M3B: A Coarse Grain Force Field for Molecular Simulations of Malto-Oligosaccharides and Their Water Mixtures

Valeria Molinero[†] and William A. Goddard III*

Materials and Process Simulation Center (MC 139-74), California Institute of Technology, Pasadena, California 91125

Received: May 27, 2003; In Final Form: November 6, 2003

We derive a chemically realistic coarse grain model and force field for the simulation of malto-oligosaccharides (α(1→4) D-glucans) and their aqueous mixtures. This coarse grain model for carbohydrates (denoted M3B) represents each glucose monomer by three beads while describing the water molecule as a single particle. M3B includes no charges or hydrogen-bonding terms, using only two-body Morse functions to describe long-range forces. The configurations obtained with the M3B model map uniquely and quickly back to a full atomistic description. M3B was parametrized to fit the results from atomistic simulations for the gas phase and amorphous bulk phase of sugars over a wide range of pressures. In particular, we required that the M3B force field provide an accurate representation of such quantities as excluded volume interactions and the distribution of torsional configurations about the intermonomer bonds. We find that M3B leads to a helical structure for polysaccharide chains and predicts left-handed helices to be more stable than right-handed ones, in agreement with the experiments. We find that parallel and antiparallel double-helical bulk structures of malto-oligosaccharides are feasible and of similar energy. The M3B model leads to a glass transition temperature (T_g) for glucose of 296 K in good agreement with experiment (304 K) and a T_g for 12 wt % water−glucose mixtures of 239 K in good agreement with experiment (240 K). These results suggest that these characteristic physical properties of carbohydrates can be described well without the use of explicit hydrogen bonds or electrostatics and also without introducing explicit directional forces in the nonbonded interactions. Molecular dynamics (MD) simulations with M3B are ∼7000 times faster than fully flexible atomistic simulations, making it practical to study large systems for long times. We expect that M3B will be of use for the study of the structure and dynamics of complex syrups and supercooled carbohydrates solutions.

1. Introduction

Carbohydrates are probably the most abundant form of organic matter in the biosphere, largely because of the abundance in the plant world of cellulose and starch, both homopolymers of glucose. The hydrolysis derivatives of starch are widely used in pharmaceutical and food technologies as syrups containing polydisperse mixtures of water and malto-oligosaccharides with degrees of polymerization (DP) ranging from 1 to ∼50. For these syrups, such physical properties as viscosity, glass transition, diffusion, hygroscopicity, and water activity depend strongly on their water content, sugar composition profile, and temperature.[1,2]

One of the most relevant properties of carbohydrates and their concentrated water mixtures is their ability to form glasses.[3] In most of their commercial applications the syrups are processed in the supercooled (rubbery) state and stored as glasses. The storage of food products in the glassy state was thought to prevent chemical degradation by the arrest of water diffusion. However, it has been proved in recent years that water diffusion in carbohydrate matrixes continues even below the glass transition.[4,5] The mechanism of water diffusion in deeply supercooled and glassy water−carbohydrate mixtures is not yet understood. We want to use computer simulations to aid in the elucidation of such processes at a molecular level. However, the limitation in the accessible simulation time-scales (on the order of a few ns, at most) with atomistic simulations constitutes a severe limitation to studying the system in the supercooled regime, even for the simplest binary water−glucose mixtures.

In this paper we present a coarse grain model and force field for α(1→4) D-glucans and their water mixtures that retains sufficient detail to represent the helical structure of the oligosaccharides and to describe glass formation (predicting a glass transition temperature in excellent agreement with experiment) while speeding up the calculations by about 7000 times.

The organization of the paper is as follows: in Section 2 we present the coarse grain model and its parametrization, in Section 3 we present and discuss results that validate and illustrate the scope of the coarse grain model, and finally, in Section 4 we present the main conclusions of this work.

2. Coarse Grain Model and Force Field Development

In recent years there has been a growing interest in the development of coarse grain superatom models for a variety of polymers. (See refs 6−8 for recent reviews.) These models consist of superatoms (also called beads) that represent groups of atoms, monomers, or even several monomeric units. The superatoms interact through effective potentials that take into account, in a mean field approach the effects of the missing degrees of freedom. The coarse grain model should reproduce relevant features of the atomistic one. In this work we consider structural and thermodynamic properties as target functions to

* Author to whom correspondence should be addressed. Phone: 1-626-395-2731. Fax: 1-626-585-0918. E-mail: wag@caltech.edu.
[†] E-mail: vale@wag.caltech.edu.

10.1021/jp0354752 CCC: $27.50
Published on Web 01/07/2004

be reproduced by the coarse grain model. The coarse grain model has two main purposes:

First, to study the structure and dynamical properties of systems for which an atomistic molecular dynamics simulation would be excessively expensive. Some of the uses we envision for the coarse grain model are (i) the analysis of the structure and dynamical properties of single oligosaccharide chains, (ii) the elucidation of the water distribution in sugar mixtures, (iii) the study of the dynamics of supercooled mixtures, and (iv) the determination of rheological properties of polydisperse mixtures with variable water content.

Second, to obtain equilibrated atomistic structures of slowly relaxing systems. The system is constructed and equilibrated with the coarse grain model and is subsequently mapped into the atomistic model. This "reconstructed" atomistic model is then relaxed with the atomistic force field to compute properties that require atomistic detail such as the structure factor *S(k)* or the free volume distribution.

There is no general recipe to construct a coarse grain model from the atomistic one. The strategy for coarsening varies according to the size associated with the mesoscopic particle and the properties of interest. The decisions to make in developing a coarse grain model are essentially related to

(i) the resolution of the model (how many atoms per bead),

(ii) the mapping of the coarse grain model from the atomistic model (i.e., defining the bead positions as a function of the atomic coordinates),

(iii) the selection of potential functions for the coarse grain Hamiltonian and,

(iv) the properties to be considered in the optimization procedure.

The systems that we aim to represent with the coarse grain model are glucose oligomers with 1→4 and/or 1→6 or 1→1 glycosidic bonds. The typical examples of homopolysaccharides with these linkages are the following:

(i) Amylose, the linear α(1→4) glucose homopolymer of starch.

(ii) Amylopectin, the α(1→6) branched α(1→4) glucan component of starch.

(iii) Glycogen, the storage form of glucose in animals that, as amylopectin, is an α(1→4) glucan with α(1→6) branching points.

(iv) Cellulose, the linear β(1→4) glucan produced by plants.

(v) Linear α(1→6) glucose homopolymers, such as those found in dextran.

In the present work, we present the complete derivation of the force field parameters for the α(1→4) D-glucans. Section 2.3.4 indicates the steps required to extend the presented parametrization to β(1→4) and α(1→6) glycosidic linkages.

We are interested not only in the modeling of the amylose polysaccharide but also of oligosaccharides such as cycloamyloses (the cyclic α(1→4) glucose oligomers) and the multiple hydrolysis products of starch (aqueous mixtures of malto-oligosaccharides with degree of polymerization DP ranging from 1 to ∼50, in varying proportions). Because of the size differences between water and the sugars, in particular for high-DP oligosaccharides, we want the coarse grain representation of the sugars to be realistic in a length scale of ∼3 Å, the size of a water molecule. A coarser representation of the sugars would drastically change the shape of the polysaccharides and, thus, the packing with water. The presence of water and the polydispersity of the system limit the degree of coarsening. The resolution must be at least that of a monomer resolution. This level of resolution is also required to study the local dynamics of the chains: at the atomistic level, they are usually mapped into the dynamics of the torsional angles around the glycosidic bonds ϕ and ψ, defined by H1−C1−O−C4 and C1−O−C4−H4, respectively. A monomer-based description allows a tight physical correspondence between the local modes in both representations.

In the following we present the development and atomistically informed parametrization of a coarse grain model for α(1→4) D-glucans. To guide us in the development of the coarse grain model, we considered that the proposed *superatom model* would fulfill the following requirements:

(1) Represent with the same level of coarsening oligosaccharides of varying degree of polymerization, from DP1 (glucose) up to any desired value.

(2) Correctly represent the connectivity of starch homopolymers (allowing both linear and branched chains).

(3) Provide unequivocal simple rules to map the atomistic model into the coarse grain.

(4) Provide a unique noniterative mapping back into the atomistic model, while leading to small uncertainty in the reconstructed atomic coordinates.

For the carbohydrate coarse grain force field, we ask that:

(1) The interaction between the coarse grain particles (beads) must be completely derived from simulations on the atomistic model. The coarse grain model must reproduce the density, cohesive energy, and structural characteristics of the atomistic model for a broad range of pressures.

(2) The interaction potential between the beads must be simple analytical functions of the bead coordinates.

(3) The model should be universal in its applicability to any polydisperse mixture of malto-oligosaccharide with the fewest number of parameters.

2.1. Mapping the Atomistic into the Coarse Grain M3B Model. We would prefer the simplicity of a single bead representation of the monomer. However, such a single spherical bead description would have several drawbacks. Because the shape of the glucose molecule departs greatly from a spheroid, a representation of the monomer by a single spherical bead would necessarily lead to a wrong shape of the glucose molecule as seen from a distance comparable to the molecular size (e.g., from another oligosaccharide or from a water molecule). One of the most important properties of carbohydrate mixtures is to form glasses. Spherical particles tend to form close-packed crystalline structures. Besides, we want the coarse grain model to be efficiently combined with atomistic simulations. This implies that the process of mapping back the position of the atoms from the beads should lead to an atomistic structure that can be relaxed in the time-scales accessible for atomistic simulations. If one monomer is mapped into one spherical bead, the reconstructed atomistic configuration would be poorly defined, and times comparable to the rotational time of the monomer would be required to relax the atomistic system. Another characteristic feature of oligo- and polysaccharides is their tendency to form helical structures. These structures have a defined handedness that cannot be distinguished if the backbone of the polymer is represented by a single bead.

The model we propose represents each monomer by three beads (M3B). We define the positions of the beads to correspond to the atoms C1, C4, and C6 in the atomistic model, as shown in Figure 1 for the dimer. This makes the mapping of the atomistic model into the coarse grain model straightforward. Intramonomer and intermonomer coarse grain bonds connect the superatoms. The latter are analogous to the glycosidic bonds in the atomistic oligosaccharides. The three beads per monomer of the M3B model constitute a well-suited representation for

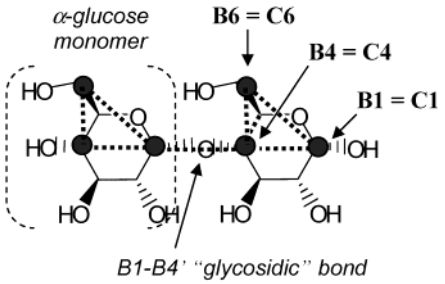


Figure 1. Overlaid representation of the atomistic and M3B model of maltose. The positions of the beads in M3B correspond to the positions of the carbons C1, C4, and C6 in the atomistic model. The three beads of each monomer are fully connected by coarse grain bonds (dashed thick lines). A coarse grain glycosidic bond links the monomers in the polymer chain.

oligosaccharides containing 1→4, 1→6, and the less frequent 1→1 linkages. M3B represents the backbone of the oligomers by two different beads (those positioned in C1 and C4 for (1→4) glucans), and thus left- and right-handed helical structures correspond to different M3B structures. The selection of beads positions on those of C1, C4, and C6 atoms provides a good representation of the shape, volume, and exposed surface of the glucose monomer, as will be shown at the end of Section 2.3.3.

As the superatoms are below the monomer level, oligosaccharides with arbitrary DP can be represented with the M3B model. The rules to convert the *atomistic* oligosaccharide into the M3B one are the following:

(1) For each monomer of the chain identify the C1, C4, and C6 atoms.

(2) Position the beads *B*1, *B*4, and *B*6 on the coordinates of *C*1, *C*4, and *C*6, respectively.

(3) Make bonds to connect each pair of beads of the same monomer.

(4) Identify the carbons involved in atomistic 1→*X* glycosidic linkages ($X = 4$, 6, or 1).

(5) Make the corresponding $B1-BX'$ intermonomer bonds (the prime implies that they belong to a different monomer).

Following this scheme, any arbitrary cellulose, starch, or glycogen hydrolysate can be constructed. Note that the model distinguishes the reducing and nonreducing end of chains of oligosaccharides, represented by different beads (*B*1 and *B*4, respectively).

Although M3B has a tight correspondence between the bead and carbon positions, we do not assume in the parametrization of the bead interactions any direct correspondence between a group of atoms of the monomer and the beads in the M3B model. Rather, the interactions between the beads are parametrized to reproduce structural and energetic properties of the *whole* atomistic model.

The configuration and dynamics of malto-oligosaccharides are usually described in terms of the atomistic glycosidic torsion angles ϕ and ψ, defined by H1−C1−O−C4 and C1−O−C4−H4, respectively.[9] These torsions, indicated for maltose (DP2) in Figure 1, correspond to the relative movement of the two relatively rigid glucose monomers around the glycosidic bonds. The rotation of ϕ and ψ can be mapped into the variation of the coarse grain angles and torsions defined by the *C*1, *C*4, and *C*6 of the two residues. If we consider the monomer to be rigid, each coarse grain angle would be a linear function of either ϕ or ψ (see Figure 1S of the Supporting Information). The coarse grain torsions $C6-C4-C1'-C6'$ (641′6) and $C4-C1-C4'-C1'$ (414′1′) are similar nonlinear functions of both ϕ and ψ (see Figure 2, or Figure 1Sb of the Supporting Information

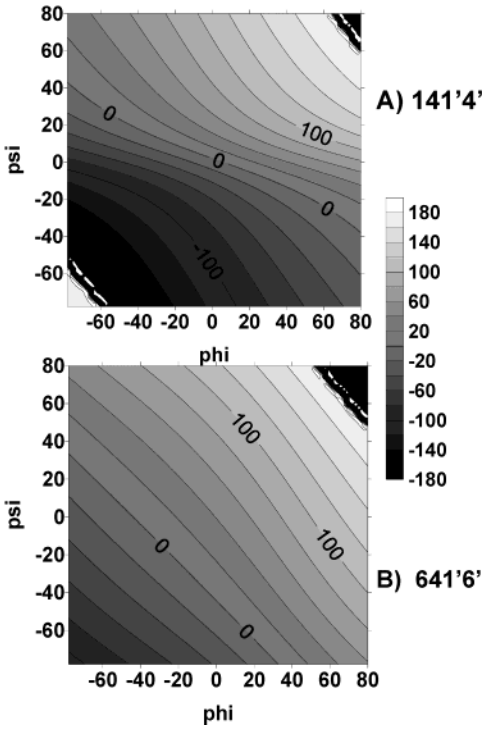


Figure 2. Contour maps showing how the torsions of crucial atoms for the coarse grain model are affected by the rotation of the glycosidic torsions ϕ and ψ. The upper panel (A) shows the torsions defined by carbons C1, C4, C1′, and C4′ (141′4′) and lower panel (B) shows the 641′6 torsion. The maps correspond to the rotation of the glycosidic torsional angles in an otherwise rigid α-maltose whose configuration was taken from the X-ray crystal structure in ref 10.

for the full [ϕ, ψ] range). The relaxation of the internal coordinates of the glucose residues will lead to variations of these simple relationships, though they can still be considered a guidance for the mapping of the atomistic model into the bead degrees of freedom. Thus, *the coarse grain angles and torsions are an alternative base to* ϕ *and* ψ for the analysis of the configurational change in oligosaccharides: In M3B, the main modes of the polysaccharide chain related to the local dynamics around glycosidic bonds are mapped into the bending of angles and torsions between three and four consecutive beads corresponding to different monomers.

2.2. Reconstruction of the Atomistic Model from M3B. The atomistic glucose molecule is chiral, that is, it cannot be overlapped with its mirror image. The M3B model of glucose is planar, and hence is not chiral, so that it can be mapped back to either L- or D-glucose. However, only one of the two enantiomers, D-glucose, exists in natural products and their degradation products. Hence, it is sufficient to know the position of the three atoms C1, C4, and C6 to completely define the orientation of the cyclic molecule.

The position of all 24 atoms of the glucose molecule can be represented as a function of three linearly independent vectors. We choose as that basis the vectors along $C4-C1$, $C4-C6$ and the normal to the plane defined by these two:

$$\vec{u} = \vec{r}_{C1} - \vec{r}_{C4}, \quad \vec{v} = \vec{r}_{C6} - \vec{r}_{C4} \quad \text{and} \quad \vec{w} = \vec{u} \times \vec{v} \qquad (1)$$

The position of each atom of the glucose molecule was written in the basis of these three vectors, via the matrix **P**. As the molecule stretches and vibrates, **P** will change with time. However, we can define the matrix corresponding to the positions of the most stable configuration of the molecule as a template for the reconstruction of the atomic coordinates,

$$\begin{bmatrix} x_1 & y_1 & z_1 \\ \dots & \dots & \dots \\ x_{24} & y_{24} & z_{24} \end{bmatrix} = \begin{bmatrix} p_1^{\,u} & p_1^{\,v} & p_1^{\,w} \\ \dots & \dots & \dots \\ p_{24}^{\,u} & p_{24}^{\,v} & p_{24}^{\,w} \end{bmatrix} \begin{bmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{bmatrix} \quad (2)$$

where x_i, y_i, z_i, are the coordinates of ith atom of the glucose molecule, u_x, u_y, u_z are the coordinates of the vectors, and p_i^u is the u component of the position of the particle i in that basis. Hence, the three beads in M3B uniquely determine the orientation of the glucose monomer, as they correspond to the positions of *C*1, *C*4, and *C*6 of the atomistic model. To reconstruct the glycosidic linkages, the same template is used though the corresponding H_2O atoms are eliminated and the residues are connected with a glycosidic bond.

Because of the relative rigidity of the ring and the impossibility of interconversion from *R* to *S* stereoisomeric forms, the mapping back from the bead model to the atoms is well defined. As an illustration, Figure 3 shows the matching of the original and reconstructed atomistic models of DP11. The root mean square (rms) displacement between the original and reconstructed DP11 models is just 0.34 Å. The only significant uncertainty while reconstructing the molecules is related to the rotation of the hydroxyl moiety around *C*6. Considering the intrinsic indeterminacy while reconstructing a flexible residue of 234 atoms from only 33, the matching is excellent.

2.3. M3B Force Field Parametrization for α(1→4) Glucans. The coarse grain force field for α(1→4) glucans was parametrized completely from atomistic simulations of the monomer (α-glucose), the dimer (α-maltose), and α-DP4 malto-oligomers in the gas and condensed phases.

2.3.1. Atomistic Simulations. Constant volume and constant pressure atomistic molecular dynamics[11] simulations were performed using Cerius2.[12] We used the DREIDING force field,[13] except that we used explicit off-diagonal van der Waals interactions between any carbohydrate oxygen and hydroxylic hydrogens with the parameters $D = 0.037\,83$ kcal/mol, $R = 2.4$ Å, and $\xi = 12.76$ based on fitting crystal structure and amorphous data for a series of carbohydrates. The hydrogen bond parameters were taken as $D_{OO} = 2.5$ kcal/mol and $R_{OO} = 3.2$ Å. Partial charges on carbohydrate atoms were obtained by charge equilibration[14] in a water box with a density of 1 g/cm^3 and $T = 300$ K, averaged over a 10 ps NVT simulation. Partial charges on water molecule were obtained by LMP2 (localized MP2) quantum mechanics calculations, $qO = -2qH = -0.7287$ eu. For periodic systems, the long-range interactions were evaluated with Ewald sums. The integration of the equations of motion was done with the Verlet Leapfrog algorithm. The time step was 1 fs for all atomistic simulations.

For atomistic glassy systems we prepared a set of independent equilibrated amorphous *atomistic* samples of DP1 (five samples), DP2 (three samples), and DP4 (five samples) at 300 K and 1 atm. Each glassy system was prepared and equilibrated following a thorough procedure involving a series of compression−expansion and annealing steps.[15] The number of molecules per cell was 32 for DP1, 20 for DP2, and 18 for DP4. The densities calculated for the five equilibrated glucose samples are 1.48 ± 0.02 g/cm^3, which compares well with the 1.52 g/cm^3 experimental value in the same thermodynamics conditions.[16] Reference atomistic structures for the parametrization of the coarse grain model were prepared optimizing the atomic and cell coordinates to minimize the energy of each cell of DP1, DP2, and DP4 in the presence of an isotropic stress field corresponding to pressures $p = -1$, 0, 1, 2, 3.5, 5, 10, 15, and 20 GPa. The minimization was considered converged when the three following criteria were satisfied: the total residual force was below 0.1 kcal/(mol Å), total rms displacement in two successive steps was below 0.003 Å, and the energy difference was below 0.001 kcal/mol.

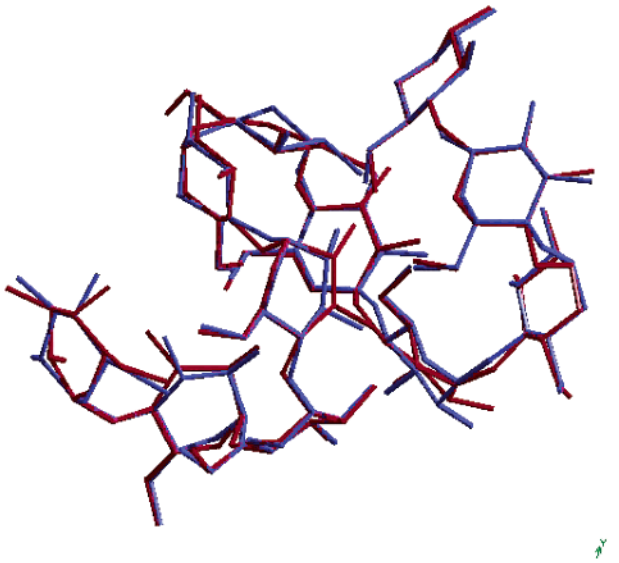


Figure 3. Original (blue) and reconstructed (red) atomistic structures of DP11. A coarse grain structure (not shown) was mapped from the original atomistic structure and further mapped back into the atomistic structure to produce the reconstructed structure. Hydrogen atoms are not shown for the sake of simplicity. The rms deviation between all the atoms of the two structures was 0.34 Å.

For atomistic single molecule simulations we carried out a reference set of gas-phase NVT simulations of the atomistic malto-oligosaccahrides DP1, DP2, and DP4 at $T = 300$ K, for times of 2 to 60 ns.

2.3.2. M3B Energy Expression. The interaction between coarse grain particles is expressed as a sum of valence and nonbonded terms,

$$E_t = E_{nb} + E_v \quad (3)$$

The valence energy terms include two-body interactions between all connected beads (bond terms), three-body interactions (bending angle between three consecutive connected beads), and four-body interactions (torsion angle between four consecutive connected beads).

$$E_v = E_{bond} + E_{angle} + E_{totsion} \quad (4)$$

We considered the coarse grain bonds and bending angles to be harmonic

$$E(r) = \frac{1}{2}k(r - r_o)^2 \quad (5a)$$

and

$$E(\theta) = \frac{1}{2}k_\theta(\theta - \theta_o)^2 \quad (5b)$$

where r and θ are the distance and angle between connected coarse grain particles, r_o and θ_o are the coarse grain bond and angle equilibrium positions, and k and k_θ are the coarse grain bond and bending angle constants.

For the four-body valence interactions, we considered each coarse grain torsional angle energy to be represented by a sum of shifted dihedral functions,

$$E(\varphi) = \sum_j \frac{1}{2} B_j (1 + \cos(j\varphi - \varphi_j^o)) \quad (6)$$

where the coarse grain torsional angles $\varphi = \varphi_{inkl}$ is the angle between the planes formed by *ink* and *nkl* (*n*, *k*, *l* are *B*1, *B*4, or *B*6), the sum is over the different *j* terms of the potential, each with a different integral periodicity *j*, barrier B_j, and phase φ_j^o. As is usual for atomistic force fields, we considered all the possible torsions around a coarse grain bond, summed them, and then scaled the total torsional potential around each coarse grain bond by the number of torsions.

The nonbonded interactions consist of two-body terms for all the pairs of beads that are not related by a bond or a bending angle. Thus, the interactions between (1−4) nearest neighbors include both the nonbonding and the torsional potential contributions. Whereas our atomistic simulations distinguish between electrostatic, van der Waals, and hydrogen-bonding terms for the nonbonding interactions, the coarse grain model uses only a single nonbonding term. Thus, we consider the superatoms to be neutral and with no special hydrogen bond terms (there are no H or O atoms in the bead description). For the nonbonded interaction between superatoms we chose the Morse potential,

$$V(R_{ij}) = D_o\{(e^{-0.5\alpha(R_{ij}/R_o-1)})^2 - 2(e^{-0.5\alpha(R_{ij}/R_o-1)})\} \quad (7)$$

where R_o is the distance of minimum energy (D_o) and α is a measure of the curvature of the potential around R_o. The lower the value of α, the softer is the potential. In the condensed phase, we expect the density of the system to be mainly affected by the value of R_o, the cohesive energy by D_o, and the compressibility by α. We parametrized the interaction for each bead type and use geometric combination rules for parameters of the cross interactions: and

$$D_{o,ij} = \sqrt{D_{o,i} D_{o,j}} \quad (8a)$$

$$R_{o,ij} = \sqrt{R_{o,i} R_{o,j}} \quad (8b)$$

$$\alpha_{ij} = \frac{1}{2}(\alpha_i + \alpha_j) \quad (8c)$$

The nonbonded interactions were truncated with a spline function. The value of the cut off radius, R_{cutoff}, considered through the parametrization was 16 Å, which is well above that required for the decay of the intermonomer interactions. The optimal cutoff value for the nonbonded interactions of the coarse grain particles was determined at the end of the parametrization procedure to be 12 Å (Section 2.3).

2.3.3. M3B Force Field Parametrization. The scheme we followed to derive the coarse grain force field is schematized in Scheme 1. The total number of parameters was not known in advance, because we did not assume, for example, that the bead *B*1 interacts in the same way when it is at the end of the chain or in the middle of it. Our procedure selected the minimum set of parameters that match the coarse grain results with those of the atomistic simulations of the same system.

We partitioned the force field development into four stages:

Step 1 computes a first estimate of the nonbonding parameters from the interaction of atomistic glucose pairs in a vacuum at different distances and orientations.

Step 2 uses a Monte Carlo simulated annealing procedure to obtain the coarse grain nonbonded parameters that minimize the difference in cohesive energy of sets of glassy structure of oligosaccharides and their M3B-mapped counterparts.

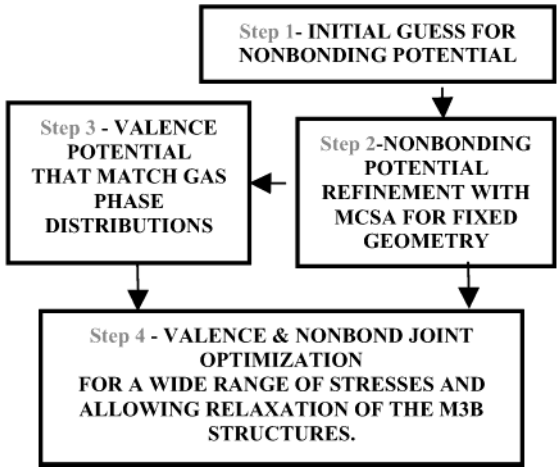
Step 3 computes the valence parameters by Boltzmann inversion of the distribution of the coarse grain bonds, angles, and torsion distribution from atomistic NVT simulations of single molecules at the atomistic level to obtain effective valence potentials for the coarse grain particles.

Step 4 refines all M3B parameters simultaneously, scaling them to minimize a cost function that accounts for the difference in cohesive energy, density, cell parameters, and bead positions evaluated for the atomistic (reference) and relaxed M3B models for a set of amorphous structures of DP1, DP2, and DP4 over a broad range of pressures, from −1 to 20 GPa.

The details of the optimization procedure are available in the Supporting Information section. The final coarse grain parameters are listed in Tables 1−4.

We found that the nonbonded Morse parameters for the same superatom (*B*1, *B*4, or *B*6) depend on whether the monomer is isolated (DP1, monomer coordination mc = 0), at the terminus

SCHEME 1: Coarse Grain Carbohydrate Parameters Were Refined by the Series of Steps Indicated in This Flow Chart[a]



[a] MCSA in step 2 stands for Monte Carlo simulated annealing.

TABLE 1: M3B Morse Parameters (Equation 7) and Masses of the Saccharide Beads

bead	mc[a]	force field type	mass (amu)	Ro (Å)	Do (kcal mol^{-1})	α
*B*1	0	*B*1_*A*	75	5.13	2.05	11
*B*4	0	*B*4_*A*	75	6.11	1.95	10.5
*B*6	0	*B*6_*A*	30	4.63	1.79	11
*B*1	1	*B*1_*B*	75	4.79	1.25	11
*B*4	1	*B*4_*B*	75	5.65	1.22	10.5
*B*6	1 or 2	*B*6_*B*	30	4.70	1.19	11
*B*1	2	*B*1_*C*	66	5.76	1.25	11
*B*4	2	*B*4_*C*	66	6.73	1.22	10.5

[a] Monomer coordination: 0 corresponds to the free monomer (DP1), 1 to the initial and terminal monomers of the chain, and 2 to the glucose residues linked to two monomers.

TABLE 2: M3B Bond Parameters (Equation 5)

bond type	R_o (Å)	k_b (kcal mol^{-1} Å^{-2})
14[a]	2.93	425
16[a]	3.69	235
46[a]	2.60	435
14′ [b]	2.49	410

[a] Intramonomer coarse grain bond. [b] Intermonomer (i.e., "glycosidic") coarse grain bond.

TABLE 3: M3B Angle Bending Parameters (Equation 5)

angle type	θ_0 (deg)	K_θ (kcal mol^{-1} radian^{-2})
141′	142	75
414′	121	150
641′	110	100
614′	105	110

TABLE 4: M3B Torsion Angle Parameters (Equation 6)

torsion type	B_1 (kcal mol^{-1})	φ_1° (deg)	B_2 (kcal mol^{-1})	φ_2° (deg)	B_3 (kcal mol^{1})	φ_3° (deg)
414′1′	1.0	55	1.8	−20		
641′6′	4.3	−135	2.8	130		
4614′	30.0	−58				
1641′	42.0	40			4.0	30
641′4′	3.6	93	2.5	−145		
614′1′	1.0	27	2.1	−115		
6141′	10.1	−65	1.3	71		
6414′	15.7	112	4.3	−29		
41′4′1″	15.0	−160				

of a chain (mc = 1), or at an internal position of the chain (mc = 2). The parameters depended on the monomer coordination but not on the degree of polymerization of the molecule. That is, the DP2 and DP4 parameters were transferable. Table 1 lists the Morse parameters for M3B model. Note that residues with different monomer coordination have different overall masses, because of the loss of one water molecule per glycosidic bond formed. We assume that the mass of *B*6 corresponds approximately to that of the exocyclic group (30 amu) and the rest of the mass is distributed equally between the ring components *B*1 and *B*4 (75 amu if terminal, 66 amu if not terminal, to account for the loss of water in the polymerization).

The equilibrium parameters for the bond and angle terms (see Tables 2−3) after optimization to fit the condensed phase equation of state are similar to those obtained from Boltzmann inversion of the populations of the gas-phase molecular dynamics (see Tables 1S, 2Sa, and 2Sb of the Supporting Information.) However, the angle and bond force constants are considerably larger. This was required to prevent the collapse of the molecules at high pressure (and an unphysical expansion for negative *p*). The net result is that the condensed phase optimized ones were just half of the atomistic value (Table 2), 700 kcal mol^{-1} Å^{-2}, whereas the value from fitting the gas-phase distributions (Table 1S) was about four times lower than the atomistic value.

Considering the masses shown in Table 1 and the force constants of Table 2, the time step for the M3B simulations can be as high as 12 fs (allowing 6 steps for the shortest vibrational period) while still providing an accurate integration of the equations of motion. In the M3B simulations presented here, we have used integration time steps of 10 fs (unless otherwise is indicated).

The final Morse parameters allowed a decrease of the spline cutoff to 12 Å. The use of spline instead of Ewald plus the increase by at least 10 times in the time step and the decrease of the number of particles to $^1/_8$ with respect to the atomistic model make M3B at least 3 orders of magnitude faster than atomistic simulations. To compare the speed of the two models consider that a 1 ps NVT simulation of a periodic system composed of 20 DP4 molecules (1566 atoms or 216 beads) using Cerius2 on an SGI-origin 10000 processor takes 2.6 s for M3B and 19 046 s for the atomistic model. The speedup is greater than 7300 times.

The optimized M3B parameters provide very good agreement with the atomistic simulations for the density, cohesive energy, and structural parameters over the full range of pressure. Figure 4 shows the ratio between the densities of amorphous glucose

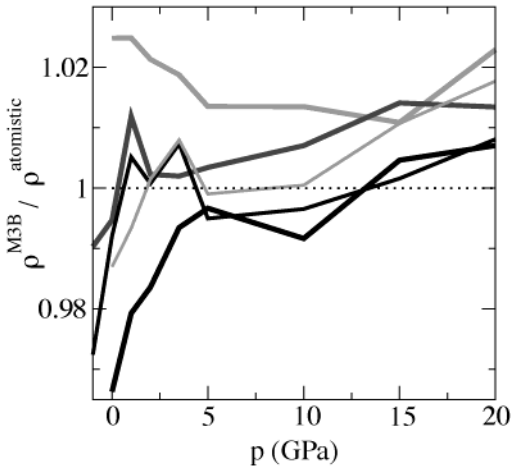


Figure 4. Ratio between the density of amorphous glucose computed for energy-minimized structures with the coarse grain (ρ^{M3B}) and the atomistic ($\rho^{atomistic}$) models in a wide range of pressures. Each of the five solid lines corresponds to a different independent amorphous cell.

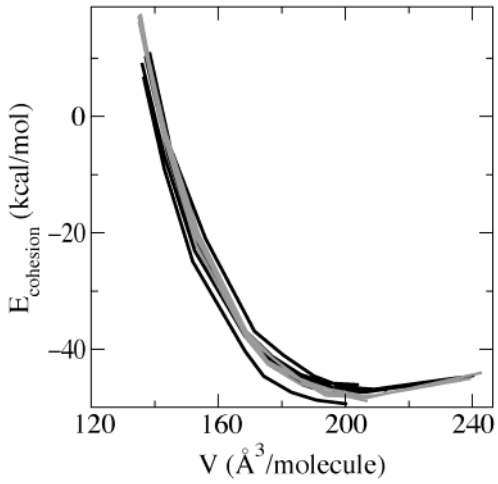


Figure 5. Cohesive energy of five independent atomistic (black lines) cells of amorphous glucose under compression and their coarse-grain-optimized counterparts (gray lines).

obtained with M3B and atomistic force fields, as a function of pressure. The overall difference in density is less than 2% up to pressures of 20 GPa. The cohesive energy and the compressibility were also in very good agreement with the atomistic model, as can be seen for amorphous glucose in Figure 5. Similar agreement was obtained for DP2 and DP4.

The final set of coarse grain parameters guarantees that the bonds and angles of the molecule are within 2% of the atomistic values for the *whole* pressure range, as can be seen in Figure 6. Figure 7 shows the average deviation of the cell parameters between the coarse grain and atomistic model, $\xi_x(p)$, as a function of pressure *p*

$$\zeta_x(p) = 100\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\frac{x_{i,p}^{atom} - x_{i,p}^{M3B}(\{V\})}{x_{i,p}^{atom}}\right)^2} \qquad (9)$$

where $x = a, b, c, \alpha, \beta, \gamma$ were evaluated for the atomistic and *relaxed* M3B model at the pressure *p* for the *m* independent amorphous cells. The mean percent deviation of the cell parameters, $\zeta_x(p)$, was below 2.5% in the entire pressure range.

The average root mean square deviation (rms) of the bead positions with respect to the atom-mapped ones computed for

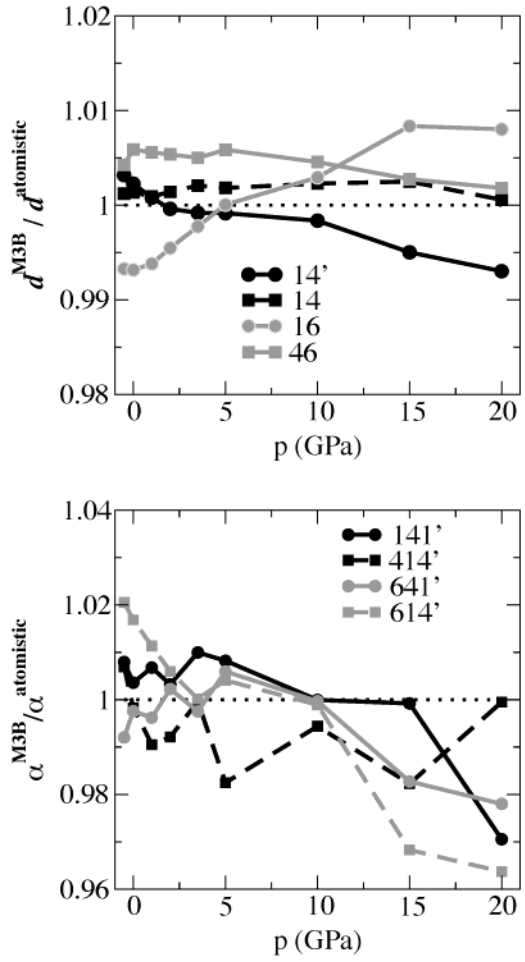


Figure 6. Ratio between coarse grain and atomistic bond distances (upper panel) and bending angles (lower panel) of DP2 as a function of pressure. The data shown correspond to the average over the amorphous cells for three independent cells.

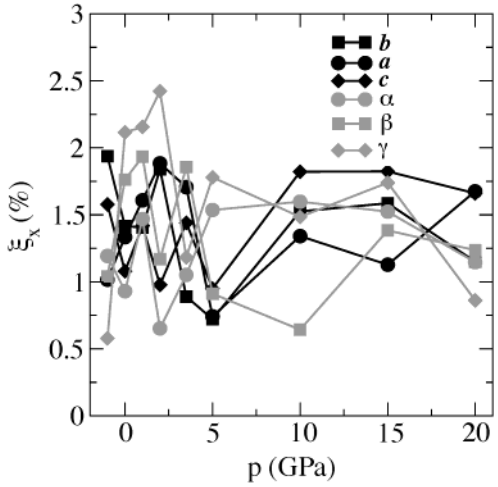


Figure 7. Mean percent deviation of the cell parameters between the atomistic and coarse grain amorphous glucose cells under compression.

each at pressure p is defined for each of the independent amorphous samples by

$$\mathrm{rms}(p) = \frac{1}{N_B}\sqrt{\sum_{k=1}^{N_B}(\vec{r}^{\,A}_{k,p} - \vec{r}^{\,M3B}_{k,p}(\{V\}))^2} \qquad (10)$$

where N_B is the number of beads. Figure 8 displays these two

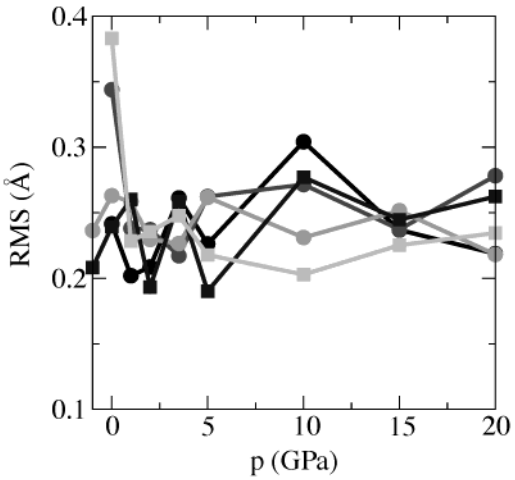


Figure 8. Root mean square difference between the atomistic- and M3B-minimized amorphous glucose structures in the range of pressure employed in the coarse grain force field parametrization. The five curves correspond to the different amorphous glucose samples employed in the parametrization.

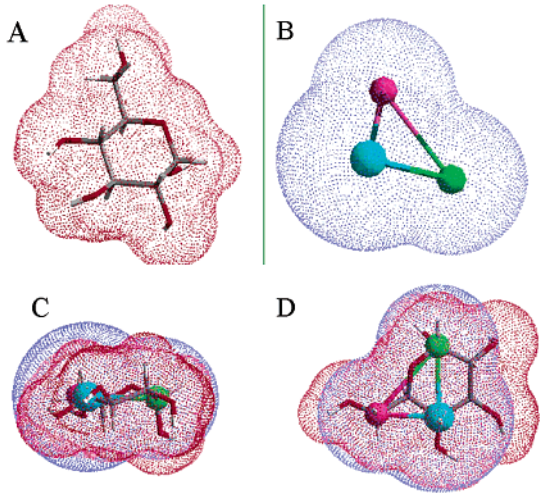


Figure 9. Connolly surfaces of the α-glucose molecule: atomistic (red) and M3B (blue). (A) Atomistic glucose. (B) M3B glucose. Lower panels, C and D, show "side" and "top" views of the two superimposed models, with their respective Connolly surfaces. The solvent probe radius was 1.4 Å.

indicators for glucose. The average deviation of the bead positions is around 0.25 Å, a very good number considering that we are coarsening the length scale 1 order of magnitude above that value (see the Morse R_o parameters, Table 1).

Although we selected the number of beads to represent each glucose molecule and their mapping from the atomistic structure in M3B, we asserted that the model is able to capture the essential shape and dimensions of the glucose molecule. Figure 9 shows the molecular surface of the glucose in the atomistic and bead representation. We defined the molecular surface as the Connolly surface[17] of the molecule rolled by a probe sphere of radius $R_p = 1.4$ Å. For both the atomistic and M3B models, the radius of the particles was taken as *R*o of the van der Waals interaction in the corresponding force field. The molecular volume and surface of the glucose molecule in the atomistic and coarse grain representation are summarized in Table 5. The good agreement in the molecular volume is not surprising, because the density was one of the properties taken into account in the parametrization. Nevertheless, the similarity in shape and the very good agreement in the value of the accessible surfaces

TABLE 5: Molecular Volume and Molecular Surface of α-glucose Obtained from the Connolly Surfaces Shown in Figure 9

model	volume (Å^3)	surface (Å^2)
atomistic	188	183
M3B	189	171

TABLE 6: Morse Parameters for Coarse Grain Water

bead type	R_0 (Å)	D_0 (kcal/mol)	α
W	3.77	1.15	8

of the molecule in the two representations are indicators of the quality of the M3B model.

2.3.4. Extension of the M3B Force Field to Other Glycosidic Linkages. Glucans containing α and β (1→*X*) linkages, with $X = 1, 4, 6$ can be mapped straightforwardly into the M3B model. We presented above a complete parametrization of the M3B model for the glucose monomer and the α(1→4) glucans. To describe the interactions of molecules containing other linkages, the force field parameters should be derived in a way analogous to that shown for the α(1→4) glucans. As an example, a complete coarse grain description of the α(1→6) glucans requires the determination of the parameters for valence terms absent in the present parametrization: the bond 16′, angles 161′, 461′, 416′, and 616′, and all torsions *Z*61′*Y*′ where $Z = 1, 4$ and $Y' = 6, 4'$. All these terms can be obtained from atomistic vacuum and bulk simulations of isomaltose (the α(1→6) dimer of glucose) following the same procedure described in the previous subsection for maltose, just by introducing a new atom type *B*6_*C* for the bead *B*6 that participates in a bead−glycosidic bond.

The extension of the M3B model to describe glucans linked by carbon atoms other than 1, 4, and 6, for example, 1→2 and 1→3 linkages, may not be straightforward. A possible extension for 1→2 linkage is to consider that C1 and C2 are both mapped into the *B*1, and thus the atomistic 1→2 glycosidic bond would be mapped into the bead 1→1, and all the parametrization should be derived, nonbonded and valence terms, from scratch from simulations of the corresponding 1→2 dimer of glucose. We have not tested the models for 1→*X* with $X = 2$ or 3, and we cannot guarantee that the representation of three beads would be enough for a decent description of the shape and flexibility of such molecules.

2.4. Force Field Parametrization for Coarse Grain Water. Most systems containing carbohydrates also contain water. To obtain a description of water compatible with M3B, we propose a simple coarse grain model in which each water (*W*) molecule is represented by a single bead interacting via a Morse potential. The three parameters of the Morse potential of *W* were adjusted to reproduce the experimental density, intermolecular energy, and diffusion coefficient of water at 300 K and 1 atm. The diffusion coefficient was considered in the parametrization because of the relevance of water transport in concentrated sugar mixtures.

To derive the parameters, we performed a series of NPT molecular dynamics simulations at 300 K and $p = 0$ using the bead model *W* with 112 molecules per cell. The density, cohesive energy, and diffusion coefficient from a 0.2 ns coarse grain simulation were compared with the experimental information available for these properties. The resulting parameters are summarized in Table 6. Table 7 compares the results obtained with this bead representation of water and the experiment[18,19] at 298 K and 1 bar. The diffusion coefficient of the coarse grain model was computed from the log−log plot of the mean square displacement with time.

TABLE 7: Density, Cohesive Energy, and Diffusion Coefficient for the Coarse Grain Model of Water and the Experimental Values at 300 K

water	ρ (g/cm^3)	*E* (kcal/mol)	*D* (10^{-5} cm^2/s)
W	0.97(2)	−10.2(1)	1.7
exptl	0.996[a]	−10.517[a]	2.4[b]

[a] From ref 18. [b] From ref 19.

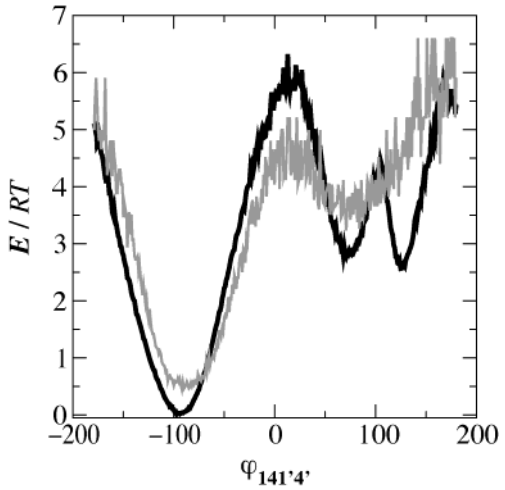


Figure 10. Potential of mean force (*E*) around the torsion 141′4′ for the atomistic (black) and M3B (gray) DP2 molecule. Curves obtained from 60 ns NVT simulations of the molecule at $T = 300$ K.

3. Validation of M3B: Results and Discussion.

3.1. Torsional Distribution for α-Maltose with M3B and Atomistic Models. Figure 10 shows good agreement between the potential of mean force *E* for the torsion angle 141′4′ obtained from the angle distributions $P(\varphi_{141'4'})$ during 60 ns NVT simulations of both the atomistic and M3B α-maltose molecule at 300 K,

$$\frac{E(\varphi_{141'4'})}{RT} \propto -\ln(P(\varphi_{141'4'})) \tag{11}$$

This plot corresponds to the projection of the conformational map on the angle 141′4′. Equivalent quality results were obtained for the other torsions. More information is obtained from the analysis of the conformational distribution as a function of two bead variables, for example, 141′4 and 641′6′, as shown in Figure 11. The distribution functions for the bead and atomistic models are very similar, though the M3B configurations show a more pronounced linear dependence of 641′6′ versus 414′1′ in the entire range than does the atomistic simulation. The agreement between the two models could be improved by introducing a coupling term between the bead angles and torsions in the M3B energy expression. However, for present purposes we prefer to keep the M3B model simple.

Although the coarse grain model reproduces well the energetics of the atomistic model, we must note that this condition is not sufficient to ensure that the two models will have the same dynamical behavior. This is because the discarded modes may have important effects on the dynamics. In particular, the reduction in the number of modes can reduce the internal friction in the dynamics. This difference is evident in the power spectrum of the velocity ($\vec{u}$) autocorrelation function for bulk DP4 in the two representations,

$$S(\nu) = \frac{2}{kT}\lim_{\tau\to\infty}\int_{-\tau}^{\tau} \langle \vec{u}(t)\vec{u}(0)\rangle e^{-i2\pi\nu t}\, dt \tag{12}$$

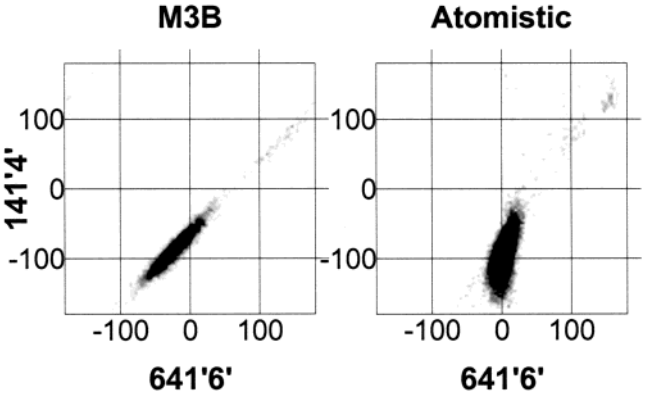


Figure 11. Joint distribution of the most relevant torsion angles of the coarse grain positions, 141′4′ and 641′6′, evaluated over 60 ns NVT trajectories of the DP2 molecule at 300 K. The density of gray indicates the population of each configuration, in a logarithmic scale. Both models predict that states with negative values of 414′1′ torsions, which give rise to left-handed helices, are more stable than the positive (right-handed) ones.

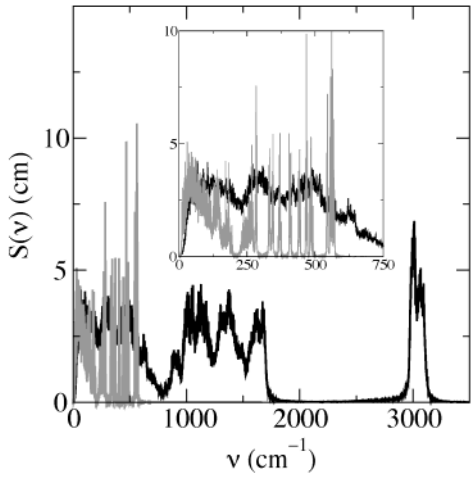


Figure 12. Velocity autocorrelation spectra for atomistic (black) and M3B (gray) amorphous bulk DP4 at 343 K and $\rho = 1.44$ g/cm^3. The time step for the coarse grain simulation was 5 fs.

shown in Figure 12. Not only the high-frequency modes (above 800 cm^{-1}) are absent in the M3B model but it also shows a spiky spectrum for the intermediate frequency range (200−1000 cm^{-1}) corresponding mainly to the soft internal modes of the molecule. The coupling between the modes for the existing particles and the one deleted in the construction of the M3B model would have the effect of a friction on the dynamics of the bead coordinates [20] and may eventually be added by a frictional term in the coarse grain dynamics.

3.2. Helical Folding in Malto-Oligosaccharides. Amylose crystallizes in a diversity of helical polymorphs.[21−23] These helices can be characterized by the number of monomers per turns, *n*, and the rise per monomer, *h*. The pitch *p* is defined as the rise per turn, $p = hn$. In principle, amylose helices can be left- or right-handed. However, the features corresponding to right-handed helices are absent from the experimental X-ray spectra of amylose crystals.[24] Double-stranded helices may exist in parallel or antiparallel configuration. The configurations found in amylose crystals are limited to left-handed single-stranded (V-amylose) and parallel double-stranded left-handed helices (A and B polymorphs). The antiparallel configuration was observed only for modified DP6 crystallized with barium triodide,[25,26] never for native starches or recrystallized amylose chains.

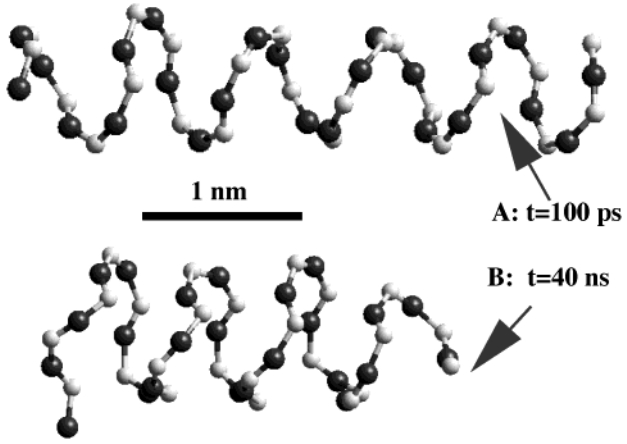


Figure 13. Helical structures obtained with M3B of DP24 in a vacuum at 300 K for two stages of the trajectory. Only the backbone (*B*1 and *B*4) beads are shown. The initial structure was constructed and minimized with M3B. The total energy of structure B is 1.2 kcal/(mol monomer) lower than that of structure A.

The configuration of the individual helices in double-stranded structures is experimentally found to be approximately twice as extended as that for single helices in V-amylose. Although single-stranded helices are stabilized by interactions between contiguous monomers and between turns, the structure of double helices is stabilized by the interactions between monomers of different chains. The crystalline behavior of amylose and malto-oligosaccharides is usually rationalized in terms of preferential hydrogen bonding. In this section we study the performance of M3B, which lacks not only hydrogen bonds but also atoms, in the prediction of the multiple helical structures of α(1 − 4) glucans.

3.2.1. Free Chain. It is generally presumed that oligosaccharides chains preserve at least some of their helical character in water solution.[24,27,28] The X-ray data of 2% w/w amylose aqueous solutions at 70 °C[24] is consistent with the existence of left-handed, single-stranded helices 6−12 monomers long. Double-helix signals were absent from the X-ray data of the solution spectra, though they were concomitant with the gelation of the amylose solution at 40 °C.[24] Right-handed double helices produce a characteristic spectrum that was never detected in the experiments.

Single chain simulations with explicit water are possible in the M3B model. However, the number of water molecules required to simulate dilute solutions of a medium size oligomer are quite high, say more than 10 000 water molecules per chain for a 2% w/w DP24 solution. In this work we restricted the single chain study to vacuum simulations of oligomers and studied the evolution of the system over dozens of nanoseconds. As the solvent is expected to screen, at least partially, the intrapolymer interactions, we do not attempt to extract quantitative information on the configuration of malto-oligosaccharides in solutions from vacuum simulation. We performed 40 ns of NVT coarse grain dynamics simulations of DP24 α(1→4) glucan vacuum at 300 K using time steps of 5 fs. The initial configuration was a left-handed helix, a local minimum of the M3B force field. Figure 13 shows two typical snapshots of a DP24 backbone, obtained after 100 ps and 40 nanoseconds of the simulation. In both cases the molecular backbone, defined by the superatoms *B*1 and *B*4, displays a helical structure, although they differ in the number of turns (*n*) and pitch (*p*): the average *n* is 5.5 for structure A and slightly less than 7 for structure B. The average pitch of the single-molecule helix we found in DP24 was 8.3 Å after 100 ps of dynamics (Figure

13A) and 7.2 Å for the structure after 40 ns of equilibration (Figure 13B). The values obtained from X-ray experiments of V_a and V_h 6-folded V-amylose crystals[22,23] were 7.91 and 8.05 Å, respectively. This compression of the helix is favored by an increase of the nonbonded interactions, at the expense of the valence energy of the molecule that favors $n \sim 6$. The addition of solvent may compensate for the nonbonded interactions, thus stabilizing the more extended helix.

Experimentally, polymorphs of amylose with $n = 4-8$ were found in different compounds.[22,23,29−33] Note that in the M3B model the rise per monomer of the helix, *h*, is directly related to the backbone torsion angle, 141′4′, where the number of monomers per turn, *n*, depends also on the backbone angles 141′ and 414′. The configurational [*n*,*h*] space available to the molecule in the M3B model depends, thus, on the range of the backbone torsion and angles and is tuned by the nonbonded interactions. The requirement that we impose on the coarse grain force field to reproduce the atomistic angles at high levels of compression produces a stiffening of the coarse grain angle force constants with respect to the values computed from gas-phase simulations (compare values in Table 2Sa and 2Sb of the Supplementary Information with Table 3) This stiffening reduces the accessible configurational space of the coarse grain model. The use of the gas-phase parameters for the coarse grain angles instead of the compression-adjusted ones might restore configurational flexibility to the model. We illustrate this point in Section 3.2.2 with the V-amylose structure.

Left- and right-handed helices correspond to different M3B structures with opposite signs of the backbone torsion 141′4′: negative for left-handed and positive for right-handed ones. We constructed right-handed DP24 and DP12 helices taking the chains to have the experimental geometry of left-handed V_h-amylose[22] and inverting the sign of the 141′4′ torsions. Then, we determined the best position of the *B*6 (nonbackbone) beads by performing an annealing molecular dynamics between 500 and 250 K with the backbone positions fixed. The values of 641′6′ adopted for fixed right-handed backbone, ∼50°, are consistent with the accessible states shown in Figure 11. These left- and right-handed DP12 models were mapped back to the atomistic structures. A simple minimization of the atomistic reconstructed structures renders an energy difference of 17.6 kcal/mol of monomer, favorable to the left-handed one. An M3B minimization was performed also in the coarse grain left- and right-handed DP12 molecules. The coarse grain results indicate that the left-handed helix is 12.8 kcal/monomer more stable. We note that the relaxation of the right-handed DP12 yielded a distorted structure in which many of the 141′4′ torsion angles had turned negative (i.e., locally left-handed) to produce a more stable structure. It may be expected that the method followed for the construction of the bead right-handed helix did not produce the optimum right-handed structure. We have found a right-handed structure that has a better hydrogen bond pattern and is 9.8 kcal/mol per monomer more stable than the reconstructed-minimized right-handed DP12. (This right-handed structure was optimized with our atomistic force field from the coordinates of a local MM3 minimum provided by an anonymous reviewer.) The energy difference between the optimized left-handed and right-handed DP12 thus decreases to 7.8 kcal per mole of glucose residue. We are not attempting in this work to perform an exhaustive search of global minima for the left- and right-handed DP12, and thus the definitive energy difference between the best left- and right-handed helices may be slightly different from that presented here. However, it should be emphasized that the M3B model correctly predicts the relative stability of the left-handed structure and provides an energy difference comparable to the atomistic models.

TABLE 8: Cell Parameters for the V_h-Amylose Crystal: Comparison of Experimental, Atomistic, and Coarse Grain Results

V_h-amylose model	ρ (g/cm^3)	*a* (Å)	*b* (Å)	*c* (Å)	α	*β*	*γ*
exptl[a]	1.426	27.3	27.3	8.05	90.0	90.0	120.0
M3B	1.435	26.23	26.28	8.54	91.0	89.5	118.7
atomistic	1.448	26.62	27.55	8.25	90.7	87.3	122.1
reconst-A[b]	1.456	26.86	25.93	8.55	89.6	89.5	121.5
M3B/exptl	1.006	0.96	0.96	1.06	1.01	0.99	0.99
atomistic/exptl	1.015	0.98	1.01	1.03	1.01	0.97	1.02
reconst-A/atomistic	1.006	1.00	0.94	1.04	0.99	1.03	1.00

[a] Experimental data from ref 22. [b] Reconst-A is the atomistic structure reconstructed from the M3B-optimized structure and further minimized with the atomistic force field.

The atomistic and coarse grain results are in agreement with the experimental lack of evidence of right-handed helices[24] and are also consistent with the potential of mean force of the 141′4′ torsional angle in the M3B and atomistic model (Figure 11). It is worth noting that although the results qualitatively agree for the M3B and atomistic right-handed DP12 relative stability, the coarse grain model does not display the variety of local minima that exist for the atomistic model: the coarse grain potential energy surface is smoother.

3.2.2. V-Amylose Crystalline Structure. Amylose can crystallize as left-handed compact single helices with water (and in some cases other small molecules) trapped inside the helical cavity. These structures are generally referred as V-amylose. Even the "anhydrous" V_a-amylose structure has roughly one water molecule per monomer in the channels formed by the helices.[23] The hydrated form, V_h-amylose,[22] has water molecules inside the helix and between them, making a total of $1^1/_3$ water molecules per glucose monomer.

We have compared the experimental X-ray crystal structure of V_h-amylose[22] with the optimized structure obtained by the atomistic and M3B model starting in both cases from the experimental configuration. The chains crystallize in a hexagonal lattice with the parameters indicated in Table 8 (see the "experimental" row) and space group $P6_522$.[22] The X-ray spectra[22] are consistent with the left-handed 6-fold helices packed with statistically random up and down chain disorder. We constructed a minimum cell with four helices using the reported symmetry and atomic positions, and assuming that two up and two down periodic chains (the diagonal pairs in the unit cell) represent the disorder of the experimental structure. Each cell consisted of 24 glucose residues and 32 water molecules. The initial M3B structure was constructed from this atomistic cell, following the rules described in Section 2.1. Particles and cell coordinates were relaxed to minimize the energy in a stepwise fashion (first, we relax the water positions, with other molecules and cell fixed; second, we relax all atoms with fixed cell, and finally free all the parameters). The densities and cell parameters that result from this minimization with the atomistic and M3B model are summarized in Table 8.

The densities predicted by the atomistic and M3B simulations are in excellent agreement with experiment. For the cell parameters, the atomistic rendered cell parameters are between 2% and 3% of the experimental values, whereas the differences between M3B and the experimental cell parameters are within 1% and 6%. Table 9 summarizes the rms displacement between the different models.

Following the mapping procedure described in Section 2.2, a fully atomistic model was *reconstructed* from the M3B-

TABLE 9: Root Mean Square Fit between V_h-Amylose Structures[a]

compared structures	total rms (Å)	helix rms (Å)	rigid-body rms (Å)
M3B/exptl	2.32	0.80 ± 0.20	1.28
atom./exptl	1.39	0.43 ± 0.03	0.46
reconst-A/exptl	2.10	0.81 ± 0.22	
reconst-A/atom	1.93	0.78 ± 0.20	

[a] Exptl: experimental X-ray structure from ref 22. M3B: coarse grain minimized from experimental structure. Atom.: atomistic minimized from experimental structure. Reconst-A: atomistic reconstructed from M3B structure and further minimized with atomistic force field. See the text for explanation of the different rms quantities.

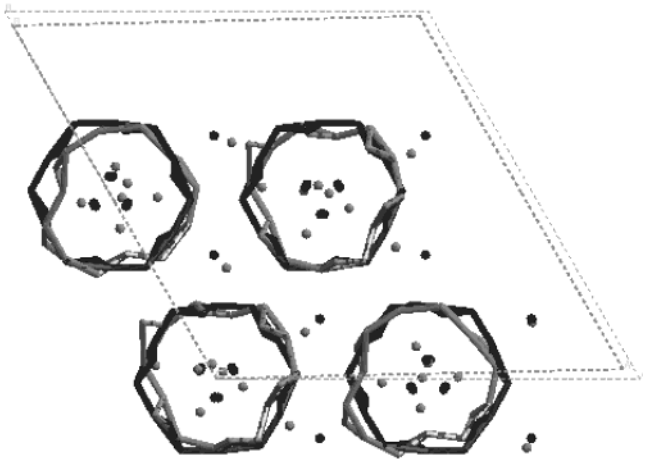


Figure 14. Experimental (black) and M3B optimized (gray) structures of V_h-amylose. Only C1, C4, and C6 of the chains and O of the water are shown for the experimental structure, to make it comparable to the M3B model.

minimized structure of V_h-amylose. We named this model "reconstructed-M3B". The structure that resulted from the energy minimization of the atomistic reconstructed-M3B with the atomistic force field was named "reconstructed-A". The density and cell parameters of reconstructed-A are indicated in Table 8. The energy of reconstructed-A is less than 1 kcal/(mol monomer) higher than the more symmetric atomistic structure minimized directly from the X-ray structure. Table 9 summarizes the rms displacement between different pairs of structures and shows the total and per helix rms of the pairs considered. Whereas "total rms" takes into account the relative displacement of the helices and water molecules, "helix rms" reflects solely the distortion of each helical molecule. In both cases, the rms fit of the atomistic description is approximately half the value of that of the coarse grain description. The low "helix rms" values indicate that for both the atomistic and coarse grain descriptions each individual helix preserves the ordered structure proposed from the X-ray data. Figure 14 shows the M3B-minimized structure and the M3B-mapped experimental cell of V_h-amylose. The main sources of mismatching between M3B and the X-ray structure are the following:

(i) The nondirectional interactions of the *W* water bead molecules. The bead model would not be able to retain the helical disposition that the atomistic water molecules presumably form via directional hydrogen bonds inside the amylose channels.

(ii) The intrinsic interactions between the helices in the coarse grain model leads to a rotation by 12° from the preferred rotation in the atomistic model. When the water beads are kept fixed in the helical configuration and each amylose helix is allowed to move only as a rigid body, we find that the M3B model favors a clockwise rotation by ~12° of the up helices around its axis and a counterclockwise rotation by ~12° of the down helices. This rotation alone is responsible for an rms difference of 1.28 Å of the difference in Table 9. The rigid-body rotation contribution is only 0.46, suggesting that the rotation of the helices in the M3B model may also be attributed to by a lack of directional forces.

(iii) The coarse grain backbone angles in the experimental structure, $141' = 170$ and $414' = 130$ depart from the equilibrium coarse grain angles in the M3B potential ($141' = 145$ and $414' = 121$). The angles of Table 3 are stiffer than the ones obtained from gas-phase simulations of DP2 (Table 2Sa) and DP4 (Table 2Sb) in order to correctly reproduce the molecular compressibility in a wide range of pressures (see the lower panel of Figure 6). The helices of V-amylose distort as the coarse grain angles try to reach their equilibrium values. The use of the gas-phase parameters of DP4 (Table 2S) for the angles renders a much improved structure for V-amylose: the agreement between the minimized M3B structure using the soft-angle parameters of Table 2Sb and the experimental one is excellent; the ratio of M3B-soft/experimental cell parameters is 1.0048, 1.0144, and 1.0007 for *a*, *b*, and *c*, 0.9988, 0.9920, and 1.0167 for α, β, and γ. The soft-angle M3B density is 100.01% of the experimental one and the rms difference between single chains of the structure is just 0.56 Å. The significant difference between the parameters obtained from vacuum simulations of DP4 and from compression of the amorphous bulk cells is that the vacuum-parametrized force constants are much lower. The results show that the angle flexibility has an important impact in (a) the breadth of accessible configurations and (b) the compressibility of the molecule. The selection of the stiffness of the coarse grain bonds and angles involves a compromise between these two. In the standard parametrization of M3B (angle parameters from Table 3) we have privileged the molecular compressibility over the conformational space. In what follows we continue to use the standard M3B parameters from Table 3.

The results obtained with M3B for V_h-amylose are surprisingly good considering the degree of coarsening and the lack of directional forces so important in hydrogen bonds and usually invoked to explain the helical structure of amylose. The reconstruction of atomistic models from M3B yields good structures for subsequent atomistic simulations, stressing the suitability of M3B for combined mesoscale/atomistic simulations.

3.2.3. Double Helices. Besides the V single-helical structures, amylose has been found experimentally to crystallize as left-handed parallel double helices.[21] Malto-oligosaccharides longer than DP10 were also found experimentally to crystallize as duplexes.[34] To examine such issues, we constructed parallel and antiparallel left-handed double-helical structures of DP12 starting from an extended configuration of the DP12 strand, a 6-folded left-handed helix with $n = 6$ and $h = 3.73$ Å mapped into the M3B model from the crystallographic structure of KOH-amylose.[35] A second DP12 strand with the same configuration was added aligned parallel or antiparallel to the first one. The best relative configuration of the helices was found by performing rigid-body optimization of each helix. The rigid-molecule systems were embedded in a periodic cell, and the cell lengths and angles were relaxed. Next, the rigidity was removed and the bead positions and cell parameters were optimized together. The resultant structure was the starting configuration for 1.2 ns NPT dynamics at $T = 300$ K and $p = 1$ atm. We found that the double-helical nature of the structure did not change during the

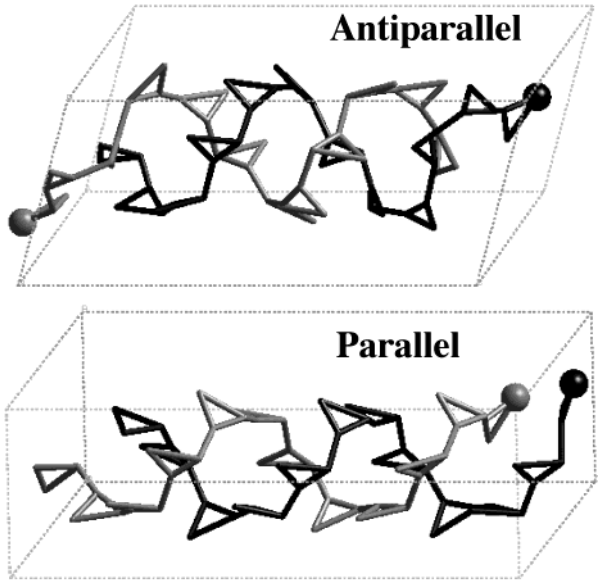


Figure 15. Antiparallel and parallel double-helical structures of DP12 after 1.2 ns NPT simulations at $p = 1$ atm and $T = 300$ K with the M3B model. The *B*1 end of each strand is indicated with a ball.

TABLE 10: Average Values of Density, Relative Energy (per Monomer), and Helical Pitch (*p*) of M3B Bulk DP12 Double-Helical Structures

configuration	ρ (g/cm^3)	E (kcal/mol)	p (Å)	n
antiparallel[a]	1.448 ± 0.007	0	16.7 ± 0.3	7.0 ± 0.2
parallel[a]	1.451 ± 0.005	0.38 ± 0.34	17.9 ± 0.2	7.0 ± 0.2

[a] Values averaged over the last 0.5 ns of the 1.2 ns NPT simulations of coarse grain DP12 at 300 K and $p = 1$ atm.

simulations, as can be seen in Figure 15. The average helix and cell parameters, density, and relative energy of the two M3B duplexes are summarized in Table 10. The energies of the parallel and antiparallel structures were indistinguishable within the uncertainty of the simulations. Both M3B duplexes are unstable in vacuum simulations, because of the unfavorable angle energy contribution required for the extended configurations in the M3B model. Vacuum atomistic simulations of DP12 parallel and antiparallel double helices were also performed, reconstructing the atomistic positions from the rigid-body optimized M3B models and performing 50 ps NVT molecular dynamics at 250 K. Contrary to the results of vacuum simulations of the M3B systems, both the parallel and antiparallel duplexes were stable in the condensed phase during the simulation time. This result confirms the importance of the multiple minima of the atomistic potential energy surface to stabilize the variety of helical structures observed for amylose and its fragments. As was observed in the M3B simulations, atomistic simulations of parallel and antiparallel configurations lead to energies (averaged over the last 25 ps of the simulations) that are indistinguishable. A negligible difference in energy was also found by other authors[36] using a search of minimum energy parallel and antiparallel configurations of amylose duplexes via optimization of the helical parameters with an atomistic model. However, the most stable parallel and antiparallel structures found in ref 36 were straight helices, whereas our simulations lead to an antiparallel atomistic DP12 duplex that bends (Figure 16) to form a "wavy" structure. The same kind of "wavy" structure was assigned from X-ray experiments on the only antiparallel structure known for malto-oligosaccharides.[25,26] The

Figure 16. Antiparallel atomistic DP12 duplex after 50 ps NVT dynamics at 250 K.

TABLE 11: Comparison of Experimental and M3B Glass-Transition Temperatures

water (wt %)	T_g^{M3B} (K)	T_g^{Exptl} (K)[a]	DP1 molecules	W molecules	ramp (K/ns)	$T_{initial}$ (K)	T_{final} (K)
0	296 ± 25	304	32	0	20	150	350
12.2	239 ± 25	240	90	125	200	150	325

[a] From ref 38. The last five columns indicate the ramp details for the T_g^{M3B} determination procedure.

M3B model did not reproduce this wavy feature of the antiparallel duplexes.

3.3. Glass Transition of Glucose and Glucose−Water Systems. Most pure carbohydrates and their mixtures with low water content form glasses when rapidly cooled. This is the case for glucose and other malto-oligosaccharides.[3] A transition between a metastable glass and a rubbery liquid occurs at the glass transition temperature, T_g, accompanied by an abrupt change in the thermal expansion coefficient. We have used the M3B model to compute the glass-transition temperature as the inflection point of V versus $1/T$ for two systems:

(1) Pure amorphous glucose.

(2) Glucose with 12.2% w/w water.

This method was also applied by other authors to estimating T_g of DP10 malto-oligosaccharide from atomistic molecular dynamics simulations.[37]

The curve V versus $1/T$ was constructed from molecular dynamics simulations of the amorphous systems in which temperature was increased at a uniform rate in steps of 25 K to convert the system from the glassy state to the rubbery state. The systems were prepared with the amorphous builder module of Cerius2,[12] followed by 5 ns NPT simulation at 500 K. They were then cooled to 200 K in steps of 50 K (0.5 ns at each temperature), equilibrated at 150 K for 10 ns, and then heated with a temperature ramp. The composition of the cells and details of the ramps are indicated in Table 11. Prior to the analysis of V versus $1/T$, the volume and temperature of the heating simulations were dynamically averaged (every 50 ps for pure glucose and every 5 ps for the mixture) to decrease the noise. The resultant V versus $1/T$ curves are plotted in Figure 17, along with the linear fit for the two portions of low- and high-temperature data. The estimated uncertainty in T_g is ~25 K. Using the M3B models, we obtained (see Table 11)

(1) $T_g = 296$ K for 0% water, in excellent agreement with the value obtained by differential scanning calorimetry experiments[38] of 304 K.

(2) $T_g = 239$ K for 12.2% water, in excellent agreement with the value obtained by differential scanning calorimetry experiments[38] of 240 K.

The accuracy of these results constitutes a further validation of the M3B coarse grain force field, suggesting that M3B is suitable for the study of supercooled and glassy carbohydrate systems. The dynamics of such systems would be extremely expensive to study with atomistic simulations.

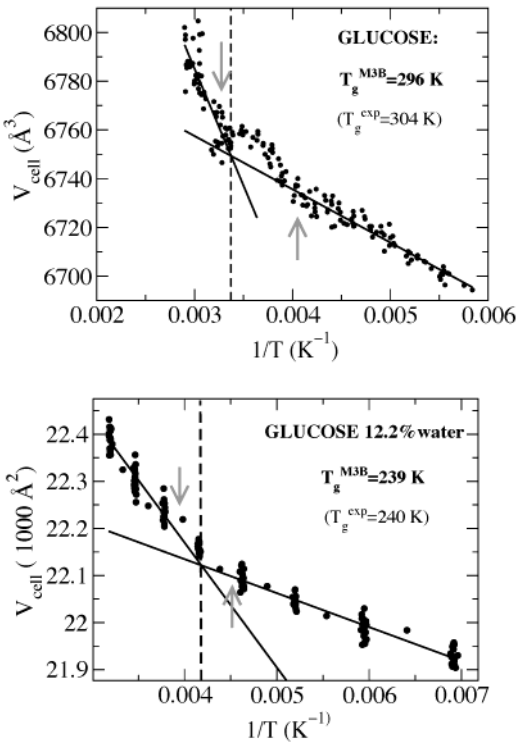


Figure 17. Determination of T_g from the discontinuity in thermal expansiveness for glucose and 12 wt % glucose in water with the M3B model. The lines correspond to linear fits of the rubbery and glassy states (using the data points to the left and right of the arrows, respectively).

4. Summary and Conclusions

In this paper we derive from atomistic simulations the M3B coarse grain model for α(1→4) D-glucans. This is the first coarse grain model derived for carbohydrates. M3B allows reliable simulations with time steps of 10 fs (10 times that for fully flexible atomistic simulations of sugars), decreases the number of particles by a factor of 8, while avoiding costly Ewald sums associated with coulomb interactions. The total gain in speed is about 7000 times for the same CPU resources and integration method. Thus, M3B makes simulations on single workstations for microseconds of simulation accessible for glucan systems with a composition equivalent to thousands of atoms.

The M3B model represents each glucose monomer by three beads while retaining enough detail to distinguish between different degrees of polymerization and the chemical coordination of the monomer. In the derivation of M3B, we paid special attention to reproducing the shape and nonbonded interactions of the atomistic molecules for a wide range of pressures, a task usually neglected in coarse grain parametrization. Our results show that a parametrization of the valence terms based exclusively on gas-phase simulations leads to extremely soft modes responsible for unphysical deformations of the molecules when the bulk system is compressed or expanded. On the other hand, we have shown that the stiffer angle force constants obtained to reproduce the molecular compressibility at high pressures restrain the configurational space available to the molecule. The price to pay for the simplification of the complex interactions of many atoms into a few coarse grain particles is that the parametrization is state dependent, and a single set of parameters may not suit as well all the purposes of the coarse grain model. We have used throughout this paper the stiffer coarse grain angle constants that were optimized to mimic the compressibility in a wide range of pressures. Although the angle parameters obtained from atomistic molecular dynamics simulations of a single DP4 (Table 2Sb of the Supplementary Information) render a wrong compressibility, they may be preferable for the study of the structure of isolated glucans or bulk structures at zero pressure.

An important difference between M3B and the usual coarse grain models for polymers is the presence of carefully parametrized torsion angle terms between the beads, which are essential for the formation of helical structures. M3B maps the essential modes characterizing the chain dynamics into coarse grain angles and torsions. The helical structures can be described in the configurational space formed by the angles and torsions of the M3B model to provide a Ramachandran plot[9,39] in terms of the rise *h* and number of monomers per turn *n* of the helices to the values of the atomistic glycosidic bonds.[39] The representation in terms of *n* and *h* is more straightforward to visualize than the standard Ramachandran plot that requires the knowledge of the details of the monomer configuration. However, the exact relationship between the helix parameters, *n* and *h*, and the coarse grain angles and torsions for the M3B models also depends on the monomer configuration (i.e., bond lengths). If the coarse grain bonds and angles are specified, the "bead-Ramachandran" representation in the [141′4,641′6′] space can be mapped into the [*h*,*n*] space and used as a tool to predict stable helical configurations of polysaccharides, in an analogous way to the atomistic study of the conformation of these molecules.[39]

The development of coarse grain models able to produce stable helical structures has become of increasing interest in recent years for use in simulations of protein folding. M3B is able to form helical structures with simple potentials and spherically symmetric nonbonded interactions. We had developed[40] a one bead per monomer (M1B) model with the same energy expression as M3B that also produced stable helices. However the M1B simplified model cannot distinguish left- and right-handed helices. The M3B model for oligosaccharides does distinguish right- and left-handed helices, and we have shown that M3B predicts the left-handed helices of α(1→4) D-glucans to be more stable, in agreement with atomistic simulations and the interpretation of X-ray spectra. We found that M3B left-handed single helices are stable at room temperature for at least several dozens of nanoseconds. Moreover, M3B is able to describe many of the multiple helical structures of amylose and malto-oligosaccharides: parallel and antiparallel left-handed double-helical structures for oligomers are predicted to be stabilized in the condensed phase and their relative energies are predicted to be indistinguishable, in agreement with detailed atomistic conformational studies.[36] Although most sugar and sugar−water mixtures properties are usually described in terms of specific and highly directional hydrogen bonds interactions, we have shown with M3B that is possible to construct a coarse grain model of these systems without the *explicit* introduction of either hydrogen bonds or directional nonbonded forces between particles. This *does not* mean that hydrogen bonds and electrostatic forces are irrelevant in the atomistic interactions that determine the configuration of glucans but that their overall effect in the molecular packing and structure can be reproduced with the mean field approach presented in this work. We consider that a high-quality description of the nonbonded interactions is the key for a faithful representation of the molecular packing, and that the inclusion of atomistically parametrized three- and four-body valence terms (angles and torsions) contribute to the stabilization of a variety of local structures arising from specific hydrogen bond interactions in

the atomistic structure. M3B succeeds in describing the main configurations of the amylose chains because it uses a realistic potential energy surface (PES) for the relative movement of the monomers. Nevertheless, it should be noted that the coarse grain PES is by construction smoother than the atomistic PES, and thus will not display the variety of local minima of the latter.

The coarse grain model presented for the glucans can be combined with its companion single bead *W* model for water (also developed here) to simulate concentrated sugar mixtures of arbitrary complex composition, such as those arising from the hydrolysis of starch. We have shown that M3B performs quite well in predicting the glass transition temperature, the single most relevant property of water−carbohydrate mixtures. The predicted T_g's for both glucose and a concentrated glucose−water mixture agree essentially exactly with the experimental values. Note that this property was *not* considered in the development of the coarse grain force field. In a recent study of glucose mixtures with 8−20 wt % water content,[41] we compared the distribution of water molecules in the mixture using the atomistic and M3B model. We found the distribution of water−water connectivity and the percolation threshold for water to be identical (within the uncertainty of the simulations) for the atomistic and coarse grain models. The distribution of water was very well reproduced despite the lack of hydrogen bonds, but this should not be so surprising if we consider that the hydrogen bonds are ubiquitous in water−sugar mixtures, and thus are not specific any more and can be treated with a mean field effective spherical potential.

This work presents a complete parametrization of M3B for linear $\alpha(1\rightarrow4)$ glucans. The same model can be parametrized for $\alpha(1\rightarrow6)$ glucans to extend the use of M3B to simulation of branched amylopectin, glycogen, and dextran. In the same way the parametrization of M3B model can be extended for $\beta(1\rightarrow4)$ glucans to perform coarse grain simulations of cellulose. We outline the required steps of these parametrizations in Section 2.3.4. The model and parametrization presented in this work is a precedent for such extensions. The accurate reconstruction of the atomic coordinates from M3B models opens the possibility to perform multiscale simulations in which the coarse grain model is used for the equilibration of the system, while allowing the full atomistic detail to be recovered for the analysis of atomistic properties. The present model for glucans should be invaluable in unraveling the structure and dynamics of water−glucan systems, in particular in conditions of low temperature and low water content, where the extremely long relaxation times discourage atomistic simulation studies.

Acknowledgment. We acknowledge Tahir Cagin, Brian Guthrie, Michael Tanoff, and Gary A. Day for their comments and in particular we thank Alejandro Strachan for his enthusiastic help with the MCSA optimization procedure. We thank referee 1 for his/her spectacular effort in checking our results, finding a numerical error, and for kindly providing his/her own atomistic structures of the right- and left-handed DP12. The facilities of the Materials and Process Simulation Center are supported by ONR-DURIP, ARO-DURIP, SIR-IBM, NSF (CHE), and additional support is provided by DOE-ASCI, DOE-FETL, ARO-MURI, ONR-MURI, NIH, NSF, Kellogg's, General Motors, ChevronTexaco, Seiko-Epson, Beckman Institute and Asahi Kasei.

Supporting Information Available: Detailed derivation procedure of the coarse grain force field. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Nowakowski, C. M.; Hartel, R. W. *J. Food Sci.* **2002**, *67*, 1419.

(2) Marchal, L. M.; Beeftink, H. H.; Tramper, J. *Trends Food Sci. Techno.l* **1999**, *10*, 345.

(3) Slade, L.; Levine, H. *Crit. Rev. Food Sci. Nutr.* **1991**, *30*, 115.

(4) Tromp, R. H.; Parker, R.; Ring, S. G. *Carbohydr. Res.* **1997**, *303*, 199.

(5) van den Dries, I. J.; van Dusschoten, D.; Hemminga, M. A. *J. Phys. Chem. B* **1998**, *102*, 10483.

(6) Kremer, K. *Macromol. Chem. Phys.* **2003**, *204*, 257.

(7) Muller-Plathe, F. *ChemPhysChem* **2002**, *3*, 754.

(8) Baschnagel, J.; Binder, K.; Doruker, P.; Gusev, A. A.; Hahn, O.; Kremer, K.; Mattice, W. L.; Muller-Plathe, F.; Murat, M.; Paul, W.; Santos, S.; Suter, U. W.; Tries, V. Bridging the Gap between Atomistic and Coarse-Grained Models of Polymers: Status and Perspectives. In *Viscoelasticity, Atomistic Models, Statistical Chemistry*; Advances in Polymer Science; Springer-Verlag: Heidelberg, 2000; Vol. 152, p 41.

(9) Rao, V. S. R. *Conformation of Carbohydrates*; Harwood Academic Publishers: Amsterdam, The Netherlands, 1998.

(10) Takusagawa, F.; Jacobson, R. A. *Acta Crystallogr., Sect. B* **1978**, *34*, 213.

(11) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, U.K. and New York, 1989.

(12) *Cerius2*, version 4.0; Accelerys: San Diego, CA, 1999.

(13) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897.

(14) Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.

(15) Goddard, W. A.; Cagin, T.; Blanco, M.; Vaidehi, N.; Dasgupta, S.; Floriano, W.; Belmares, M.; Kua, J.; Zamanakos, G.; Kashihara, S.; Iotov, M.; Gao, G. H. *Comput. Theor. Polym. Sci.* **2001**, *11*, 329.

(16) Parks, G. S.; Huffmann, H. H.; Cattoir, F., *J. Phys. Chem.* **1928**, *32*, 1366.

(17) Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548.

(18) *CRC Handbook of Chemistry and Physics*, 81st ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 2000−2001.

(19) *CODATA Key Values for Thermodynamics*; Cox, J. D., Wagman, D., Medeyev, V., Eds.; Hemisphere Publishing Corp.: New York, 1984.

(20) Akkermans, R. L. C.; Briels, W. J. *J. Chem. Phys.* **2000**, *113*, 6409.

(21) Imberty, A.; Chanzy, H.; Perez, S.; Buleon, A.; Tran, V. *J. Mol. Biol.* **1988**, *201*, 365.

(22) Brisson, J.; Chanzy, H.; Winter, W. T. *Int. J. Biol. Macromol.* **1991**, *13*, 31.

(23) Winter, W. T.; Sarko, A. *Biopolymers* **1974**, *13*, 1447.

(24) Muller, J. J.; Gernat, C.; Schulz, W.; Muller, E. C.; Vorwerg, W.; Damaschun, G. *Biopolymers* **1995**, *35*, 271.

(25) Hinrichs, W.; Buttner, G.; Steifa, M.; Betzel, C.; Zabel, V.; Pfannemuller, B.; Saenger, W. *Science* **1987**, *238*, 205.

(26) Hinrichs, W.; Saenger, W. *J. Am. Chem. Soc.* **1990**, *112*, 2789.

(27) Shimada, J.; Kaneko, H.; Takada, T.; Kitamura, S.; Kajiwara, K. *J. Phys. Chem. B* **2000**, *104*, 2136.

(28) Carriere, C. J.; Bagley, E. B. *J. Rheol.* **1999**, *43*, 753.

(29) Simpson, T. D.; Dintzis, F. R.; Taylor, N. W. *Biopolymers* **1972**, *11*, 2591.

(30) Sarko, A.; Marchessault, R. H. *J. Am. Chem. Soc.* **1967**, *89*, 6454.

(31) Hinkle, M. E.; Zobel, H. F. *Biopolymers* **1968**, *6*, 1119.

(32) French, A. D.; Zobel, H. F. *Biopolymers* **1967**, *5*, 457.

(33) Senti, F. R.; Witnauer, L. P. *J. Polym. Sci.* **1952**, *9*, 115.

(34) Gidley, M. J.; Bulpin, P. V. *Carbohydr. Res.* **1987**, *161*, 291.

(35) Sarko, A.; Biloski, A. *Carbohydr. Res.* **1980**, *79*, 11.

(36) Schulz, W.; Sklenar, H.; Hinrichs, W.; Saenger, W. *Biopolymers* **1993**, *33*, 363.

(37) Momany, F. A.; Willett, J. L. *Biopolymers* **2002**, *63*, 99.

(38) Roos, Y. *Carbohydr. Res.* **1993**, *238*, 39.

(39) French, A. D. Conformational Accessibility of Polyglucosides. In *Polymers for Fibers and Elastomers*; Arthur, J. C., Ed.; American Chemical Society: Washington D. C., 1983; Vol. 260, p 43.

(40) Molinero, V.; Goddard, W. A., III. To be submitted for publication.

(41) Molinero, V.; Cagin, T.; Goddard, W. A., III. *J. Phys. Chem.*, submitted for publication