–901.
- Hansen, H. S.; Hünenberger, P. H. *J. Comput. Chem.* **2009**. DOI: 10.1002/jcc.21253.
- Stenutz, R.; Widmalm, G. *Glycoconjugate J.* **1998**, *15*, 415–418.
- Kozar, T.; Tvaroska, I.; Carver, J. P. *Glycoconjugate J.* **1998**, *15*, 187–191.
- Pereira, C. S.; Kony, D.; Baron, R.; Müller, M.; van Gunsteren, W. F.; Hünenberger, P. H. *Biophys. J.* **2006**, *90*, 4337–4344.
- Sugiyama, H.; Nitta, T.; Horii, M.; Motohashi, K.; Sakai, J.; Usui, T.; Hisamichi, K.; Ishiyama, J. I. *Carbohydr. Res.* **2000**, *325*, 177–182.
- Yu, H.; Amann, M.; Hansson, T.; Köhler, J.; Wich, G.; van Gunsteren, W. F. *Carbohydr. Res.* **2004**, *339*, 1697–1709.
- Okobira, T.; Miyoshi, K.; Uezu, K.; Sakurai, K.; Shinkai, S. *Biomacromolecules* **2008**, *9*, 783–788.
- Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press: Boca-Raton, 2008.
- Liu, P.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 11566–11575.
- Molinero, V.; Goddard, W. A. *J. Phys. Chem. B* **2004**, *108*, 1414–1427.
- Bathe, M.; Rutledge, G. C.; Grodzinsky, A. J.; Tidor, B. *Biophys. J.* **2005**, *88*, 3870–3887.
- Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.
- Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- Risselada, H. J.; Marrink, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17367–17372.
- Baoukina, S.; Monticelli, L.; Risselada, H. J.; Marrink, S. J.; Tieleman, D. P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10803–10808.
- Yefimov, S.; van der Giessen, E.; Onck, P. R.; Marrink, S. J. *Biophys. J.* **2008**, *94*, 2994–3002.
- Rzepiela, A. J.; Schafer, L. V.; Goga, N.; Risselada, H. J.; de Vries, A. H.; Marrink, S. J. *J. Comput. Chem.* **2009**, in press.

- (40) Best, S. A.; Merz, K. M.; Reynolds, C. H. *J. Phys. Chem. B* **1999**, *103*, 714–726.
- (41) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (42) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (43) Costa, M. T. M. *Carbohydr. Res.* **2005**, *340*, 2185–2194.
- (44) Engelsen, S.; Herve du Penhoat, C.; Perez, S. *J. Phys. Chem.* **1995**, *99*, 13334–13351.
- (45) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*, in *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981.
- (46) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
- (47) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (48) van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed.* **1990**, *29*, 992–1023.
- (49) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5459.
- (50) van Gunsteren, W. F.; Berendsen, H. J. C. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 171–176.
- (51) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- (52) Ji, P.; Feng, W.; Tan, T. *J. Chem. Eng. Data* **2007**, *52*, 135–140.
- (53) Mazzobre, M. F.; Román, M. V.; Mourelle, A. F.; Corti, H. R. *Carbohydr. Res.* **2005**, *340*, 1207–1211.
- (54) Gessler, K.; Uson, I.; Takaha, T.; Krauss, N.; Smith, S. M.; Okada, S.; Sheldrick, G. M.; Saenger, W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4246–4251.
- (55) Lu, Z.; Nowak, W.; Lee, G.; Marszalek, P. E.; Yang, W. *J. Am. Chem. Soc.* **2004**, *126*, 9033–9041.
- (56) di Luzio, N. R.; Williams, D. L.; McNamee, R. B.; Edwards, B. F.; Kitahama, A. *Int. J. Cancer* **1979**, *24*, 773–779.
- (57) Helbert, W.; Chanzy, H. *Int. J. Biol. Macromol.* **1994**, *16*, 207–213.
- (58) Maciejewska, W.; KaczmarSKI, M. *J. Raman Spectrosc.* **1989**, *20*, 413–418.
- (59) Deslandes, Y.; Marchessault, R. H.; Sarko, A. *Macromolecules* **1980**, *13*, 1466–1471.
- (60) Okuyama, K.; Otsubo, A.; Fukuzawa, Y.; Ozawa, M.; Harada, T.; Kasai, N. *J. Carbohydr. Chem.* **1991**, *10*, 645–656.
- (61) Chuah, C. T.; Sarko, A.; Deslandes, Y.; Marchessault, R. H. *Macromolecules* **1983**, *16*, 1375–1382.
- (62) Montiel, P. *Cryo Lett.* **2000**, *21*, 83–90.
- (63) Pereira, C. S.; Lins, R. D.; Chandrasekhar, I.; Freitas, L. C. G.; Hünenberger, P. H. *Biophys. J.* **2004**, *86*, 2273–2285.
- (64) Marrink, S. J.; Risselada, J.; Mark, A. E. *Chem. Phys. Lipids* **2005**, *135*, 223–244.
- (65) Pereira, C. S.; Hünenberger, P. H. *Mol. Simul.* **2008**, *34*, 403–420.
- (66) Sum, A. K.; Faller, R.; de Pablo, J. J. *Biophys. J.* **2003**, *85*, 2830–2844.
- (67) Kabayama, M. A.; Paterson, D. *Can. J. Chem.* **1958**, *36*, 563–573.
- (68) Reeves, R. E. *J. Am. Chem. Soc.* **1950**, *72*, 1499–1506.
- (69) Franks, F.; Lillford, P. J.; Robinson, G. *J. Chem. Soc., Faraday Trans.* **1989**, *85*, 2417–2426.
- (70) Höög, C.; Landersjö, C.; Widmalm, G. *Chemistry* **2001**, *7*, 3069–3077.
- (71) Vidal, P.; Vauzeilles, B.; Bieriot, Y.; Sollogoub, M.; Sinay, P.; Jimenez-Barbero, J.; Espinosa, J. F. *Carbohydr. Res.* **2007**, *342*, 1910–1917.
- (72) DeMarco, M. L.; Woods, R. J. *Glycobiology* **2009**, *19*, 344–355.

CT900313W