JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

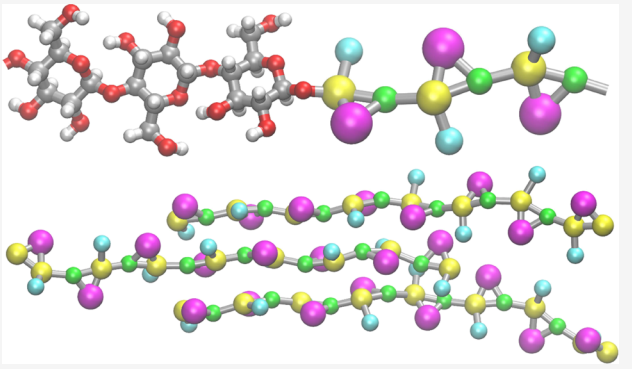
Extending the Martini 3 Coarse-Grained Force Field to Carbohydrates

Valery Lutsyk, Pawel Wolski, and Wojciech Plazinski*

Cite This: J. Chem. Theory Comput. 2022, 18, 5089−5107 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Carbohydrates play an essential role in a large number of chemical and biochemical processes. High structural diversity and conformational heterogeneity make it problematic to link their measurable properties to molecular features. Molecular dynamics simulations carried out at the level of classical force fields are routinely applied to study the complex processes occurring in carbohydrate-containing systems, while the usefulness of such simulations relies on the accuracy of the underlying theoretical model. In this article, we present the coarse-grained force field dedicated to glucopyranose-based carbohydrates and compatible with the recent version of the Martini force field (v. 3.0). The parameterization was based on optimizing bonded and nonbonded parameters with a reference to the all-atom simulation results and the experimental data. Application of the newly developed coarse-grained carbohydrate model to oligosaccharides curdlan and cellulose displays spontaneous formation of aggregates of experimentally identified features. In contact with other biomolecules, the model is capable of recovering the protective effect of glucose monosaccharides on a lipid bilayer and correctly identifying the binding pockets in carbohydrate-binding proteins. The features of the newly proposed model make it an excellent candidate for further extensions, aimed at modeling more complex, functionalized, and biologically relevant carbohydrates.



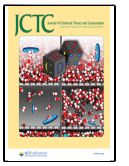
1. INTRODUCTION

Carbohydrates (saccharides) play an essential and widely recognized role in numerous chemical,[1,2] biochemical,[3,4] and technological processes[5,6] as well as display potential for the design of new materials.[7−9] A series of carbohydrate properties (chemical heterogeneity, variable glycosidic linkage types, variable functionalization patterns, nonuniform chain length, and high conformational heterogeneity even at the monomer level) make studying them at the molecular level problematic.[10,11] As an alternative to experimental approaches, the theoretical methods of molecular simulations can be proposed in order to quantitatively investigate the carbohydrate-containing systems.

Molecular dynamics (MD) simulations are a commonly applied computational tool capable of providing the connection between the molecular-level structure of the given system and its measurable properties. The usefulness of the MD simulations relies primarily on the accuracy of the underlying potentials of interactions assigned to describe the system (force fields). Numerous atomistic (all-atom, AA) and united-atom force fields have been parametrized for either unfunctionalized or functionalized carbohydrates and have been used to provide conformational, structural, and thermodynamic details related to saccharide properties.[12,13] Inherent to MD simulations, the system size often becomes a bottleneck for simulation efficiency. Thus, simulations of long polysaccharide chains or large, carbohydrate-containing systems may be very challenging when considering the atomistic level of resolution.

A convenient alternative to the AA-based simulations is the use of coarse-grained (CG) force fields.[14−17] Such choice significantly enhances the computational efficiency associated with simulations because of both reducing the system size and increasing the timestep parameter. As a consequence, large systems can be studied on a microsecond-long timescale and at reasonable computational cost. There exist several carbohydrate-dedicated CG models.[18−22] However, in most of the cases, a serious limitation is that the given model is restricted to the sole class of carbohydrates and there is no compatible set of models capable of describing other types of molecules and biomolecules that often coexist with carbohydrates in real systems. In 2009, the Martini-based CG model for

Received: May 25, 2022
Published: July 29, 2022



https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

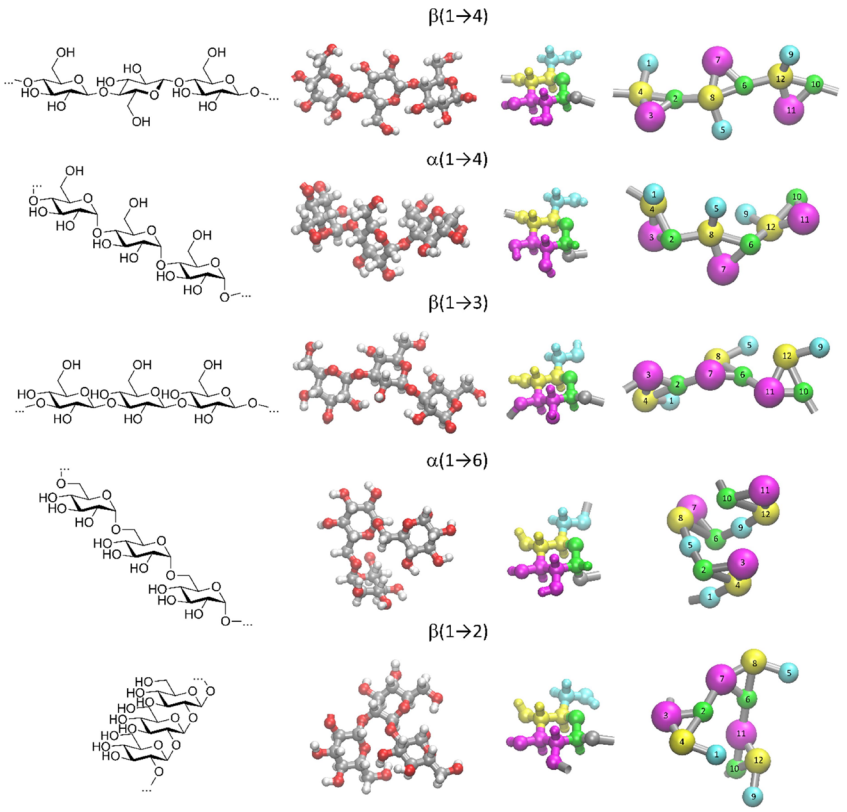


Figure 1. Chemical formulas of the considered glucose-based saccharides exploiting different linkage types, atomistic structures of the corresponding trimers, and the illustration of the mapping scheme on the example of a single residue. The shown bead numbering is used to define the force field terms (Tables 3−7). Analogous bead numbering is applied in the case of monosaccharides (Table 2).

unfunctionalized, glucopyranose-based carbohydrates has been proposed.[23] This model, its subsequent extensions[24−26] and modifications[27] are compatible with the earlier version of the Martini 2 force field, enabling simulations of complex systems, containing, for example, lipid bilayers and disaccharides,[28] covalently bonded glycans,[29] crystalline carbohydrates,[26] and many other carbohydrate-containing systems.[30,31] At the same time, several shortcomings of this model have been reported,[32] including, for example, overestimated self-aggregation properties.[27] Most of these shortcomings resulted from the underlying inaccuracies of the nonbonded interactions in the Martini 2-inherent, Lennard-Jones (LJ) parameters for certain bead types combined with the center-of-mass approach used for bead mapping. Recently, an updated version of the Martini force field has been developed,[33] differing substantially from the previous one and improving most of the spurious effects. However, the presently distributed force field parameters lack more complex carbohydrates and are limited to only two monosaccharide units (glucose and ribose). It is also worth noting that possible directions of extending the Martini 3.0 model toward the carbohydrates have been explored in a previous study;[34] however, none of the solutions proposed there has been included in the Martini distribution so far.

In this work, we present the extension of the recently developed Martini 3.0 CG force field, including the parameters for glucopyranose-based carbohydrates. The parameterized group comprised the carbohydrate di-, oligo-, and polysaccharides exploiting the most important glycosidic linkages (i.e., $\beta(1 \rightarrow 2)$, $\beta(1 \rightarrow 3)$, $\alpha(1 \rightarrow 4)$, $\beta(1 \rightarrow 4)$, and $\alpha(1 \rightarrow 6)$). The set of compounds considered by us is given in Figure 1. The procedure of parameterization relied on the use of atomistic data (MD simulations carried out at an AA level of accuracy) and experimental measurements (log P values, structural data). The final parameter set was tested by carrying out a series of CG MD simulations of different, glucose-based saccharides in water as well as in the presence of other biomolecules (lipid bilayers and proteins).

The article is organized as follows: The Theoretical Grounds section contains the accepted mapping procedure, the brief description of the force field functional form and the force field build-up rules. The Methods section includes the detailed description of the computational methodology used throughout the study and concerning either AA or CG simulations. The Model section, divided into several subsections with respect to considered aspects, reports the final CG parameters. The section entitled Properties of Studied Systems includes a series of simulation results that provide a validation for the newly proposed force field with respect to various carbohydrate-containing systems (including the comparison with the available experimental data and AA simulations). The discussion on the force field applicability and its possible limitations is given in the subsequent section. Finally, Conclusions summarize the main findings and provide concluding remarks.

2. THEORETICAL GROUNDS

2.1. General Remarks. The proposed CG model concerns the glucose-containing saccharides, in particular, glucopyranose monomers (as either α or β anomers), disaccharides containing the linkages of the $\beta(1 \rightarrow 2)$, $\beta(1 \rightarrow 3)$, $\alpha(1 \rightarrow 4)$, $\beta(1 \rightarrow 4)$, or $\alpha(1 \rightarrow 6)$ types, as well as oligo- and polysaccharides exploiting such linkages. The model relies on

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

the nonbonded parameters introduced for the Martini 3.0 force field,[33] which is developed by using the Martini-inherent strategies of parameterization, designed to be compatible with the Martini water model and was tested in combination with existing Martini models, describing other types of systems (e.g., lipid bilayers and proteins). Therefore, it can be considered as compatible with the Martini 3.0 family of CG force fields. In spite of that, the presently proposed approach substantially differs from the prospective carbohydrate-dedicated models discussed in the previous study.[34] In the following subsection, we provide a brief overview of the basic parametrization methodology, including procedures followed for the definition of the mapping scheme and parametrization of nonbonded and bonded interactions.

2.2. Mapping. According to the mapping scheme accepted in Martini 3.0, the three different CG beads can be used to represent two (tiny bead, T), three (S, small bead), or four (R, regular bead) heavy atoms. Therefore, at least three CG particles are required to model a single glucose monosaccharide. In contrast to the previous edition of the Martini carbohydrate-dedicated model, compatible with Martini 2,[23] we decided to use uniform, four-bead representation for a single glucose residue either as monosaccharide or being a part of a longer, oligo/polysaccharide chain. According to the proposed mapping, the carbohydrate residue is divided into four beads, representing the following molecular regions: (1) the hydroxymethyl group; (2) the ring oxygen, anomeric carbon and (in the case of reducing end or monosaccharide) anomeric hydroxyl group; (3) the vicinal diol or hydroxyethyl group (depending on the linkage type); and (4) the hydroxyethyl group. The graphical illustration of the accepted mapping scheme is given in Figure 1.

Although such scheme increases the number of beads per one unfunctionalized hexopyranose residue from 3 to 4 and, subsequently, slightly decreases the computational efficiency inherent to the model, it also has several advantages concerning both the present work and prospective model extensions:

1. Increasing the resolution of the model that represents the polysaccharide backbone which is now composed of either two ((1 → 2), (1 → 3), (1 → 4) linkages), or three ((1 → 6) linkages) bonds per residue in a chain. This increases the accuracy of the model with respect to structural rearrangements along the carbohydrate main chain. In particular, the rotation around the (1 → 6) linkages can now be properly modeled, which was impossible in the previous version of the carbohydrate-dedicated Martini 2 model.[23]
2. Reflecting the structural and topological features of carbohydrate polymers. This includes the symmetry of the carbohydrate residue and its basic geometrical features. For instance, the central part of the glucose ring is composed of the two nonbonded groups of atoms; therefore, it is represented by two separate beads.
3. Uniform mapping for either monosaccharides or residues in a chain. This feature facilitates perspective polysaccharide-related model extensions and its refinements based on the single-residue models.
4. Facilitating the procedure of prospective parameterization for functionalized polymers composed of pyranose residues. In the case of adding some ring-substituents, only minor refinements in the parameters concerning the closest neighbors of those substituents will be required. The most essential part of the model, that is, that describing the properties of glycosidic bonds, will remain nearly unaffected. This is analogous to the parameterization procedures developed for carbohydrates in some of the atomistic force fields.[35,36]

The mapping relied on the center-of-geometry (COG) approach introduced for Martini 3.0.[33] The COG approach includes all atoms contributing to a given CG bead, together with aliphatic hydrogens. In order to account for this, the AA MD simulations, performed according to the methodology described below, were selected as the source of the target data, concerning mainly the bonded CG parameters.

2.3. Bonded Interactions. The functional form of the potential-energy term, associated with the stretching of CG bonds and applied to all unique pairs of covalently linked beads, is given by eq 1:

$$V(b) = \frac{1}{4}k_b(b^2 - b_0^2)^2 \tag{1}$$

where b is the bond-length distance, b_0 is its reference value, and k_b is the corresponding force constant.

The functional form of the potential-energy term, associated with the bending of bond angles and applied to selected triplets of covalently linked beads, is given by eq 2:

$$V(\theta) = \frac{1}{2}k_\theta(\cos\theta - \cos\theta_0)^2 \tag{2}$$

where θ is the bond-angle value, θ_0 is its reference value, and k_θ is the corresponding force constant. In certain cases, especially when the bond angle becomes close to 180°, the special type of bond-angle bending potential (restricted bending potential)[37] is used in order to prevent the numerical errors and simulation instabilities:

$$V(\theta) = \frac{1}{2}k_\theta\frac{(\cos\theta - \cos\theta_0)^2}{\sin^2\theta} \tag{3}$$

where all variables are defined as in eq 2.

The functional form of the potential-energy term, associated with the deformation of improper-dihedral angles and applied to the subset of bead quadruplets in order to control out-of-plane distortions, is given by eq 4:

$$V(\xi) = \frac{1}{2}k_\xi(\xi - \xi_0)^2 \tag{4}$$

where ξ is the improper-dihedral angle value, ξ_0 is its reference value, and k_ξ is the corresponding force constant.

Finally, the functional form of the potential-energy term, associated with the torsion around dihedral angles, is given by eq 5:

$$V(\varphi) = k_\varphi[1 + \cos(m\varphi - \varphi_0)] \tag{5}$$

where φ is the dihedral angle value, m is the multiplicity of the term, φ_0 is the associated phase shift, and k_ϕ is the corresponding force constant. This term is applied to a subset of all possible dihedral angles as specified in Section 4 and Tables 34567.

Functions defined by eqs 1−5 are invoked by using the following types of interaction functions in GROMACS: eq 1: bonds, type 1; eq 2: angles, type 2; eq 3: angles, type 10; eq 4: dihedrals, type 2; eq 5: dihedrals, type 1.

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

2.4. Nonbonded Interactions. The considered nonbonded interactions are represented solely by LJ potentials because of electrical neutrality of all considered compounds and their building blocks. The nonbonded interactions are calculated as a sum over all interacting nonbonded pairs (i,j) using the following 12/6 interaction function with parameters C_{12} and C_6:

$$V(r_{ij}) = \sum_{\text{pairs},ij} \left(\frac{C_{12}}{r_{ij}^{12}} - \frac{C_6}{r_{ij}^{6}} \right) \tag{6}$$

where r_{ij} is the distance between interacting beads. The parameters $C_{12,ij}$ and $C_{6,ij}$ depend on the type of involved beads and those used in the present work were chosen among the bead types derived for Martini 3. Following the Martini convention, the first covalent neighbors are excluded from nonbonded interactions.

During selection of the optimal LJ parameters, we considered the chemical character of the mapped functional groups, the AA-derived solvent-accessible surface area (SASA) values, the experimental log P values measured for glucose monosaccharides, and the theoretically (KOWWIN and ALOGPS) predicted log P for selected di- and oligosaccharides.

3. METHODS

3.1. CG Simulations. The detailed list of the systems studied at the CG level is given in Table 1. The initial structures were either drawn manually or generated by using the hand-written *python3* program *carbo2martini3.py* (included in the Supporting Information). The *insane* tool was used to solvate the considered solute molecule and construct the initial configuration of the lipid bilayer wherever needed. The parameters concerning saccharides and constituting the currently proposed force field are given and discussed in detail in further sections. The remaining types of molecules (lipids, proteins) were modeled by using the Martini 3.0 parameters[33] and prepared by using the web-available tools (*martinize2* and *insane*). Simulations were carried out with the GROMACS 2016.4 package,[38] under periodic boundary conditions and in the isothermal−isobaric ensemble. The temperature was maintained close to the reference value (298, if not indicated otherwise) by applying the V-rescale thermostat,[39] whereas for the constant pressure (1 bar), the Parrinello-Rahman barostat[40] was used with a relaxation time of 40 ps. Either semi-isotropic (bilayer systems) or isotropic (remaining systems) pressure scaling was applied. The equations of motion were integrated with a time step of 10 ($\beta(1 \rightarrow 3)$-glucan- and $\beta(1 \rightarrow 2)$-glucan-containing systems), 30 (monosaccharides), or 20 fs (remaining systems) using the leap-frog scheme.[41] The translational center-of-mass motion was removed every timestep separately for the solute and the solvent. The Martini 3 water model[33] was applied. The van der Waals interactions (LJ potentials) are shifted to zero beyond the cutoff distance, that is, 1.1 nm. For Coulomb interactions, the reaction-field approach was used with a cutoff of 1.1 nm and $\varepsilon_r = 15$. Details of the MD parameters were kept according to the sample *mdp* files deposited at the *cgmartini.nl* website. Production simulations were carried out for a duration of 100−10,000 ns (depending on the system), and the data were saved to trajectory every 10−50 ps.

The log P values were calculated as the Gibbs free energy difference corresponding to the transfer of the saccharide molecule from water to *n*-octanol. The calculations concerned the following systems: monosaccharide of β-D-glucopyranose and $\alpha(1 \rightarrow 4)$-linked octamer of glucopyranose. In order to construct the thermodynamic cycle, the carbohydrate molecule was decoupled (removed) from both water and *n*-octanol solvents. Decoupling the carbohydrate molecule from either type of system was carried out by scaling down to zero all nonbonded interactions involving carbohydrate atoms in a stepwise manner as a function of a coupling parameter λ. The associated free energy changes were calculated with the Bennett acceptance ratio (BAR) method,[42] implemented in the GROMACS *gmx bar* subroutine, including the error estimation determined by using the default criteria. The 21 evenly spaced λ-points were accepted, and the data from equilibrated systems were collected every 1 ps for a duration of 100 (monomers) or 500 ns (octamers) in each λ window. A soft-core function was used for the van der Waals interactions to prevent energy singularities.

3.2. AA Simulations. The following saccharides were considered during AA simulations: (1) monosachcarides of α- and β-D-glucopyranose; (2) $\alpha(1 \rightarrow 4)$-linked octamers of glucopyranose; (3) $\beta(1 \rightarrow 4)$-linked octamers of glucopyranose; (4) $\beta(1 \rightarrow 3)$-linked octamers of glucopyranose; (5) $\beta(1 \rightarrow 2)$-linked octamers of glucopyranose; (6) $\alpha(1 \rightarrow 6)$-linked octamers of glucopyranose; (7) disaccharides linked by each of the above linkage types. Additionally, the data from our previous carbohydrate-oriented studies were occasionally used to validate the CG model. The CHARMM36[43,44] force field was used in all AA simulations. The initial structures of saccharides as well as the GROMACS-readable parameters were generated by the www.charmm-gui.org online server.[45,46] Simulations were carried out with the GROMACS 2016.4 package.[38] The saccharide molecules were placed in simulation boxes of dimensions dependent on the system type and surrounded by a number of explicit water molecules approximately accounting for the system density of 1 g/cm^3. The MD simulations were carried out under periodic boundary conditions and in the isothermal−isobaric ensemble. The temperature was maintained close to its reference value (298 K) by applying the V-rescale thermostat,[39] whereas for the constant pressure (1 bar, isotropic coordinate scaling), the Parrinello−Rahman barostat[40] was used with a relaxation time of 0.4 ps. The equations of motion were integrated with a time step of 2 fs using the leap-frog scheme.[41] The translational center-of-mass motion was removed every timestep separately for the solute and the solvent. The TIP3P model of explicit water[47] was applied and the full rigidity of the water molecules was enforced by the application of the SETTLE procedure.[48] The hydrogen-containing solute bond lengths were constrained by the application of the LINCS procedure with a relative geometric tolerance of 10^{-4}.[49] The electrostatic interactions were modeled by using the particle-mesh Ewald method[50] with cutoff set to 1.2 nm, while van der Waals interactions (LJ potentials) were switched off between 1.0 and 1.2 nm. Production simulations were carried out for a duration of 100 ns, and the data were saved to trajectory every 2 ps.

4. MODEL

4.1. Parameters. *4.1.1. Nonbonded Interactions.* The chemical character of glucopyranose monomer partially dictates the assignment of nonbonded parameters that should take into account the following factors: (1) the similar or the same polarity of beads representing the hydroxymethyl and

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

Table 1. CG Systems Considered in the Present Study[a]

purpose	compositon	no. of solvent molecules	box vector lengths [nm^3]	simulation time	remarks
conformational properties, SASA	α-Glc monomer	323	3.0 × 3.0 × 3.0 nm^3	5000 ns	unbiased MD
conformational properties, SASA, log P	β-Glc monomer	323	3.0 × 3.0 × 3.0 nm^3	5000 ns unbiased MD,20 ns per TI window	Unbiased MD, TI, various nonbonded parameters
log P	β-Glc monomer	450	4.8 × 4.8 × 4.8 nm^3	20 ns per TI window	TI, octanol solvent, various nonbonded parameters
solution density	100−2000 β-Glc monomers	1514−14,177	12 × 12 × 12 nm^3	10 ns	unbiased MD, various Glc content
lipid bilayer	DPPC lipid bilayer	11,645−11,663	10 × 10 × 18 nm^3	500 ns	unbiased MD,various temperatures
influence on the lipid bilayer	4200 β-Glc monomers + DPPC lipid bilayer	4242−4299	10 × 10 × 23.5 nm^3	500 ns	unbiased MD, various temperatures
log P	$\alpha(1 \rightarrow 4)$-linked Glc octamer	8100	9.9 × 9.9 × 9.9 nm^3	50 ns per TI window	TI, various nonbonded parameters
log P	$\alpha(1 \rightarrow 4)$-linked Glc octamer	5064	10.8 × 10.8 × 10.8 nm^3	50 ns per TI window	TI, octanol solvent, various nonbonded parameters
conformational properties, SASA	$\beta(1 \rightarrow 4)$-linked Glc dimer	3354	7 × 7 × 7 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\alpha(1 \rightarrow 4)$-linked Glc dimer	3353	7 × 7 × 7 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\beta(1 \rightarrow 2)$-linked Glc dimer	3348	7 × 7 × 7 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\beta(1 \rightarrow 3)$-linked Glc dimer	3350	7 × 7 × 7 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\alpha(1 \rightarrow 6)$-linked Glc dimer	3351	7 × 7 × 7 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\beta(1 \rightarrow 4)$-linked Glc tetramer	4871	8 × 8 × 8 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\alpha(1 \rightarrow 4)$-linked Glc tetramer	4872	8 × 8 × 8 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\beta(1 \rightarrow 2)$-linked Glc tetramer	4870	8 × 8 × 8 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\beta(1 \rightarrow 3)$-linked Glc tetramer	4874	8 × 8 × 8 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\alpha(1 \rightarrow 6)$-linked Glc tetramer	4874	8 × 8 × 8 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\beta(1 \rightarrow 4)$-linked Glc octamer	9187	10 × 10 × 10 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\alpha(1 \rightarrow 4)$-linked Glc octamer	9195	10 × 10 × 10 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\beta(1 \rightarrow 2)$-linked Glc octamer	9181	10 × 10 × 10 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\beta(1 \rightarrow 3)$-linked Glc octamer	9187	10 × 10 × 10 nm^3	500 ns	unbiased MD
conformational properties, SASA	$\alpha(1 \rightarrow 6)$-linked Glc octamer	9197	10 × 10 × 10 nm^3	500 ns	unbiased MD
aggregation properties	3 $\beta(1 \rightarrow 3)$-linked Glc chains (26 residues)	29,364	15 × 15 × 15 nm^3	200 ns	unbiased MD, initiated from triple-helix structure
aggregation properties	1 $\alpha(1 \rightarrow 4)$-linked Glc chain (32 residues)	132,476	25 × 25 × 25 nm^3	250 ns	unbiased MD, initiated from V-helix structure, various solvents: water and 1-bead solvents (C1 to C6)
aggregation properties	50 $\alpha(1 \rightarrow 4)$-linked Glc octamers	271,906	32 × 32 × 32 nm^3	1500 ns	unbiased MD

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

Table 1. continued

purpose	compositon	no. of solvent molecules	box vector lengths [nm^3]	simulation time	remarks
aggregation properties	100 $\beta(1 \rightarrow 4)$-linked Glc octamers	112,182	24 × 24 × 24 nm^3	4575 ns	unbiased MD
aggregation properties	10 $\beta(1 \rightarrow 3)$-linked Glc octamers	35,376	16 × 16 × 16 nm^3	10,000 ns	unbiased MD
aggregation properties	50 $\alpha(1 \rightarrow 4)$-linked Glc polymers (26 residues)	606,152	42 × 42 × 42 nm^3	170 ns	unbiased MD
aggregation properties	50 $\beta(1 \rightarrow 4)$-linked Glc polymers (26 residues)	606,248	42 × 42 × 42 nm^3	780 ns	unbiased MD
aggregation properties	40 $\beta(1 \rightarrow 3)$-linked Glc polymers (26 residues)	124,888	25 × 25 × 25 nm^3	650 ns	unbiased MD
interactions with proteins	1 protein (PDB: 1I8A) and + 1 monomer of β-Glc	8579 + 98 Na^+ + 91 Cl^-	10 × 10 × 10 nm^3	10,000 ns	unbiased MD
Interactions with proteins	1 protein (PDB: 4O7P) and + 25 maltose molecules	27,444 + 310 Na^+ + 294 Cl^-	15 × 15 × 15 nm^3	5000 ns	unbiased MD
Interactions with proteins	1 protein (PDB: 2OVW) and + 25 cellobiose molecules	27,750 + 305 Na^+ + 307 Cl^-	15 × 15 × 15 nm^3	10,000 ns	unbiased MD

[a]If not indicated otherwise, the solvent was Martini 3 water, for which 1 CG bead = 4 water molecules.

diol groups (beads of odd numbers; see Figure 1); (2) the similar or the same polarity of beads representing the central part of the monosaccharide or polysaccharide backbone (beads of even numbers; see Figure 1); (3) more polar character of odd beads in comparison to less polar even beads. Multiple combinations fulfilling these conditions have been tested. The results for selected combinations which provided the best agreement with the reference data are illustrated in Figure 2.

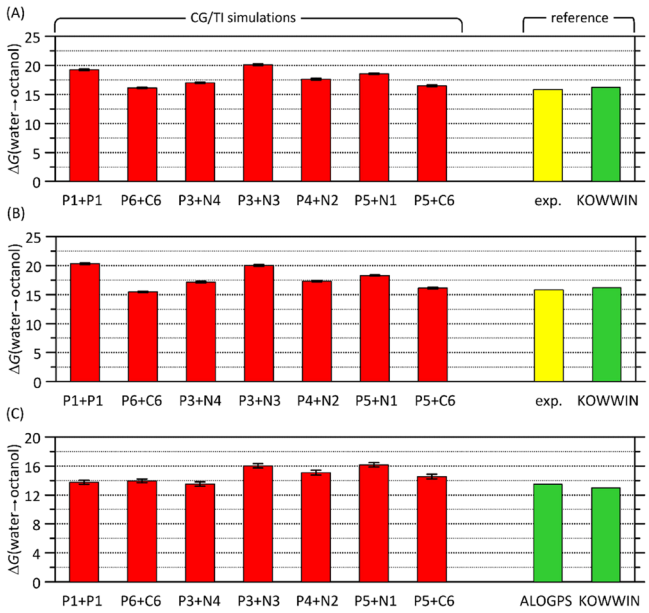


Figure 2. Results of the CG TI simulations aimed at determining the free energies associated with the transfer of glucose mono- and oligosaccharides from water to *n*-octanol (equivalent to log *P*). (A) Results for monosaccharides with rings composed of T, S, and R beads. (B) Results for monosaccharides with rings composed of T, S, and S beads. (C) Results for octamers with rings composed of T, S, and R beads (recalculated with respect to a single residue). Theoretical data are compared with predictions of KOWWIN (Estimation Programs Interface Suite for Microsoft Windows, v 4.11) and ALOGPS 2.1[51] programs as well as with the experimental data.[52]

A very similar level of agreement with the experimental data was achieved in each of the cases shown in Figure 2. Moreover, the trend of decreasing polarity of the glucose oligomer in comparison to the monomer was correctly reflected. In view of these similarities, we decided to accept the TP3 and P3 bead types to describe the more polar, outward regions of the Glc molecule, whereas the inner part, belonging to the polymer backbone in the case of the Glc-based chains, was represented by the SN4 and TN4 beads. The choice of the P3 bead is in agreement with the polar character of hydroxymethyl and diol groups and with the 3-bead model for the glucose monomer existing in the current implementation of Martini 3.0. Its complementation by the N4 bead type is dictated by the match of the total polarity of the Glc molecule with the reference data.

Finally, it is worth noting that the change of one of the ring bead types from R to S has only marginal influence on the calculated log *P* values. The same can be stated in a more general context, about nearly all combinations of bead types, shown in Figure 2. This is discussed in further sections.

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

Table 2. Parameters for Glucopyranose Monomers (Bond Stretching, Bond-Angle Bending, and Improper-Dihedral Distortion) in the Presently Proposed Force Field[a]

type	topological pattern	parameters			
		α-anomer		β-anomer	
		k_b [kJ mol^{-1} nm^{-4}]	b_0 [nm]	k_b [kJ mol^{-1} nm^{-4}]	b_0 [nm]
bonds:	B1-B4	12,000	0.268	12,000	0.268
	B2-B3	20,000	0.284	24,000	0.291
	B3-B4	28,000	0.291	28,000	0.291
	B2-B4	28,000	0.342	32,000	0.355
		k_θ [kJ mol^{-1}]	θ_0 [deg]	k_θ [kJ mol^{-1}]	θ_0 [deg]
angles	B1-B4-B2[b]	340	85	450	81
	B1-B4-B3[b]	580	132	580	132
		k_ξ [kJ mol^{-1} deg^{-2}]	ξ_0 [deg]	k_ξ [kJ mol^{-1} deg^{-2}]	ξ_0 [deg]
improper dihedrals	B4-B3-B2-B1	200	9	200	9

[a]Force constants correspond to eqs 1−4. [b]Eq 2.

4.1.2. Bonded Interactions. Table 2 contains the optimized bonded parameters proposed for α- and β-glucopyranose monomers. Because of structural similarities of both anomers, only some minor differences concerning the B2 bead are involved. Consequently, the parameters for remaining D- or L-aldohexopyranoses can straightforwardly be obtained by using only a limited set of alterations in the currently proposed parameters in accordance to the corresponding AA simulations.

Parameters for saccharides containing two or more glucose units were developed separately for each of the following linkage types: $\beta(1 \rightarrow 2)$, $\beta(1 \rightarrow 3)$, $\alpha(1 \rightarrow 4)$, $\beta(1 \rightarrow 4)$, and $\alpha(1 \rightarrow 6)$. They are collected in Tables 34567. As mentioned, the accepted strategy of parameterization allows for exploiting the monosaccharide parameters in constructing the CG models for more complex saccharides. Thus, developing parameters for saccharide chains relied on adopting the monosaccharide-derived parameters (Table 2) in a nearly unchanged form and adding components responsible for modeling the glycosidic bonds. Thus, apart from relatively minor alterations resulting from reassignments of atoms contributing to beads connecting two adjacent residues, the monosaccharide parameters are fairly well conserved during this stage of parameterization.

The conformation of CG saccharide chains can be described by the following two determinants: (1) the geometry of the backbone, defined, depending on the linkage type, by the following beads: 4−2−8−6−12−10−... ($\alpha(1 \rightarrow 4)$ and $\beta(1 \rightarrow 4)$ linkages); 3−2−7−6−11−10−... ($\beta(1 \rightarrow 2)$ and $\beta(1 \rightarrow 3)$ linkages); and 1−4−2−5−8−6−9−12−10−... ($\alpha(1 \rightarrow 6)$ linkage); (2) the orientation of the rings (represented by the triplets of atoms: 2−3−4, 6−7−8, ...) with respect to the backbone. In the case of non-$\alpha(1 \rightarrow 6)$ linkages, the conformation of the backbone can be described by the two different types of quadruplets of atoms, creating repeating motifs: (1) quadruplet with two central atoms within the ring; (2) quadruplet with two central atoms creating the residue-residue linkage. The geometry of the dihedral angle included in the first of these motifs is controlled by the two improper-dihedral terms involving atoms that create linkages. Thus, because of the functional forms of the corresponding potential-energy terms (eq 4), the conformation of those backbone fragments will be represented by a single energy well. In contrast, the conformation of the dihedral angle associated with the second topological motif will be correlated with a rotation around the glycosidic linkage. However, the geometry of the same dihedral angle is convoluted in the rotation of neighboring rings around the backbone because of the restricted shapes of those rings. Thus, both abovementioned determinants are often treated collectively by introducing a series of improper-dihedral terms, controlling geometry within residue and adjacent, covalently bound beads and regular-dihedral term(s) controlling the rotations around glycosidic linkages. In the most complex case of the $\alpha(1 \rightarrow 6)$ linkage, the additional regular-dihedral terms are required, combined with bonds between nonadjacent beads, in order to maintain proper conformation of a chain.

Table 3. Parameters for $\alpha(1 \rightarrow 4)$-Linked Glucopyranose Di-, Oligo-, and Polysaccharides (Bond Stretching, Bond-Angle Bending, and Improper- and Regular-Dihedral Distortion) in the Presently Proposed Force Field[a]

type	topological pattern	parameters		
		k_b [kJ mol^{-1} nm^{-4}]	b_0 [nm]	
bonds:	B1-B4	18,000	0.251	
	B2-B3	32,000	0.280	
	B3-B4	34,000	0.279	
	B2-B4	60,000	0.292	
	B2-B8	32,000	0.280	
		k_θ [kJ mol^{-1}]	θ_0 [deg]	
angles	B1-B4-B2[b]	340	86	
	B1-B4-B3[b]	580	142	
	B3-B2-B8[b]	310	103	
	B4-B2-B8[c]	180	103	
	B1-B2-B8[b]	210	105	
	B2-B8-B5[b]	30	107	
	B2-B8-B6[c]	240	148	
	B2-B8-B7[b]	120	99	
		k_ξ [kJ mol^{-1} deg^{-2}]	ξ_0 [deg]	
improper dihedrals	B4-B3-B2-B1[d]	200	9	
	B8-B2-B7-B6	170	22.3	
	B2-B8-B4-B3	300	−67	
		k_ϕ [kJ mol^{-1}]	ϕ_0 [deg]	m
regular dihedrals	B4-B2-B8-B6	−12.8	161	1
	B3-B2-B8-B7	−2.2	165	1
	B1-B3-B7-B5	−3.6	−72	5

[a]Force constants correspond to eqs 1−5. Atom numbering concerns only the first linkage; the parameters for any subsequent, nth linkage can be obtained by increasing the corresponding bead numbers by $4n$. [b]Eq 2. [c]Eq 3. [d]Concerns only the nonreducing end.

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

Table 4. Parameters for $\beta(1 \rightarrow 4)$-Linked Glucopyranose Di-, Oligo-, and Polysaccharides (Bond Stretching, Bond-Angle Bending, and Improper- and Regular-Dihedral Distortion) in the Presently Proposed Force Field[a]

type	topological pattern	parameters		
		k_b [kJ mol^{-1} nm^{-4}]	b_0 [nm]	
bonds:	B1-B4	14,100	0.250	
	B2-B3	37,500	0.268	
	B3-B4	27,000	0.273	
	B2-B4	53,200	0.257	
	B2-B8	7500	0.267	
	B2-B6	16,300	0.520	
	B4-B8	3770	0.542	
		k_θ [kJ mol^{-1}]	θ_0 [deg]	
angles	B1-B4-B2[b]	220	91	
	B1-B4-B3[c]	159	143	
	B3-B2-B8[b]	245	115	
	B1-B2-B8[b]	350	127	
	B2-B8-B5[b]	16	123	
	B2-B8-B7[b]	52	93	
		k_ξ [kJ mol^{-1} deg^{-2}]	ξ_0 [deg]	
improper dihedrals	B4-B3-B2-B1[d]	200	9	
	B8-B2-B7-B6	212	11	
	B2-B3-B8-B4	229	9	
		k_ϕ [kJ mol^{-1}]	ϕ_0 [deg]	m
regular dihedrals	B3-B2-B8-B7	−35	−135	1

[a]Force constants correspond to eqs 1−5. Atom numbering concerns only the first linkage; the parameters for subsequent, nth linkage can be obtained by increasing the corresponding bead numbers by $4n$. [b]Eq 2. [c]Eq 3. [d]Concerns only the nonreducing end.

Let us note that in some cases, the bonds between nonadjacent beads were introduced in order to control the geometry of the triplets of atoms. This was done mainly in those systems where the geometry of those triplets becomes close to linear, resulting in potential numerical errors and simulation instability. For the same reasons, the regular bond angle term (eq 2) was replaced by the restricted-angle term (eq 3) for flexible bond angles with optimal values close to 180°.

4.2. General Remarks. The parameters collected in Tables 2−7 were developed, tested, and validated under assumption that the first covalently bound neighbors are excluded from any nonbonded interactions.

In most of the systems, the developed parameters allow to carry out the numerically stable MD simulations with a timestep of 0.02 ps or even 0.03 ps (monosaccharides). The exception are the systems containing either the $\beta(1 \rightarrow 2)$ or $\beta(1 \rightarrow 3)$ linkages where the probability of linearization of atoms triplets involved in either bond angle of the dihedral type of bonded interactions is too large. In the case of these two system types, the stability of MD simulations was achieved upon decreasing the timestep to 0.01 ps.

Interestingly, several features that seem to be largely dependent on the bead type (e.g., log P, SASA, density of glucose solutions; see details in further sections of the article) were relatively weakly affected by alternative assignments of the bead type. This includes both bead types and sizes (R, S, and T, according to the MARTINI-characteristic distinction). Moreover, the conformational properties of both monomers and octamers, relying mainly on the bonded parameters, are fairly independent of the choice of nonbonded parameters.

Table 5. Parameters for $\beta(1 \rightarrow 3)$-Linked Glucopyranose Di-, Oligo-, and Polysaccharides (Bond Stretching, Bond-Angle Bending, and Improper- and Regular-Dihedral Distortion) in the Presently Proposed Force Field[a]

type	topological pattern	parameters		
		k_b [kJ mol^{-1} nm^{-4}]	b_0 [nm]	
bonds:	B1-B4	10,000	0.268	
	B2-B3	38,000	0.240	
	B3-B4	24,000	0.287	
	B2-B4	50,000	0.287	
	B2-B7	12,000	0.250	
		k_θ [kJ mol^{-1}]	θ_0 [deg]	
angles	B1-B4-B2[b]	220	78	
	B1-B4-B3[b]	400	123	
	B3-B2-B7[b]	60	75	
	B1-B2-B7[c]	80	148	
	B2-B7-B6[b]	110	128	
	B2-B7-B8[b]	40	65	
		k_ξ [kJ mol^{-1} deg^{-2}]	ξ_0 [deg]	
improper dihedrals	B4-B3-B2-B1	120	15	
	B7-B2-B8-B6	200	22	
	B2-B7-B3-B4	200	3	
		k_ϕ [kJ mol^{-1}]	ϕ_0 [deg]	m
regular dihedrals	B3-B2-B7-B6	−20	−174	1

[a]Force constants correspond to eqs 1−5. Atom numbering concerns only the first linkage; the parameters for subsequent, nth linkage can be obtained by increasing the corresponding bead numbers by $4n$. [b]Eq 2. [c]Eq 3.

Table 6. Parameters for $\beta(1 \rightarrow 2)$-Linked Glucopyranose Di-, Oligo-, and Polysaccharides (Bond Stretching, Bond-Angle Bending, and Improper-Dihedral Distortion) in the Presently Proposed Force Field[a]

type	topological pattern	parameters		
		k_b [kJ mol^{-1} nm^{-4}]	b_0 [nm]	
bonds:	B1-B4	10,000	0.265	
	B2-B3	50,000	0.263	
	B3-B4	24,000	0.272	
	B2-B4	50,000	0.290	
	B2-B7	6000	0.274	
	B4-B7	600	0.530	
		k_θ [kJ mol^{-1}]	θ_0 [deg]	
angles	B1-B4-B2[b]	200	80	
	B1-B4-B3[b]	400	132	
	B3-B2-B7[b]	30	88	
	B1-B2-B7[c]	60	165	
	B2-B7-B6[b]	50	97	
	B2-B7-B8[c]	40	148	
		k_ξ [kJ mol^{-1} deg^{-2}]	ξ_0 [deg]	
improper dihedrals	B4-B3-B2-B1	150	11	
	B7-B6-B8-B2	300	15	
	B2-B4-B3-B7	200	1	
		k_ϕ [kJ mol^{-1}]	ϕ_0 [deg]	m
regular dihedrals	B3-B2-B7-B6	−22	−141	1

[a]Force constants correspond to eqs 1−5. Atom numbering concerns only the first linkage; the parameters for subsequent, nth linkage can be obtained by increasing the corresponding bead numbers by $4n$. [b]Eq 2. [c]Eq 3.

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

Table 7. Parameters for $\alpha(1 \rightarrow 6)$-Linked Glucopyranose Di-, Oligo-, and Polysaccharides (Bond Stretching, Bond-Angle Bending, and Improper- and Regular-Dihedral Distortion) in the Presently Proposed Force Field[a]

type	topological pattern	parameters		
		k_b [kJ mol^{-1} nm^{-4}]	b_0 [nm]	
bonds:	B1-B4	21,000	0.251	
	B2-B3	35,000	0.280	
	B2-B4	60,000	0.322	
	B3-B4	24,000	0.289	
	B2-B5	13,000	0.225	
	B3-B7	2400	0.820	
	B1-B9	180	0.790	
		k_θ [kJ mol^{-1}]	θ_0 [deg]	
angles	B1-B4-B2[b]	600	80	
	B1-B4-B3[b]	560	135	
	B3-B2-B5[b]	180	109	
	B4-B2-B5[b]	100	105	
	B2-B5-B6[c]	28	122	
		k_ξ [kJ mol^{-1} deg^{-2}]	ξ_0 [deg]	
improper dihedrals	B4-B2-B3-B1	250	−7	
	B2-B3-B5-B4	110	−74	
		k_ϕ [kJ mol^{-1}]	ϕ_0 [deg]	m
regular dihedrals	B3-B2-B5-B6	−3	155	2
	B2-B5-B6-B7	−3	170	5
	B2-B5-B6-B7	3	60	1
	B2-B4-B8-B6	5	−25	1
	B2-B4-B8-B6	−4	165	3

[a]Force constants correspond to eqs 1−5. Atom numbering concerns only the first linkage; the parameters for subsequent, nth linkage can be obtained by increasing the corresponding bead numbers by $4n$. [b]Eq 2. [c]Eq 3.

Thus, although we have accepted the P3 + N4 combination as the final one, some other, chemically sound choices may also provide a reasonable modification of the model. The main motivation for introducing such alterations may be the modification of interaction strength either within the carbohydrate−carbohydrate pairs or between carbohydrates and other biomolecules or solvents.

5. PROPERTIES OF THE STUDIED SYSTEMS

The subsequent subsections report the results obtained by using the newly developed set of parameters to study various carbohydrate-containing systems. These results aim to provide a validation of the force field and, in part, were used to adjust the final parameters.

5.1. SASA. The SASA parameter values were calculated at both CG and AA levels of resolution for mono-, di-, and octasaccharides. For nonmonomeric saccharides, the calculations involved all considered glycosidic linkage types. The results are illustrated in Figure 3. This part of the study confirms that the CG model is capable of accurately reproducing the molecular surface and molecular volume properties, as essential for Martini 3.0-based models.[33,53] Interestingly, replacing the R bead in the ring structures into the corresponding S one has a negligible influence on the calculated SASA values. In both cases, the relative and absolute deviations from the atomistic data are extremely small and of comparable magnitude.

5.2. Conformational Properties. The parameters collected in Tables 2−7 were used to produce a series of models for carbohydrate mono- and octamers, subsequently applied to validate the conformational properties against the predictions of AA simulations. This was done by comparing the distribution of selected conformational descriptors: (1) distances between selected CG bead pairs; (2) angles between triples of beads; (3) improper dihedrals defined by quadruplets of beads; (4) regular torsional angles defined by quadruplets of beads; (5) end-to-end distances; (6) gyration radii. Because of numerous possible combinations of beads defining a given descriptor, we limited ourselves to analyzing mainly those that can be directly traced back to certain terms in the force field (e.g., given bead-bead bond or dihedral angle defined on a rotatable bond). In the case of octamers, the analysis mentioned in points (1)−(4) always concerned central residue(s) of the chain. Covalent bead-bead bonds, bond angles, and dihedral terms are, as a rule, modeled by the harmonic potential (see the Model section). Thus, in order to simplify the comparison for descriptors of those types, we provide only the average values and the accompanying fluctuation measures (standard deviations on the set of values). Independent of that, the full comparison of the distributions was carried out during the stage of parametrization and data analysis. On the contrary, the regular-dihedral angles are reported as distributions recovered from AA and CG simulations.

The results are given in Figure 4 (bonds, angles, and improper-dihedral average values), 5 (regular-dihedral distri-

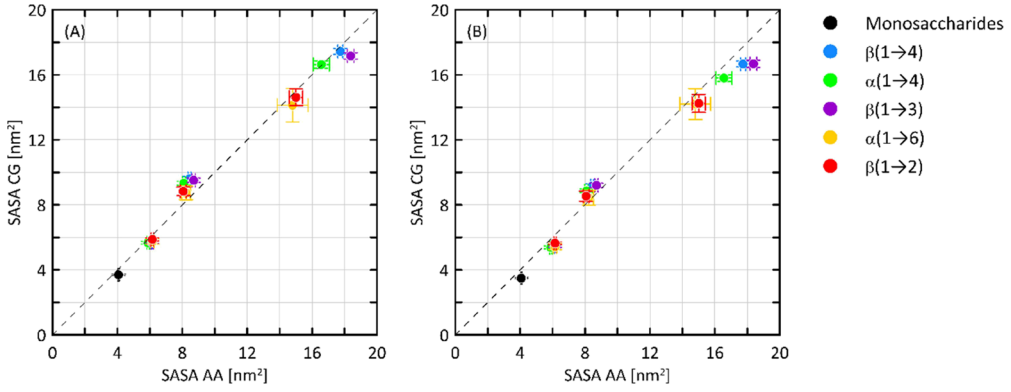


Figure 3. Comparison of the SASA values obtained from either AA of CG simulations. Two alternative assignments of the bead type for one of the ring beads were accepted: (A) S bead; (B) R bead.

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

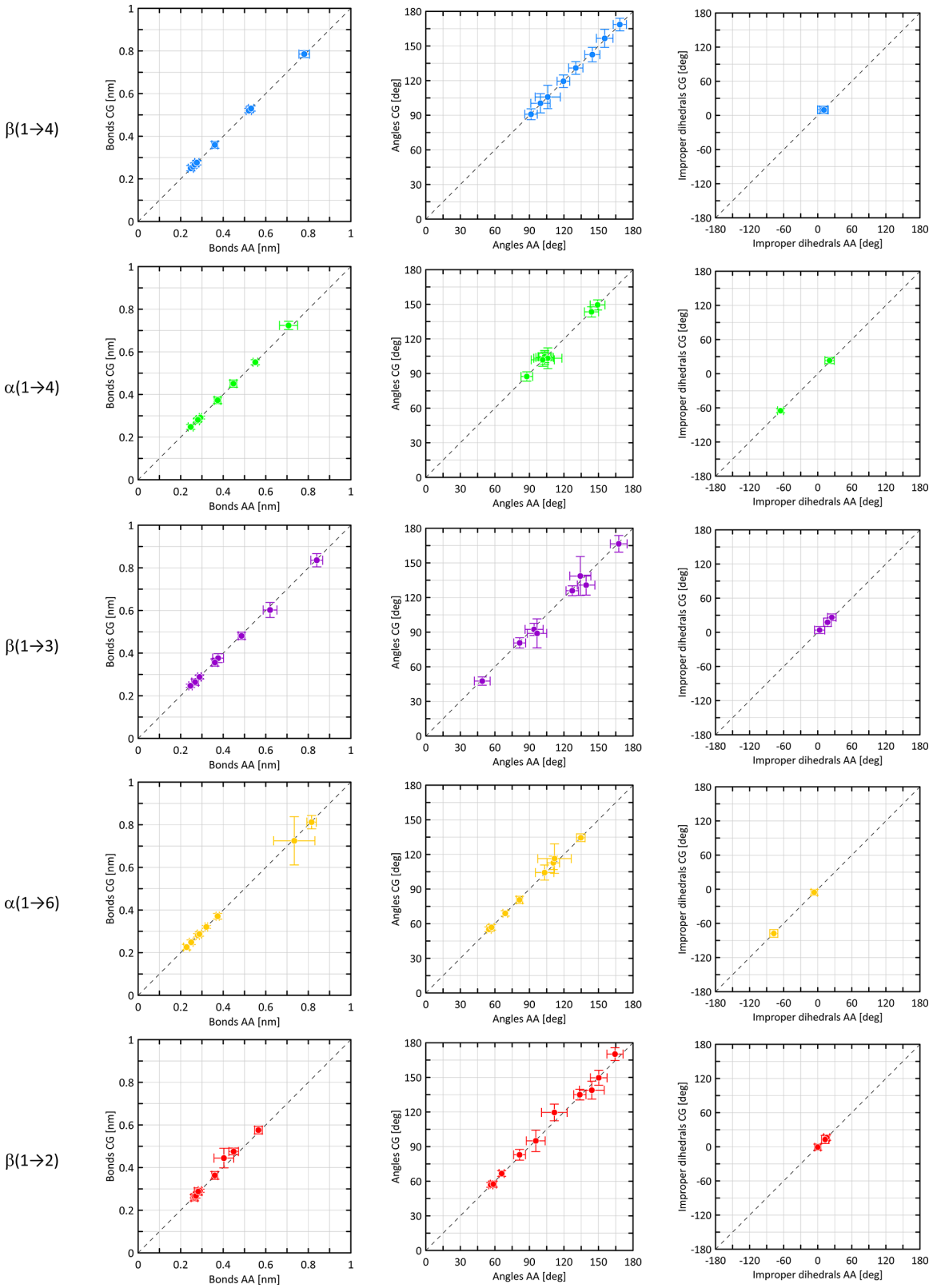


Figure 4. Comparison of the average values of the selected bead-bead distances, bead-bead-bead regular angles, and improper-dihedral angles calculated from unbiased MD simulations within either the AA or CG force field for different types of glycosidic linkages. Horizontal and vertical bars denote the fluctuation associated with a given descriptor, expressed as the standard deviation on the data set.

butions), and 6 (gyration radii and end-to-end value distributions). In most of the cases, the agreement between AA and CG data is excellent or at least satisfactory. In the case of bonds, bond angles, and improper dihedrals, most essential deviations are the consequence of the inherent asymmetry of the AA distributions approximated by symmetrical, parabolic potentials. However, deviations of this type are not frequent and the level of asymmetry as well as deviations from unimodal

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

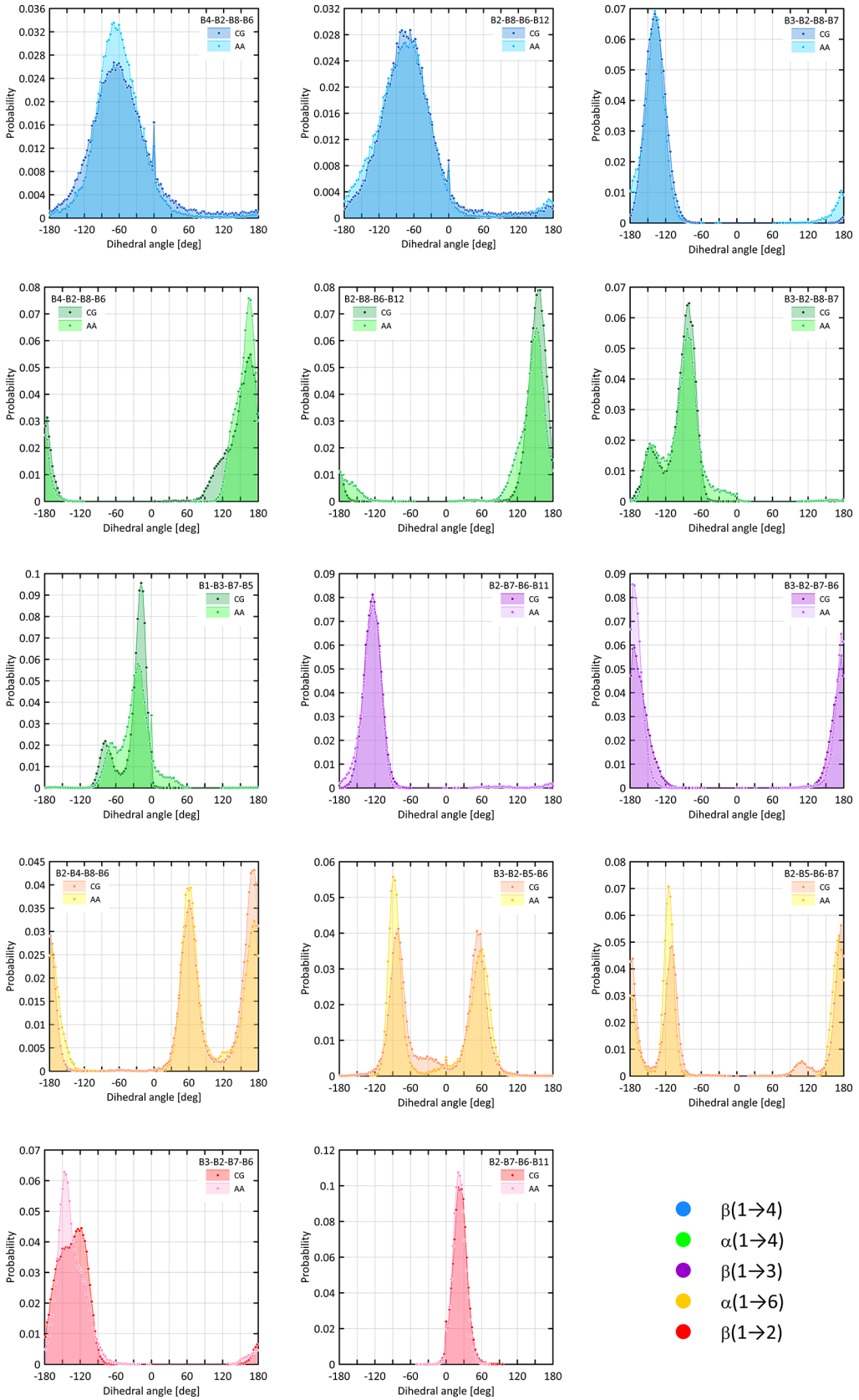


Figure 5. Distributions of dihedral angle values calculated within either the AA or CG force field. The angles correspond to the selected rotation around glycosidic linkages of different types.

character of distributions are not large, which results in a reasonable level of agreement between AA and CG data (Figure 4). The most pronounced consequence in this context is ignoring the secondary conformers for polysaccharide

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

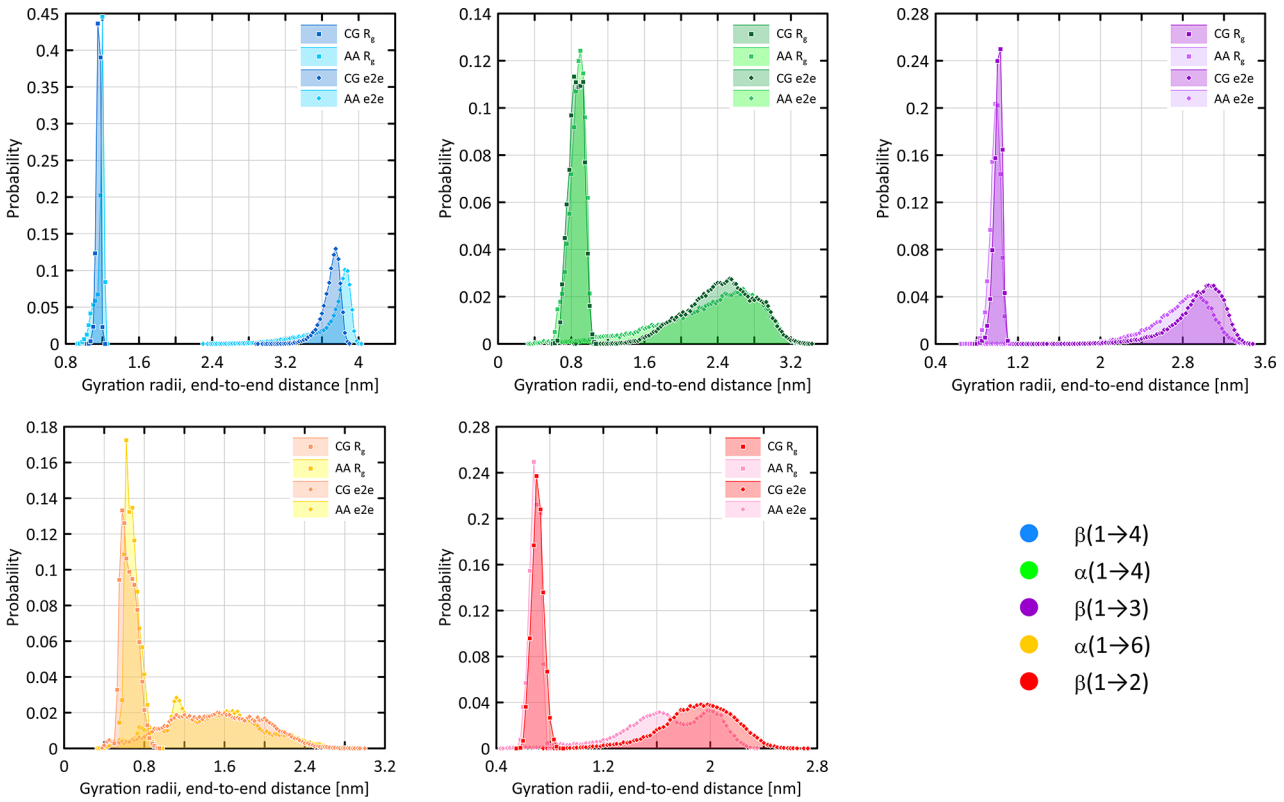


Figure 6. Distributions of gyration radii (R_g) and end-to-end distance (e2e) values calculated within either the AA or CG force field. Calculations concerned homooctasaccharides containing different types of glycosidic linkages.

backbone in chains composed of $\beta(1 \rightarrow 4)$ linkages. This approximation is quite accurate, as the population of neglected conformers is ca. 2.5% but can have some influence when studying backbone kinks in extremely long chains. At a smaller scale, it underestimates the rigidity of oligomeric chains containing $\beta(1 \rightarrow 4)$ linkages expressed by the end-to-end distance (deviation between average values equal to ca. 3.5%) but has very little influence on the gyration radii (deviation equal to ca. 1%). On the other hand, a series of independent CG simulations carried out for very long cellulose chains (>50 units) shows that the alternative measure of chain rigidity, that is, the persistence length, is overestimated in comparison to the available experimental data.

This effect and other potential inaccuracies concerning the case of long saccharide chains may result from the fact that the current CG model partially ignores the possibility of "kinks" along the carbohydrate backbone. Such kinks represent the secondary or tertiary conformers of glycosidic linkages (anti-ϕ and anti-ψ rotamers). However, the conformational flexibility within the main conformational state is reproduced very well, as shown in Figures 4–6. Moreover, the presence of anti-ϕ and anti-ψ rotamers in some of the glucose-based polysaccharides is partially accounted for by possible rotations of residues around the backbone. By allowing for that, we preserve the dynamic equilibrium between relative orientations of the B1 and B3 (or only B3 in the case of $(1 \rightarrow 6)$ linkages) beads of the neighboring residues even if the backbone shape does not reflect the correlated reorientation.

The accurate implementation of the "kinks" in the saccharide backbone will be the subject of our future studies, focused on those members of the carbohydrate family for which the flexibility of both backbone and rings plays a more significant role (e.g., the iduronate-containing glycosaminoglycans[11]).

The agreement between AA and CG data in the context of regular-dihedral angles is shown in Figure 5. Distributions obtained for linkages of the $\beta(1 \rightarrow 2)$, $\beta(1 \rightarrow 3)$, and $\beta(1 \rightarrow 4)$ types have a relatively simple character, indicating the existence of a single free energy well. In contrast, the distributions calculated for $\alpha(1 \rightarrow 4)$ and $\alpha(1 \rightarrow 6)$ linkages are more complex and associated with more than one well-defined conformational states. However, both these tendencies are satisfactory, reproduced by the CG model. The most significant deviations are often a consequence of inability to find the function modeling the CG torsion of multiplicity, which cannot be expressed by an integer (see eq 5).

Finally, the good agreement between polymer properties (gyration radii and end-to-end distance) calculated at CG and AA levels is worth mentioning (Figure 6). The larger deviations between average values are characteristic of the $\beta(1 \rightarrow 2)$, $\beta(1 \rightarrow 3)$, and $\beta(1 \rightarrow 4)$ linkages, exhibiting unimodal character of distributions of glycosidic torsional angle values (Figure 5). Nevertheless, in all of the cases, the average value and even the nonsymmetric character of AA distributions are reasonably well reflected by the corresponding CG model.

5.3. Density. Our CG model was tested against its capability to reproduce the density of aqueous solutions of glucose. A systematic set of CG simulations was carried out at different glucose concentrations and compared with the experimental data. Figure 7 shows the graphical illustration of such comparison. The experimental densities are reproduced with a satisfactory accuracy up to maximal experimentally measured concentrations of ca. 0.1 molar fraction.

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

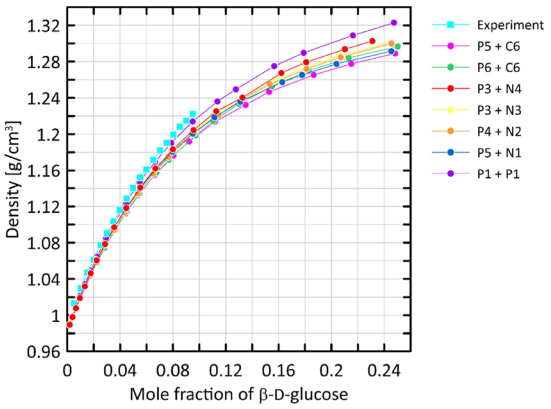


Figure 7. Density of aqueous solutions of β-D-glucopyranose as a function of the molar ratio. Results from CG simulations (circles) are compared with the experimental data[54] (squares).

The final P3 + N4 combination of bead types is associated with the error of prediction of ca. 1.5% and is the second best among all tested nonbonded parameters. However, as previously mentioned, all investigated combinations offer a similar level of agreement with reference data. Moreover, we did not observe any spurious sugar-sugar aggregation at high concentrations, up to 0.25 molar ratio.

5.4. Self-Assembly Properties. Because of a large polarity of carbohydrate molecules, the self-aggregation properties can often be seen as spurious artifact being a result of inaccuracies in the force field parameters.[27,32] However, there exist a series of supramolecular structures created by self-interacting carbohydrate oligo- and polymers. The examples include helical structures created by curdlan chains,[55,56] double amylose helix,[57] and amylose V-helix present in some nonaqueous solutions[58] or cellulose aggregation resulting in cellulose II-like complexes.[59] The capabilities of the newly developed model were tested in the context of either spontaneous creating the carbohydrate aggregates or maintaining the structures initiated from the molecular geometries roughly corresponding to the given complex.

It was found that the double-stranded amylose helix of the structural features based on the XRD data is unstable and undergoes unfolding within several nanoseconds. This observation relies on the unbiased MD simulation in aqueous solution. The capability of the model to maintain the amylose V-helix in nonaqueous solutions was investigated in a systematic manner by dissolving the V-helix-like structure in a series of artificial solvents, composed of uniform beads of the following types: C1, C2, C3, C4, C5, and C6. However, in none of the cases, the initial structure was maintained. Instead, more compact aggregates have been formed in the presence of less polar solvents (beads C1 and C2), whereas unfolding of the V-helix occurred in polar solvents (beads C3 to C6).

On the other hand, the unbiased MD simulations of concentrated solutions of octamers with the $\beta(1 \rightarrow 4)$ (cellulose) and $\beta(1 \rightarrow 3)$ (curdlan) linkages resulted in spontaneous formation of structures corresponding to curdlan triple helices and cellulose II-like sheets. In the latter case, the sheets exhibited a tendency to further aggregation, creating clusters that resemble the structure of the cellulose II crystal. The exemplary snapshots from MD simulations are shown in Figure 8. Additionally, we have studied the structural parameters of the curdlan triple helix composed of longer 32-residue-long chains. The helix length (per 1 helix turn) predicted by CG simulation is equal to ca. 1.95 nm, which is in between the predictions of the GROMOS $56A6_{CARBO\_R}$ united-atom force field,[60] that is, 1.79 nm for a single helix in solution and the value of 2.02 nm for helices packed in parallel in the supramolecular sheet.[61]

At the stage of testing various nonbonded parameters, we have observed that exchanging the R bead type (one of the beads defining the glucose ring) into the S one resulted in alterations of the aggregation properties. Namely, the carbohydrate−carbohydrate interactions become less attractive and, as a consequence, the events of self-aggregation less frequent. Still, at sufficiently high concentrations, aggregates of the types described above were possible to observe.

Thus, concluding, our CG model is capable of predicting the carbohydrate self-assembly properties in the cases where

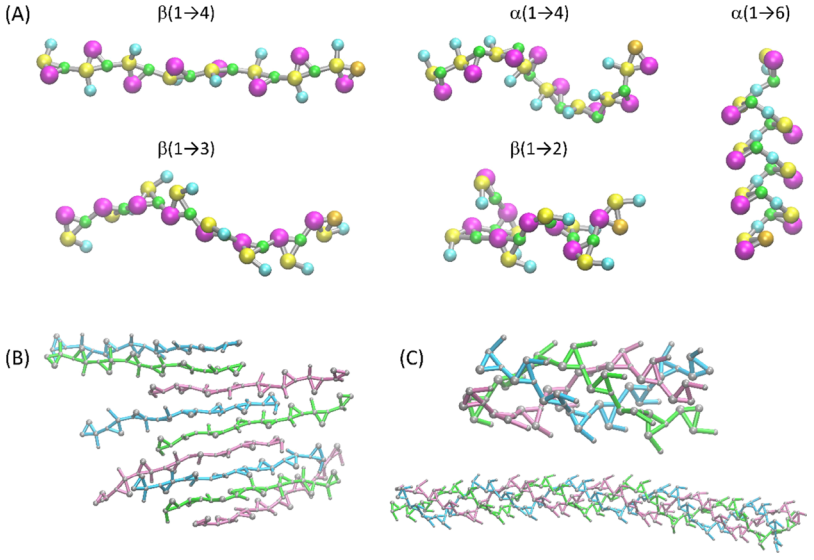


Figure 8. (A) Exemplary snapshots from the MD simulation of octamers of glucopyranose, exploiting different types of glycosidic linkages. (B) Cellulose II-like sheet formed during the simulation of aqueous solution of cellulose oligomers. (C) Triple helices of curdlan formed during the simulation of the aqueous solution of curdlan oligomers (*up*) or obtained as a result of simulation of the XRD data-based structure (*down*).

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

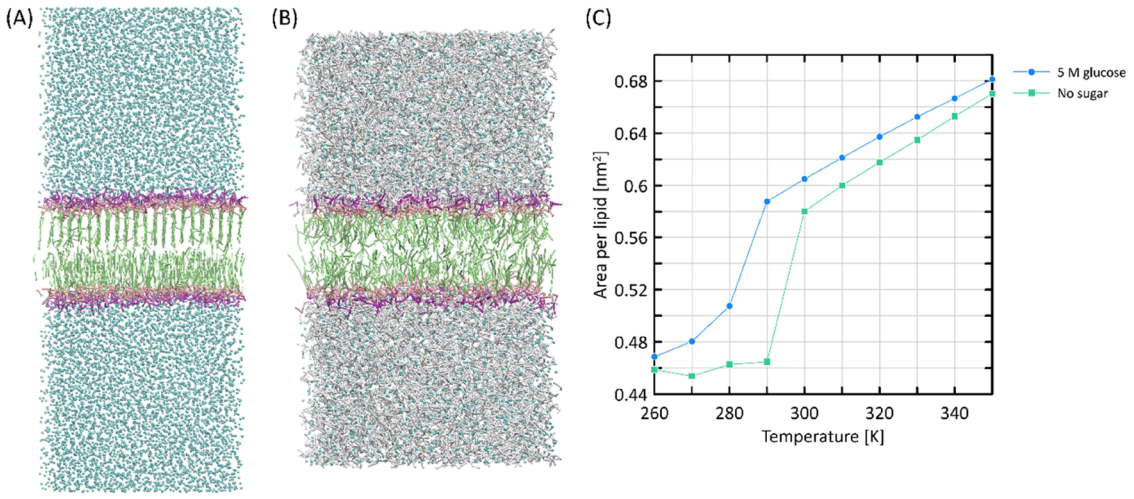


Figure 9. (A, B) Snapshots of CG DPPC bilayers at T = 290 K: (A) gel phase formed in pure water; (B) fluid phase stabilized in the presence of 4 M glucose. Lipid tails are depicted in green, head groups are shown in pink and violet, and glucose molecules shown in white. Water beads are represented by blue balls. (C) Average area per lipid of a DPPC bilayer as a function of temperature. Glucose-free (green) and 5 M glucose solution (blue) systems were considered.

aggregation is driven by the contact of aliphatic patches on the pyranose rings and, possibly, supported by hydrogen interactions. Such a situation occurs primarily in the case of chains containing equatorial-equatorial glycosidic linkages. On the other hand, the structures supported mainly by hydrogen bonding are poorly reflected by the model, probably due to the directional character of such interactions, which is hardly reproduced by the LJ interactions between particular beads.

5.5. Interactions with the Lipid Bilayer. Carbohydrates are capable of acting as cryo- and anhydro-protective agents for lipid membranes.[62] One of the possible mechanisms of this effect relies on alterations of the main phase transition temperature of the lipid bilayer, which is lowered in the presence of carbohydrates. This phenomenon has been studied by means of both AA[63−65] and CG[23] simulations. Our model was tested against the capabilities to reproduce this effect. Several CG simulations of the DPPC lipid bilayer were carried out either in the absence or the presence of ca. 5 M glucose and with the temperature varying in the range of 260−350 K. The effect of carbohydrate influence was monitored by the area per lipid parameter. Figure 9 illustrates the results.

There exists a striking difference between the behavior of the two systems: the pure DPPC system adopts a sharp increase of area per lipid (APL) from ca. 0.46 to ca. 0.58 nm^2 around 295 K, characteristic of a transition from a gel phase to a fluid phase. The glucose-containing system exhibits similar phase transition but at lower temperature, equal to ca. 283 K. At even lower temperatures (260−280 K), the glucose-containing system displays systematically higher APL values, whereas at the elevated temperatures (300−350 K), the thermal expansivity (i.e., the slope of the APL vs temperature curves) is slightly smaller in the presence of the sugars. All these findings confirm that glucose influences the structure of the lipid bilayer, increasing the stability of the fluid phase. The insight into MD trajectories clearly shows that the lower APL values are correlated with the conformation of the lipid tails forming the inner part of the bilayer (Figure 9). In the pure system, the DPPC tails below 290 K adopted am extended, straight structure, whereas in the presence of glucose, they remain more disordered.

In analogy to the results reported for Martini 2,[23] also in this case, the analysis of MD trajectories indicates that water-replacement hypothesis[64,65] can be evoked to explain the mechanism of the observed results. The carbohydrate molecules are able to bind to the lipid/water interface (via interaction with the lipid head groups) and can penetrate the membrane up to the level of the carbonyl groups. By the competition effect, the amount of water in the water/membrane interface is reduced. Although the physical mechanism of interactions between lipid bilayers, monosaccharides, and water is more complex and dependent on additional factors (e.g., sugar concentration),[66] these findings demonstrate that the current model can reflect the carbohydrate-induced alteration in the structure of lipid bilayers.

5.6. Interactions with Proteins. Interactions of carbohydrates with proteins are essential for numerous processes occurring in living organisms. Thus, the protein−carbohydrate interactions are extensively studied by MD simulations at either the AA or CG level or resolution. In addition to the routine procedure relying on initiating the simulation with the correct, experimentally determined position of the ligand in the protein binding cavity, the extensive MD simulations may sporadically be used to blindly identify binding sites. The recent version of Martini proved to be capable of successfully carrying out such predictions.[67] Presently, we have tested if the carbohydrate-binding cavities in several different proteins can be identified by using the Martini 3.0 force field in combination with our model. Carbohydrate−protein binding may, in principle, be problematic for classical force fields due to the fact that in the case of nonfunctionalized carbohydrates, binding is driven by the CH-π interactions.[68,69] In spite of the lack of explicit functional form capable of mimicking this type of interactions, most of modern atomistic, biomolecular force fields offer reasonable predictions of the binding free energy, but some divergences arise when considering the quantitative descriptors of the binding pattern and strength.[70] Additionally, either functionalized or unfunctionalized carbohydrates are highly hydrophilic, and binding to protein is associated with small or moderate changes in free energy. Thus, even small

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

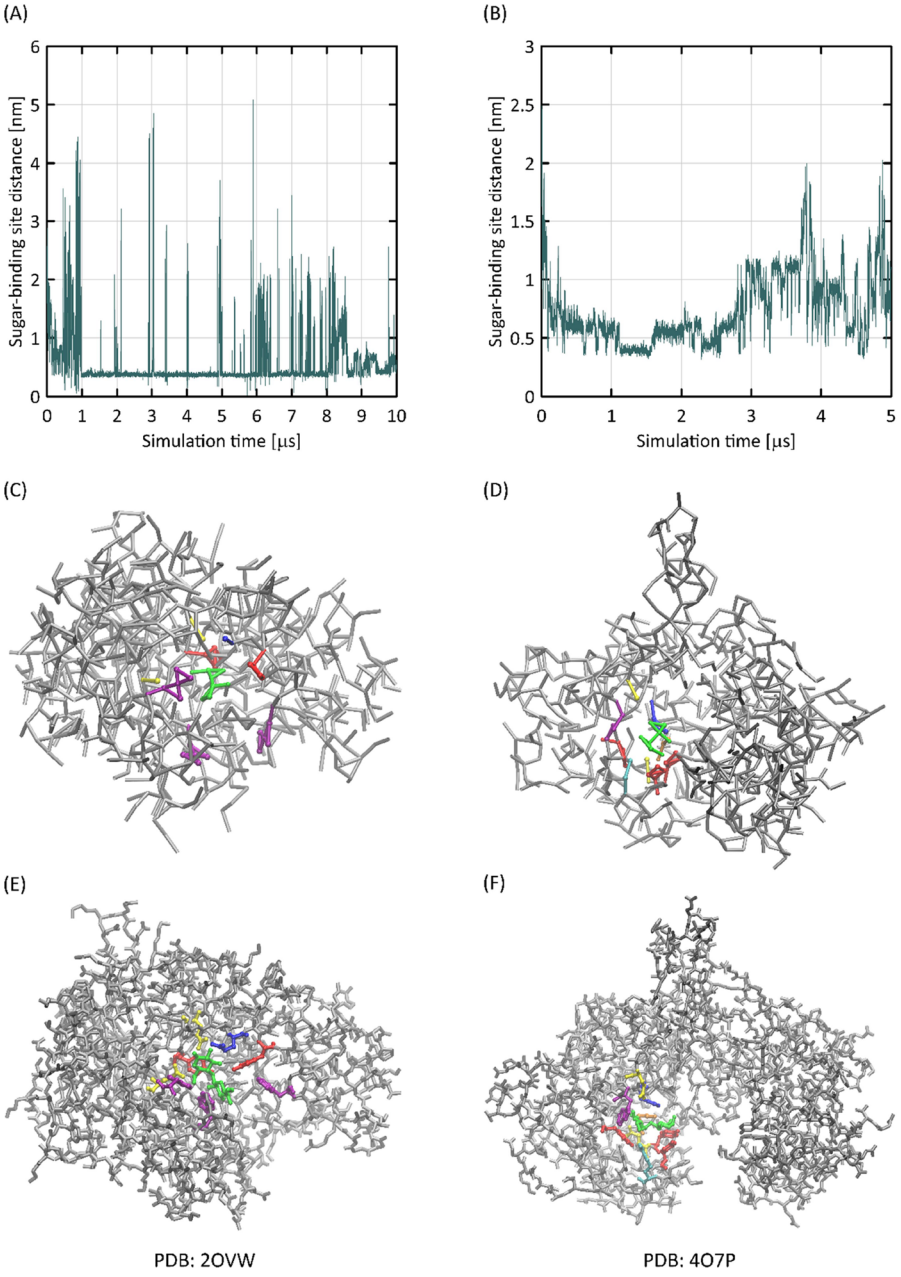


Figure 10. (A, B) Binding cavity-carbohydrate molecule distance recorded during MD simulations for protein−carbohydrate systems. Lowest values correspond to the binding event. (C, D) CG structures of the protein−carbohydrate complex and (E, F) atomistic, XRD structures, corresponding to them. In the case of panels (C−F), the same color code was used, highlighting, for example, carbohydrates (green) and two types of aromatic amino-acid residues, responsible for CH-π binding: tryptophane (violet) and tyrosine (red).

inaccuracies in the force field can result in its inapplicability to correctly predict any carbohydrate binding-related properties.

At this stage of the study, we investigated carbohydrate−protein binding in the three distinct systems: (1) endoglucanase I interacting with cellobiose (PDB: 2OVW); (2) maltose kinase interacting with maltose (PDB: 4O7P); (3) xylanase 10A interacting with glucose (PDB: 1I8A). Our aim was to check if the few tens microseconds-long MD simulation will allow to identify the experimentally determined carbohydrate binding sites.

The results for systems (1) and (2) are illustrated in Figure 10. It appears that in both cases, the binding site is correctly identified with neither a priori assumed configuration nor any bias imposed to the system. The binding pattern is driven by interactions of "mixed" nature, that is, the CH-π interactions with tyrosines and tryptophans and hydrogen bonding with, for example, arginine or aspartic acid. Thus, even in spite of highly hydrophilic nature of unfunctionalized glucose di- and oligomers, the newly developed model is capable of correctly predicting the binding pocket, in analogy to noncarbohydrate ligands and their protein molecular targets, as described in a previous study.[67]

Interestingly, the same set of parameters applied for the system in which carbohydrate binding is driven mainly by the CH-π interactions with only a minor (or none) contribution of polar interactions gives opposite results. Such an example is xylanase 10A (structure deposited at PDB:1I8A), where the bound ligand is β-glucopyranose and the binding pocket is

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

created by the two tryptophan sidechains. In this case, we did not observe any specific binding with the contribution of the experimentally identified binding pocket. Some binding events were observed when artificially reducing the hydrophilicity of the glucose beads. However, as such reduction is significant (e.g., converting N4 bead types into C4 and P3 into P2), this type of alteration is not recommended, as it will likely lead to numerous artifacts, associated with the presence of "hydrophobic sugars" in the system.

In conclusions, we state that our CG model is capable of identifying the carbohydrate binding pockets in the "blind" MD simulations, but the necessary condition for that is a large contribution of polar interactions that provide the driving force for binding. Otherwise, when binding is driven solely by the CH-π interactions, such predictions may not be successful. On the other hand, quantitative prediction of the energy involved in the CH-π interactions in carbohydrate−protein complexes can be problematic even in the case of atomistic force fields.[70]

6. MODEL LIMITATIONS AND POTENTIAL REFINEMENTS/EXTENSIONS

The developed CG force field contains a series of limitations which, as a rule, are a consequence of coarse-grained resolution of the underlying model.

1. The interactions with a strong orientation dependence (e.g. hydrogen bonding) are modeled by orientation-free LJ interactions between beads. As a consequence, hydrogen bonding-mediated self-assembly in some carbohydrate-containing systems may not be accurately reproduced. This concerns mainly the formation of double-helix of amylose in polar solvents and amylose V-helix in non- or weakly polar solvents. On the other hand, self-assembly of "hybrid" nature, induced by both polar- and non-polar contacts (e.g., triple helices of curdlan and cellulose II structure) are reproduced correctly (Section 5.3),
2. Binding of small carbohydrate molecules by proteins driven nearly exclusively by CH-π interactions is not correctly reflected. The reason is the lack of sufficiently hydrophobic beads representing the nonpolar patch of the glucose ring, responsible for CH-π binding. On the contrary, carbohydrate−protein binding occurring with larger contribution of polar interactions is reflected properly (Section 5.4).
3. Some of the conformational rearrangements occurring within the carbohydrate backbone are correlated with the alterations of the geometry of pseudo-chiral centers, described by the "improper-dihedral" force field terms. These terms are represented by parabolic potentials with single minimum; thus, they are not capable of reflecting any nonharmonic conformational changes. This issue concerns mostly the $\beta(1 \rightarrow 4)$ linkages. As a consequence, the long polysaccharide chains containing this linkage type (cellulose) will suffer from the lack of inherent conformational "kinks", which will result in apparent stiffening of the chain and influence several polymer properties (e.g., persistence length, end-to-end distance). However, in the studied oligomeric systems, this effect is likely to have a limited influence.
4. The conformational transitions of the chair-to-inverted chair or chair-to-boat type occurring within pyranose ring are not modeled at all. This is partially due to the complexity of the problem considered in the context of limitations of the functional forms of the bonded potential combined with the simplicity of the accepted CG representation. Nevertheless, this issue does not seem to be especially important for currently considered glucopyranose-based saccharides because of the fact that the glucopyranose ring belongs to the most rigid ones and the population of non-$^4C_1$ conformers is usually negligible under standard conditions.
5. In connection to points 3 and 4, it is worth mentioning that the atomistic force field differ significantly in their predictions related to free energy levels associated with some secondary or tertiary conformers. This includes, for example, the nonstandard pyranose ring conformers[71] and rotamers of glycosidic bonds (Lutsyk and Plazinski, unpublished data). As CG models rely usually on the AA reference data, some of these uncertainties may also be treated as a weak point of the current CG model.
6. The developed set of parameters was tested only in the context of various homopolymers. Although the currently proposed set is theoretically capable to simulate unbranched heteropolymers as well as branched polysaccharides (under assumption that the branching type corresponds to any of the glycosidic linkages considered in this work), systems of those character were not studied so far. Therefore, it is expected that some refinements may be necessary, at least at the level of bond-length alterations resulting from minor redefinitions of beads.

The issues mentioned in points 1 and 2 may potentially be addressed by introducing alternative sets of parameters designed with the aim to deal with those problematic situations. For instance, a separate set of parameters can be introduced to simulate amylose in nonpolar solvents (in analogy to the previous version of Martini), whereas CH-π binding can be enforced by intentionally decreasing the polarity of certain beads in the carbohydrate model (ignoring the potential consequences for, for example, log P value). Furthermore, maintaining the supramolecular structures of long carbohydrate chains may be supported by introducing rationally designed constraints between particular fragments of the system (in analogy to the elastic network approach routinely applied in the protein-dedicated editions of Martini). In order to capture the rare conformational transitions (points 3 and 4), nonstandard functional forms of some bonded potential may be necessary. This includes, for example, GROMACS-compatible tabulated potentials. Some of these solutions will be explored in our future studies.

7. CONCLUSIONS

The CG model, compatible with the new version (3.0) of the Martini biomolecular force field and concerning glucopyranose-based saccharides has been developed. The proposed parameters that constitute the model are described and discussed in the previous sections of the article, whereas the Supporting Information contains the program facilitating the automated generation of the force field files (topologies and structures in the GROMACS format) for carbohydrate chains of user-defined lengths. The model covers the parameters for glucose-containing saccharides, in particular, glucopyranose monomers (separately for α or β anomers), disaccharides

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

containing the linkages of the $\beta(1 \rightarrow 2)$, $\beta(1 \rightarrow 3)$, $\alpha(1 \rightarrow 4)$, $\beta(1 \rightarrow 4)$, or $\alpha(1 \rightarrow 6)$ types, as well as oligo- and polysaccharides exploiting such linkages. The capabilities of our model were tested through multiple MD simulations, confirming that the properties of glucose monomers and polymers simulated at the CG level are in accordance with the predictions of AA models and with the available experimental data. More precisely, we were able to satisfactorily reproduce the density profiles of glucose solutions, experimental log P values, and a series of conformational properties of homo-oligomeric chains. The compatibility of our newly proposed CG model with the Martini 3.0 CG force field arises from adopting a similar parameterization strategy and using the bead definition introduced for Martini 3.0. In this context, we also tested our model in the system where carbohydrates can interact with other biomolecules, such as lipid bilayers and proteins. In the context of sugar−lipid bilayer interactions, our model correctly predicts the stabilizing effect of the concentrated solution of glucose on the bilayer structure, protecting it from transition to gel phase. In the case of interactions with proteins, the model is capable of correctly identifying the binding pocket of carbohydrate-binding proteins when the binding occurs via both polar and CH-π interactions. However, it seems that the model has limited capabilities to identify the binding pockets in system where carbohydrate binding is driven primarily through CH-π interactions. The final model may serve as a convenient starting point to further extensions, involving either glucopyranose-based functionalized saccharides or other group of saccharides exploiting different residues and glycosidic linkage types. Developing and validating such extensions will be the subject of our future investigations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00553.

Python script (*carbo2martini3.py*) allowing to generate primitive starting geometries and force field files (in the GROMACS format) for various homopolymers of glucopyranose (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Wojciech Plazinski − *Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, 30-239 Krakow, Poland; Department of Biopharmacy, Medical University of Lublin, 20-093 Lublin, Poland;* orcid.org/0000-0003-1427-8188; Email: wojtek_plazinski@o2.pl

Authors

Valery Lutsyk − *Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, 30-239 Krakow, Poland;* orcid.org/0000-0003-3112-8560

Pawel Wolski − *Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, 30-239 Krakow, Poland;* orcid.org/0000-0002-5970-4542

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00553

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was funded by the National Science Centre, Poland (contract financed in 2020−2024 under no. 2019/35/B/ST4/01149 OPUS 18).

■ REFERENCES

(1) Delbianco, M.; Bharate, P.; Varela-Aramburu, S.; Seeberger, P. H. Carbohydrates in Supramolecular Chemistry. *Chem. Rev.* **2016**, *116*, 1693−1752.

(2) Boysen, M. M. K. Carbohydrates as Synthetic Tools in Organic Chemistry. *Chem. − Eur. J.* **2007**, *13*, 8648−8659.

(3) Smith, B. A. H.; Bertozzi, C. R. Author Correction: The Clinical Impact of Glycobiology: Targeting Selectins, Siglecs and Mammalian Glycans. *Nat. Rev. Drug Discovery* **2021**, *20*, 244−244.

(4) Dwek, R. A. Glycobiology: Moving into the Mainstream. *Cell* **2009**, *137*, 1175−1176.

(5) Phan, D.-N.; Khan, M. Q.; Nguyen, N.-T.; Phan, T.-T.; Ullah, A.; Khatri, M.; Kien, N. N.; Kim, I.-S. A Review on the Fabrication of Several Carbohydrate Polymers into Nanofibrous Structures Using Electrospinning for Removal of Metal Ions and Dyes. *Carbohydr. Polym.* **2021**, *252*, No. 117175.

(6) Sharma, K. Carbohydrate-to-Hydrogen Production Technologies: A Mini-Review. *Renew. Sustainable Energy Rev.* **2019**, *105*, 138−143.

(7) de Wit, D.; Maat, L.; Kieboom, A. P. G. Carbohydrates as Industrial Raw Materials. *Ind. Crops Prod.* **1993**, *2*, 1−12.

(8) Ma, Z.; Zhu, X. X. Copolymers Containing Carbohydrates and Other Biomolecules: Design, Synthesis and Applications. *J. Mater. Chem. B* **2019**, *7*, 1361−1378.

(9) Su, L.; Feng, Y.; Wei, K.; Xu, X.; Liu, R.; Chen, G. Carbohydrate-Based Macromolecular Biomaterials. *Chem. Rev.* **2021**, *121*, 10950−11029.

(10) Woods, R. J. Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. *Chem. Rev.* **2018**, *118*, 8005−8024.

(11) Lutsyk, V.; Plazinski, W. Conformational Properties of Glycosaminoglycan Disaccharides: A Molecular Dynamics Study. *J. Phys. Chem. B* **2021**, *125*, 10900−10916.

(12) Foley, B. L.; Tessier, M. B.; Woods, R. J. Carbohydrate Force Fields: Carbohydrate Force Fields. *WIREs* **2012**, *2*, 652−697.

(13) Xiong, X.; Chen, Z.; Cossins, B. P.; Xu, Z.; Shao, Q.; Ding, K.; Zhu, W.; Shi, J. Force Fields and Scoring Functions for Carbohydrate Simulation. *Carbohydr. Res.* **2015**, *401*, 73−81.

(14) Liwo, A.; Czaplewski, C.; Sieradzan, A. K.; Lipska, A. G.; Samsonov, S. A.; Murarka, R. K. Theory and Practice of Coarse-Grained Molecular Dynamics of Biologically Important Systems. *Biomolecules* **2021**, *11*, 1347.

(15) Joshi, S. Y.; Deshmukh, S. A. A Review of Advancements in Coarse-Grained Molecular Dynamics Simulations. *Mol. Simul.* **2021**, *47*, 786−803.

(16) Singh, N.; Li, W. Recent Advances in Coarse-Grained Models for Biomolecules and Their Applications. *Int. J. Mol. Sci.* **2019**, *20*, 3774.

(17) Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The Power of Coarse Graining in Biomolecular Simulations: The Power of Coarse Graining in Biomolecular Simulations. *WIREs* **2014**, *4*, 225−248.

(18) Rusu, V. H.; Baron, R.; Lins, R. D. PITOMBA: Parameter Interface for Oligosaccharide Molecules Based on Atoms. *J. Chem. Theory Comput.* **2014**, *10*, 5068−5080.

(19) Wu, Z.; Beltran-Villegas, D. J.; Jayaraman, A. Development of a New Coarse-Grained Model to Simulate Assembly of Cellulose Chains Due to Hydrogen Bonding. *J. Chem. Theory Comput.* **2020**, *16*, 4599−4614.

(20) Mehandzhiyski, A. Y.; Rolland, N.; Garg, M.; Wohlert, J.; Linares, M.; Zozoulenko, I. A Novel Supra Coarse-Grained Model for Cellulose. *Cellulose* 2020, *27* (8), 4221−4234. DOI: 10.1007/s10570-020-03068-y.

(21) Gatsiou, C. A.; Bick, A.; Krokidis, X.; Economou, I. G. Atomistic and Coarse-Grained Simulations of Bulk Amorphous Amylose Above and Below the Glass Transition. *Macromolecules* **2022**, *55*, 2999−3010.
(22) Pieczywek, P. M.; Płaziński, W.; Zdunek, A. Dissipative Particle Dynamics Model of Homogalacturonan Based on Molecular Dynamics Simulations. *Sci. Rep.* **2020**, *10*, 14691.
(23) López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hünenberger, P. H.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195−3210.
(24) Wohlert, J.; Berglund, L. A. A Coarse-Grained Model for Molecular Dynamics Simulations of Native Cellulose. *J. Chem. Theory Comput.* **2011**, *7*, 753−760.
(25) López, C. A.; Sovova, Z.; van Eerden, F. J.; de Vries, A. H.; Marrink, S. J. Martini Force Field Parameters for Glycolipids. *J. Chem. Theory Comput.* **2013**, *9*, 1694−1708.
(26) López, C. A.; Bellesia, G.; Redondo, A.; Langan, P.; Chundawat, S. P. S.; Dale, B. E.; Marrink, S. J.; Gnanakaran, S. MARTINI Coarse-for Crystalline Cellulose Microfibers. *J. Phys. Chem. B* **2015**, *119*, 465−473.
(27) Schmalhorst, P. S.; Deluweit, F.; Scherrers, R.; Heisenberg, C.-P.; Sikora, M. Overcoming the Limitations of the MARTINI Force Field in Simulations of Polysaccharides. *J. Chem. Theory Comput.* **2017**, *13*, 5039−5053.
(28) Moiset, G.; López, C. A.; Bartelds, R.; Syga, L.; Rijpkema, E.; Cukkemane, A.; Baldus, M.; Poolman, B.; Marrink, S. J. Disaccharides Impact the Lateral Organization of Lipid Membranes. *J. Am. Chem. Soc.* **2014**, *136*, 16167−16175.
(29) Shivgan, A. T.; Marzinek, J. K.; Huber, R. G.; Krah, A.; Henchman, R. H.; Matsudaira, P.; Verma, C. S.; Bond, P. J. Extending the Martini Coarse-Grained Force Field to *N* -Glycans. *J. Chem. Inf. Model.* **2020**, *60*, 3864−3883.
(30) Mehandzhiyski, A. Y.; Zozoulenko, I. Computational Microscopy of PEDOT:PSS/Cellulose Composite Paper. *ACS Appl. Energy Mater.* **2019**, *2*, 3568−3577.
(31) López, C. A.; de Vries, A. H.; Marrink, S. J. Computational Microscopy of Cyclodextrin Mediated Cholesterol Extraction from Lipid Model Membranes. *Sci. Rep.* **2013**, *3*, 2071.
(32) Alessandri, R.; Souza, P. C. T.; Thallmair, S.; Melo, M. N.; de Vries, A. H.; Marrink, S. J. Pitfalls of the Martini Model. *J. Chem. Theory Comput.* **2019**, *15*, 5448−5460.
(33) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics. *Nat. Methods* **2021**, *18*, 382−388.
(34) Punt, M. "Sweet" Martini 3 Guidelines for a transferable sugar model in Martini 3, *PhD Thesis* , 2021.
(35) Hansen, H. S.; Hünenberger, P. H. A Reoptimized GROMOS Force Field for Hexopyranose-Based Carbohydrates Accounting for the Relative Free Energies of Ring Conformers, Anomers, Epimers, Hydroxymethyl Rotamers, and Glycosidic Linkage Conformers. *J. Comput. Chem.* **2011**, *32*, 998−1032.
(36) Nester, K.; Gaweda, K.; Plazinski, W. A GROMOS Force Field for Furanose-Based Carbohydrates. *J. Chem. Theory Comput.* **2019**, *15*, 1168−1186.
(37) Bulacu, M.; Goga, N.; Zhao, W.; Rossi, G.; Monticelli, L.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Angle Potentials for Coarse-Grained Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 3282−3292.
(38) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19−25.
(39) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(40) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(41) Hockney, R. W. The potential calculation and some applications. *Methods Comput. Phys.* **1970**, *9*, 135−211.
(42) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(43) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. Additive Empirical Force Field for Hexopyranose Monosaccharides. *J. Comput. Chem.* **2008**, *29*, 2543−2564.
(44) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5*, 2353−2370.
(45) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405−413.
(46) Park, S.-J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI *Glycan Modeler* for Modeling and Simulation of Carbohydrates and Glycoconjugates. *Glycobiology* **2019**, *29*, 320−331.
(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(48) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(49) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(50) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(51) Tetko, I. V.; Tanchuk, V. Y. Application of Associative Neural Networks for Prediction of Lipophilicity in ALOGPS 2.1 Program. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1136−1145.
(52) Mazzobre, M. F.; Román, M. V.; Mourelle, A. F.; Corti, H. R. Octanol−Water Partition Coefficient of Glucose, Sucrose, and Trehalose. *Carbohydr. Res.* **2005**, *340*, 1207−1211.
(53) Alessandri, R.; Barnoud, J.; Gertsen, A. S.; Patmanidis, I.; de Vries, A. H.; Souza, P. C. T.; Marrink, S. J. Martini 3 Coarse-Grained Force Field: Small Molecules. *Adv. Theory Simul.* **2022**, *5*, No. 2100391.
(54) Ji, P.; Feng, W.; Tan, T. Density Calculation of Sugar Solutions with the SAFT Model. *J. Chem. Eng. Data* **2007**, *52*, 135−140.
(55) Chuah, C. T.; Sarko, A.; Deslandes, Y.; Marchessault, R. H. Packing Analysis of Carbohydrates and Polysaccharides. Part 14. Triple-Helical Crystalline Structure of Curdlan and Paramylon Hydrates. *Macromolecules* **1983**, *16*, 1375−1382.
(56) Okuyama, K.; Otsubo, A.; Fukuzawa, Y.; Ozawa, M.; Harada, T.; Kasai, N. Single-Helical Structure of Native Curdlan and Its Aggregation State. *J. Carbohydr. Chem.* **1991**, *10*, 645−656.
(57) Hinrichs, W.; Büttner, G.; Steifa, M.; Betzel, C.; Zabel, V.; Pfannemüller, B.; Saenger, W. An Amylose Antiparallel Double Helix at Atomic Resolution. *Science* **1987**, *238*, 205−208.
(58) Zhang, Q.; Lu, Z.; Hu, H.; Yang, W.; Marszalek, P. E. Direct Detection of the Formation of V-Amylose Helix by Single Molecule Force Spectroscopy. *J. Am. Chem. Soc.* **2006**, *128*, 9387−9393.
(59) Zhang, L.; Ruan, D.; Zhou, J. Structure and Properties of Regenerated Cellulose Films Prepared from Cotton Linters in NaOH/Urea Aqueous Solution. *Ind. Eng. Chem. Res.* **2001**, *40*, 5923−5928.

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107

(60) Plazinski, W.; Lonardi, A.; Hünenberger, P. H. Revision of the GROMOS 56A6 $_{CARBO}$ Force Field: Improving the Description of Ring-Conformational Equilibria in Hexopyranose-Based Carbohydrates Chains: Revision of the GROMOS 56A6 $_{CARBO}$ Force Field. *J. Comput. Chem.* **2016**, *37*, 354−365.
(61) Pylkkänen, R., Mohammadi, P., Liljeström, V., Plazinski, W., Beaune, G., Timonen, J.V.I., Wagermaier, W., Penttilä, M. *β*-1,3-glucan Synthesis, Novel Supramolecular Self-Assembly, Characterization and Application, *Submitted for Publication*.
(62) Dashnau, J. L.; Sharp, K. A.; Vanderkooi, J. M. Carbohydrate Intramolecular Hydrogen Bonding Cooperativity and Its Effect on Water Structure. *J. Phys. Chem. B* **2005**, *109*, 24152−24159.
(63) You, X.; Lee, E.; Xu, C.; Baiz, C. R. Molecular Mechanism of Cell Membrane Protection by Sugars: A Study of Interfacial H-Bond Networks. *J. Phys. Chem. Lett.* **2021**, *12*, 9602−9607.
(64) Pereira, C. S.; Hünenberger, P. H. Effect of Trehalose on a Phospholipid Membrane under Mechanical Stress. *Biophys. J.* **2008**, *95*, 3525−3534.
(65) Kapla, J.; Wohlert, J.; Stevensson, B.; Engström, O.; Widmalm, G.; Maliniak, A. Molecular Dynamics Simulations of Membrane−Sugar Interactions. *J. Phys. Chem. B* **2013**, *117*, 6667−6673.
(66) Andersen, H. D.; Wang, C.; Arleth, L.; Peters, G. H.; Westh, P. Reconciliation of Opposing Views on Membrane−Sugar Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 1874−1878.
(67) Souza, P. C. T.; Thallmair, S.; Conflitti, P.; Ramírez-Palacios, C.; Alessandri, R.; Raniolo, S.; Limongelli, V.; Marrink, S. J. Protein−Ligand Binding with the Coarse-Grained Martini Model. *Nat. Commun.* **2020**, *11*, 3714.
(68) Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. Carbohydrate−Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **2015**, *137*, 15152−15160.
(69) Hsu, C.-H.; Park, S.; Mortenson, D. E.; Foley, B. L.; Wang, X.; Woods, R. J.; Case, D. A.; Powers, E. T.; Wong, C.-H.; Dyson, H. J.; Kelly, J. W. The Dependence of Carbohydrate−Aromatic Interaction Strengths on the Structure of the Carbohydrate. *J. Am. Chem. Soc.* **2016**, *138*, 7636−7648.
(70) Plazinska, A.; Plazinski, W. Comparison of Carbohydrate Force Fields in Molecular Dynamics Simulations of Protein−Carbohydrate Complexes. *J. Chem. Theory Comput.* **2021**, *17*, 2575−2585.
(71) Plazinski, W.; Plazinska, A. Molecular Dynamics Simulations of Hexopyranose Ring Distortion in Different Force Fields. *Pure Appl. Chem.* **2017**, *89*, 1283−1294.

https://doi.org/10.1021/acs.jctc.2c00553
J. Chem. Theory Comput. 2022, 18, 5089−5107