Published in final edited form as:

J Phys Chem B. 2010 October 14; 114(40): 12981–12994. doi:10.1021/jp105758h.

CHARMM Additive All-Atom Force Field for Glycosidic Linkages in Carbohydrates Involving Furanoses

E. Prabhu Raman¹, Olgun Guvench², and Alexander D. MacKerell Jr.^{1,*}

¹ Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, 20 Penn Street HSF II, Baltimore MD 21201

² Department of Pharmaceutical Sciences, University of New England College of Pharmacy, 716 Stevens Ave, Portland ME 04103

Abstract

Presented is an extension of the CHARMM additive carbohydrate all-atom force field to enable modeling of polysaccharides containing furanose sugars. The new force field parameters encompass 1 $\leftrightarrow$ 2, 1 $\rightarrow$ 3, 1 $\rightarrow$ 4 and 1 $\rightarrow$ 6 pyranose-furanose linkages and 2 $\rightarrow$ 1 and 2 $\rightarrow$ 6 furanose-furanose linkages, building on existing hexopyranose and furanose monosaccharide parameters. The model compounds were chosen to be monomers or glycosidic-linked dimers of tetrahydropyran (THP) and tetrahydrofuran (THF) as to contain the key atoms in full carbohydrates. Target data for optimization included two-dimensional quantum mechanical (QM) potential energy scans of the Φ/Ψ glycosidic dihedral angles, with geometry optimization at the MP2/6-31G(d) level followed by MP2/cc-pVTZ single point energies. All possible chiralities of the model compounds at the linkage carbons were considered, and, for each geometry, the THF ring was constrained to the favorable South or North conformation. Target data also included QM vibrational frequencies and pair interaction energies and distances with water molecules. Force field validation included comparison of computed crystal properties, aqueous solution densities and NMR *J*-coupling constants to experimental reference values. Simulations of infinite crystals showed good agreement with experimental values for intramolecular geometries as well as for crystal unit cell parameters. Additionally, aqueous solution densities and available NMR data were reproduced to a high degree of accuracy, thus validating the hierarchically optimized parameters in both crystalline and aqueous condensed phases. The newly developed parameters allow for the modeling of linear, branched, and cyclic pyranose/furanose polysaccharides both alone and in heterogeneous systems including proteins, nucleic acids and/or lipids when combined with existing additive CHARMM biomolecular force fields.

Introduction

Carbohydrates play numerous roles in biology with implications for nutrition, medicine, infection and immunity among many other subjects. Crucial to their function is the flexibility of polysaccharide chains about the glycosidic linkage, which enables them to have a broad conformational distribution¹. This work focuses on furanose sugar-containing polysaccharides, the most common of which is the disaccharide sucrose, which consists of

* Corresponding author: Alexander D. MacKerell, Jr., Room 629, HSF II, 20 Penn Street, Baltimore, MD 21201, Tel: 410-706-7442, Fax: 410-706-5017, amackere@rx.umaryland.edu.

Supporting Information **Available**: Supporting information includes additional methodological details, 5 figures and 6 tables, including the topology and parameter information which is also available at the MacKerell group web page at http://mackerell.umaryland.edu/CHARMM_ff_params.html. This information is available free of charge via the Internet at <http://pubs.acs.org>.

the monosaccharides glucose and fructose connected by a glycosidic linkage. In addition to being a major economic commodity, sucrose has been used as a model system to understand the conformational preferences of carbohydrates and for method development towards that goal^{2,3}. Sucrose has been found to bind non-covalently to several proteins, and crystal structures of these reveal diverse conformations of the sugar around the glycosidic linkage (see Venable et. al.⁴ and references therein). Other examples of furanose-containing polysaccharides include raffinose, lactulose, melezitose, 1-kestose, 6-kestose, isomaltulose, planteose, nystose and arabinoxylan. One class of furanose-furanose linked polysaccharides are fructans. These carbohydrates, which serve as an alternative to starch for carbohydrate storage in plants, are exclusively composed of D-fructofuranosyl residues joined by β -2 $\rightarrow$ 6 and β -2 $\rightarrow$ 1 linkages and are one of the most highly distributed biopolymers in nature⁵. Levans, a sub-class of fructans, are composed of D-fructofuranosyl residues joined by β -2 $\rightarrow$ 6 linkages with multiple β -2 $\rightarrow$ 1 branches and are used in the pharmaceutical (ie. in drug delivery formulation) and chemical industries, and have also been indicated to have anti-tumor activity⁶. Due to the various roles carbohydrates play in recognition phenomena, understanding their conformational properties as well as their interactions with proteins and other biomolecules is of scientific interest and may be of value, for instance, in carbohydrate-based drugs⁷. Molecular mechanics based methods offer excellent tools to probe both carbohydrate conformational properties and intermolecular interactions; however, to be of value, such methods require accurate force field models of carbohydrates.

Significant efforts have gone into the development of molecular mechanics force fields for carbohydrates⁸⁻¹¹. However, relatively few parameter sets exist for pyranose-furanose and furanose-furanose glycosidic linkages¹². Building on existing hexopyranose¹³, aldopentofuranose and fructofuranose¹⁴ monosaccharide parameters recently developed in our lab, and following methodology used to develop parameters for pyranose-pyranose glycosidic linkages¹⁵, we present CHARMM additive force field parameters for an extensive set of linkage types involving pyranose and furanose monosaccharides. These include 1 $\leftrightarrow$ 2, 1 $\rightarrow$ 3, 1 $\rightarrow$ 4 and 1 $\rightarrow$ 6 pyranose-furanose linkages and 2 $\rightarrow$ 1 and 2 $\rightarrow$ 6 furanose-furanose linkages with all possible combinations of chiralities for the two ring carbons in each glycosidic link (carbon numberings are per the IUPAC standard for the hexopyranose glucose and the hexofuranose fructose). Parameters are also developed purely based on QM data for 1 $\rightarrow$ 2 and 1 $\rightarrow$ 3 furanose-pyranose linkages as seen in arabinoxylan (carbon numberings are per the IUPAC standard for the pentofuranose arabinose and the hexopyranose glucose) due to the lack of available condensed phase experimental data for validation. In combination with the recently developed hexopyranose disaccharide parameters¹⁵, these new parameters allow the modeling of linear, cyclic and branched complex polysaccharides composed of both pyranose and furanose monosaccharides. Furthermore, as the parameters have been developed using the same protocol as for the CHARMM additive force fields for proteins,¹⁶ nucleic acids,¹⁷ lipids¹⁸ and pharmaceuticals¹⁹, they can be used for simulating carbohydrate-containing systems that include any or all of these other classes of molecules.

Our parametrization strategy involves the use of high-level QM calculations of simple model compounds representative of disaccharides with the resulting data used as target data for the optimization of parameters associated with the glycosidic linkage. Specifically, glycosidic heterodimers of tetrahydropyran (THP), tetrahydrofuran (THF) and cyclohexane are chosen that contain the key ring carbon and oxygen atoms of six and five-membered cyclic monosaccharides, but not the hydroxyls (Figure 1). Optimized two-dimensional (2D) potential energy scans at the MP2/6-31G(d) level for the full range of the Φ/Ψ dihedral angles of a glycosidic linkage are performed, and MP2/cc-pVTZ single point energies are computed for these geometry-optimized conformations ("MP2/cc-pVTZ//MP2/6-31G(d)"). Due to the inherent conformational flexibility of five-membered rings²⁰, two sets of QM

calculations are performed with the ring conformation of THF constrained to geometries representative of either the favorable North or South conformation populated by furanose sugars^{21,22}. In total, over 12600 MP2/cc-pVTZ//MP2/6-31G(d) conformational energies are used as target data in the parametrization process, which encompasses all types and chiralities of linkages. Also included in the target data are interaction energies and distances with water at the HF/6-31(d) level for charge optimization of the glycosidic linkage oxygen atom and vibrational frequencies at the MP2/6-31G(d) level. The hierarchically developed parameters are applied to full di/tri/tetra-saccharides and validated by their ability to reproduce conformational and bulk properties in the crystalline and aqueous phases.

Methods

All molecular mechanics calculations were carried out using the CHARMM²³ program and used the same potential energy function as in the CHARMM protein¹⁶, nucleic acid¹⁷ and lipid²⁴ all-atom additive force fields.

$$\begin{aligned}
 U(r) = & \sum_{\text{bonds}} K_b (b - b_0)^2 \\
 & + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 \\
 & + \sum_{\text{Urey-Bradley}} K_{UB} (S - S_0)^2 \\
 & + \sum_{\text{dihedrals}} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{\text{impropers}} K_{imp} (\phi - \phi_0) + \sum_{\text{nonbonded}} \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}}.
 \end{aligned} \tag{1}$$

In Equation 1, K_b , K_θ , K_{UB} , K_χ and K_{imp} are bond, valence angle, Urey-Bradley, dihedral angle, and improper dihedral angle force constants, respectively. b , θ , S , χ , and ϕ are the bond distance, valence angle, Urey-Bradley 1,3-distance, dihedral angle, and improper dihedral angle values, and the subscript 0 represents an equilibrium value. For the dihedral term, n is the multiplicity and δ is the phase angle. The nonbonded interaction energy between pairs of atoms i and j consists of the Lennard-Jones (LJ) 6-12 term and the Coulomb term. $\bullet_{ij}$ is the LJ well depth, $R_{\min,ij}$ is the interatomic distance at the LJ energy minimum, q_i and q_j are the partial atomic charges, and r_{ij} is the distance between atoms i and j . The Lorentz-Berthelot combining rules determine LJ parameters between different atom types²⁵.

Water was represented using a rigid three-site modified TIP3P model^{26,27}. Dynamics were performed with the SHAKE algorithm,²⁸ which constrains covalent bonds involving hydrogen atoms to their equilibrium value and was used to keep water molecules rigid through the introduction of a virtual bond between the hydrogen atoms in a water molecule. Energy calculations in gas phase were performed with an infinite non-bonded cutoff while simulations in aqueous phase and crystal environments were performed using periodic boundary conditions²⁵. LJ interactions in these simulations employed a force-switch smoothing function in the range of 10 to 12 Å with a long range correction to the energy and pressure applied to account for LJ interactions beyond 12 Å²⁵. Particle mesh Ewald²⁹ with a real-space cutoff of 12 Å was used to treat long-range electrostatics. The “leapfrog” integrator³⁰ with a 1 fs timestep was used in molecular dynamics (MD) simulations. Experimental temperature and pressure were used in the crystal and aqueous phase simulations and were maintained by a Nose-Hoover thermostat^{31,32} and a Langevin piston barostat³³, respectively. Aqueous phase simulations were done with a periodic cubic unit cell with edge lengths allowed to vary isotropically. Crystal simulations used the

experimental unit cell geometries where the edge lengths were allowed to vary independently. Unit cell lattice angles of 90° were constrained whereas others were allowed to vary independently thereby maintaining the unit cell symmetry.

All QM calculations were performed using Gaussian 03 program³⁴. Geometry optimizations and vibrational calculations were performed using MP2/6-31G(d) model chemistry^{35,36}, with tight tolerances applied when optimizing structures for vibrational calculations. QM frequencies were scaled by a factor of 0.9434 in order to account for limitations in the level of theory to reproduce experimental frequencies³⁷. Potential energy decomposition analysis was performed using the MOLVIB module in CHARMM using internal coordinates³⁸. Relaxed potential energy scans were performed by geometry optimization at the MP2/6-31G(d) level followed by MP2/cc-pVTZ single point calculations (MP2/cc-pVTZ//MP2/6-31G(d))^{35,36,39}. All potential energy scans were performed in 15° increments with the scanned dihedrals constrained. QM calculations of model compound interactions with water were performed at the HF/6-31(d) level following the standard procedure used for the CHARMM additive force-fields⁴⁰. Per this procedure, only the solute:water interaction distance was allowed to vary in the calculation with all other degrees of freedom constrained. The solute was constrained to the gas phase optimized geometry in the QM or the CHARMM representation, and the TIP3P water geometry was used in both MM and QM calculations. The QM energy of the solute:water system E_{pair} was computed following the interaction distance optimization. The target energy used was calculated as $1.16 * (E_{\text{pair}} - E_{\text{solute}} - E_{\text{water}})$, with no basis-set superposition-error correction and an empirical scaling factor of 1.16 introduced to yield parameters appropriate for the condensed phase and compatible with the TIP3P water model^{16,41}. Target data for interaction distances were the QM optimized distances minus 0.2 Å, again to yield parameters appropriate for the condensed phase.

For the optimization of dihedral parameters, relaxed QM potential energy surfaces at the MP2/cc-pVTZ//MP2/6-31G(d) level of theory were used as target data. Analogous adiabatic MM energy surfaces were computed with the to-be-fit dihedral force constant parameters set to zero. The difference energy (MM-QM) was fit using the freely-available Monte Carlo simulated annealing (MCSA) dihedral parameter fitting program `fit_dihedral.py`⁴². In this procedure, three multiplicities n of 1, 2 and 3 were included and the corresponding K_{χ} values (Eqn 1.) were optimized to minimize the root-mean-square error (RMSE) and therefore the difference energy. K_{χ} values were constrained to be no more than 3 kcal/mol, and the phase angles δ were limited to 0 and 180° as required to insure applicability of the parameters to stereoisomers of a chiral center. For multiple stereoisomers of the same molecule, separate RMS alignment of MM surfaces to their corresponding QM surfaces was performed¹⁵, thereby not requiring the reproduction of QM energy difference across different isomers.

The aqueous solution density ρ was calculated from the average volume $\langle V \rangle$, obtained from the MD simulations, using equations 2 and 3:

$$\rho = \frac{N}{\langle V \rangle} \quad (2)$$

$$N = \frac{(N_{\text{water}} + N_{\text{solute}}) * \langle MW \rangle}{N_{\text{Avogadro}}} \quad (3)$$

where N_{water} , N_{solute} , and N_{Avogadro} are the number of water molecules, solute molecules and Avogadro's number, respectively, and $\langle MW \rangle$ is the average molecular weight of the solution. For all aqueous density simulations, $N_{\text{water}} = 1100$, while N_{solute} was adjusted to give the appropriate molality at which the experiment was performed. These simulations were equilibrated for 0.5 ns followed by 5 ns of production run, with the volume of the system calculated every 10ps and averaged over the production simulation time to give $\langle V \rangle$.

NMR heteronuclear three-bond proton-carbon coupling constants for the glycosidic dihedral angle $^3J_{\text{COCH}}$ were computed from the simulations using the modified Karplus equation developed by Tvaroska et al.⁴³

$$^3J_{\text{COCH}} = 5.7 \cos^2 \phi - 0.6 \cos \phi + 0.5, \quad (4)$$

where ϕ is the $^6\text{H}_1\text{-}^6\text{C}_1\text{-O}_{\text{link}}\text{-}^5\text{C}_2$ dihedral angle in a pyranose-furanose disaccharide and where the superscript number 5 or 6 represents five or six ring membership, respectively. $^3J_{\text{H-3f-H-4f}}$ and $^3J_{\text{H-4f-H-5f}}$ associated with the fructofuranose ring pucker geometry were calculated using the Altona and Hasnoot equations⁴⁴

$$^3J_{\text{HCCH}} = 4.35 \cos^2 \phi - 0.91 \cos \phi + 2.34 \sin^2 \phi + 4.83, \quad (5)$$

$$^3J_{\text{HCCH}} = 3.97 \cos^2 \phi - 0.91 \cos \phi - 0.10 \sin^2 \phi + 4.67, \quad (6)$$

where ϕ in equations 5 and 6 is $^5\text{H}_3\text{-}^5\text{C}_3\text{-}^5\text{C}_4\text{-}^5\text{H}_4$ and $^5\text{H}_4\text{-}^5\text{C}_4\text{-}^5\text{C}_5\text{-}^5\text{H}_5$, respectively. To compute J values from MD simulations, a separate J value from each snapshot was calculated and an average of these over the production simulation was compared to the experiment. The simulation system consisted of one sugar molecule in a box of 1100 water molecules, resulting in a concentration of approximately 50 mM and minimizing solute-solute interactions between adjacent periodic images. Simulations were equilibrated for 0.5 ns, followed by 20 ns of production, during which structures were saved every 1 ps for analysis. The convergence of calculated properties was confirmed by performing a second independent set of simulations of equal time length, but with a different random seed number. For kestose, the experimental concentration of 300mM⁴⁵ was significantly higher than 50mM. Therefore, an additional 20ns simulation was performed at the experimental concentration to check for any dependence of conformational properties on concentration.

The conformational distribution of sucrose in vacuum was examined using temperature replica exchange Langevin dynamics simulations⁴⁶. The “leapfrog” integrator³⁰ was used with an integration time step of 2 fs. Eight replicas of a single sucrose molecule were simulated with temperatures exponentially varying between 300K and 500K for 400 ns each. Properties were computed from snapshots saved every 5 ps from the 300 K replica.

Results

The focus of this work was the development of parameters for glycosidic linkages present in polysaccharides containing furanose sugars. This includes pyranose (P) - furanose (F) linkages P1 $\leftrightarrow$ F2, P1 $\rightarrow$ F3, P1 $\rightarrow$ F4, P1 $\rightarrow$ F6 where the numbers in the linkage type notation refer to the IUPAC carbon numbering for the six-carbon monosaccharides glucopyranose and fructofuranose, respectively. Arrows point from the anomeric center, and

double-headed arrows indicate a linkage between two anomeric centers. Two types of furanose-furanose linkages are also covered, namely $F2 \rightarrow F1$ and $F2 \rightarrow F6$. Good agreement with NMR data for a model disaccharide composed of the five-carbon furanose arabinose shows that the parameters can also be used for polysaccharides composed of five-carbon furanoses. Parameters are also presented for $F1 \rightarrow P2$ and $F1 \rightarrow P3$ linkages present in the polysaccharide arabinoxylan (here the numbers associated with the furanose moiety are per IUPAC carbon numbering of arabinose), though without the benefit of MD validation studies like for the other systems due to the absence of condensed phase experimental data. Because of the hierarchical parametrization methodology, in principle, the optimized parameters are also applicable to $P1 \rightarrow F1$ and $F2 \rightarrow P4$ linkages.

Parameter optimization

Available parameters for hexopyranose monosaccharides,¹³ furanose monosaccharides,¹⁴ and associated model compounds were used directly; only those involving the glycosidic linkage oxygen atom (O_{link}) were optimized. Starting values for the bond, valence angle, dihedral angle, Lennard-Jones and partial charge parameters were transferred from those optimized for model compounds used in the above mentioned works and from glycosidic linkages between hexopyranoses¹⁵ previously optimized in our laboratory. The following describes the optimization of the transferred parameters. To develop parameters for the different types of linkages possible between a six- and a five-membered ring, four model compounds (Figure 1) containing an ether-linked pair of tetrahydropyran (THP) and tetrahydrofuran (THF) molecules were used. Model compound 2 (Figure 1e) is used to parametrize pyranose-furanose or furanose-furanose three-bond linkages. Note that in all subsequent atom notations, the atoms in THP are indicated by a superscript “6” and atoms in THF by a superscript “5,” denoting their location in a six- or five-membered ring, respectively. Although the linkage is not between two anomeric carbons, for model compounds 3, 4 and 6 (supplementary information) we still use the α/β nomenclature to denote stereoisomers in order to be consistent with other model compounds.

Parametrization was done in a self-consistent fashion such that whenever one parameter was changed, all properties were first recomputed and then the particular parameter and any other parameters re-optimized as necessary so as to simultaneously fit all the target data. Conformational energies for all compounds were studied using the MP2/cc-pVTZ//MP2/6-31G(d) model chemistry, which provides a reasonable compromise between accuracy and computational expediency. Previously it was seen that energies for hexopyranose monosaccharides using this model chemistry were in excellent agreement compared to MP2/cc-pVTZ optimized structures¹³. Owing to the inherent flexibility of the five-membered ring^{14,20}, the THF ring conformation in all of the model compounds was constrained. Two sets of conformations were optimized for all the dihedral scan points obtained, one in which THF dihedral angles ${}^5O_{\text{ring}}-{}^5C_2-{}^5C_3-{}^5C_4$ and ${}^5C_2-{}^5C_3-{}^5C_4-{}^5C_5$ were constrained to $40^\circ/-35^\circ$ and the second to $-40^\circ/35^\circ$, to approximate the South and North conformational preferences of aldopentofuranoses and ketohexofuranoses¹⁴. It was important to constrain the ring pucker geometry in order to avoid Φ/Ψ dihedral fitting being influenced by ring energetics. The exact location of the energy minimum in the ring pucker landscape varies between different furanose sugars and their anomers¹⁴. These values of the ring dihedral constraints were chosen as they are representative of preferred ring conformations for aldopentofuranoses and ketohexofuranoses, and constraints were applied for all model compounds involving tetrahydrofuran (THF). It should be noted that when optimized, the ring dihedral angles of THF with O-methylation at the C2 position acquire values of $40^\circ/-35^\circ$ and $-40^\circ/35^\circ$ for the α and β anomers, respectively. Ring dihedral parameters for the full sugars (with hydroxyls) were recently optimized in the context of

aldopentofuranoses and ketohexofuranoses¹⁴ and therefore were transferred to the di/poly-saccharides used for validation, as detailed below in the Validation section.

Model Compound 1—Model compound 1 involves a glycosidic linkage between ${}^6\text{C}_1$ (of THP) and ${}^5\text{C}_2$ (of THF) and was used to parametrize $1 \leftrightarrow 2$ linkages between pyranose and furanose monosaccharides. QM MP2/cc-pVTZ//MP2/6-31G(d) two-dimensional dihedral scans of ${}^6\text{O}_{\text{ring}}\text{-}{}^6\text{C}_1\text{-O}_{\text{link}}\text{-}{}^5\text{C}_2$ (Φ) / ${}^6\text{C}_1\text{-O}_{\text{link}}\text{-}{}^5\text{C}_2\text{-}{}^5\text{O}_{\text{ring}}$ (Ψ) were computed for the four possible anomers ($\alpha\alpha$, $\alpha\beta$, $\beta\alpha$ and $\beta\beta$, Figure 1a-d) in two frozen THF conformations as described above with 15° increments, yielding $4 \times 2 \times 24 \times 24 = 4608$ conformations. An initial set of parameters was developed by transferring analogous parameters from the $1 \leftrightarrow 1$ linked hexopyranose disaccharide model compound THP-O-THP¹⁵ and O-methyl THF¹⁴, leaving ${}^6\text{O}_{\text{ring}}\text{-}{}^6\text{C}_1\text{-O}_{\text{link}}\text{-}{}^5\text{C}_2$, ${}^6\text{C}_2\text{-}{}^6\text{C}_1\text{-O}_{\text{link}}\text{-}{}^5\text{C}_2$, ${}^6\text{C}_1\text{-O}_{\text{link}}\text{-}{}^5\text{C}_2\text{-}{}^5\text{O}_{\text{ring}}$ and ${}^6\text{C}_1\text{-O}_{\text{link}}\text{-}{}^5\text{C}_2\text{-}{}^5\text{C}_3$ to be optimized using a first round of MCSA fitting. Only conformations with relative energies < 12 kcal/mol were considered for fitting and for each anomer, a separate RMS alignment of the MM and QM surfaces was performed. The transferred ring pucker parameters underestimated the QM energy difference between the two pucker conformations by about 3 kcal/mol. As detailed in Supplementary materials, the associated dihedrals were re-parametrized in the context of the simpler model compound 5 (Figures S1a and S2) and the optimized parameters were transferred to model compound 1. Following this, a second round of MCSA fitting was applied to the 2-D Φ/Ψ scan for fitting the glycosidic dihedral parameters. Using these as initial parameters, the four anomers were fully geometry-optimized in the QM MP2/6-31G(d) and MM representations starting from their respective Φ/Ψ QM minima. Errors in ${}^6\text{C}_1\text{-O}_{\text{link}}$ bond were corrected by an adjustment of the equilibrium bond length value to 1.425 from 1.395 Å (also validated by sucrose crystal simulations described below). The QM vibrational frequencies were well reproduced by the parameters (Supplementary information Table S1). In an effort to improve the fit, low energy points (< 6 kcal/mol) were given 5 times more weight than the higher ones in the MCSA procedure. The Φ/Ψ parameters were refit following the adjustment of the bond length parameter. The RMSE value, calculated over energy points < 12 kcal/mol, improved from 3.00 to 1.27 kcal/mol. Figure 2 shows that the fit MM parameters capture the Φ/Ψ energy surface to a high degree of accuracy. In order to be able to use the same set of parameters for all anomers, it was necessary to slightly overestimate the $90^\circ/60^\circ$ local minimum in the $\alpha\beta$ South conformations (Figure 2e,f) and underestimate the $-90^\circ/60^\circ$ local minimum in the $\beta\alpha$ South conformations (Figure 2i,j). Table 1 shows that the fit parameters reproduce the QM optimized geometries for the four anomers. For the $\alpha\beta$ anomer a discrepancy of nearly 10° in Ψ dihedral is observed. This can be explained by the relatively flat region of the energy landscape along $\Phi \approx 60$, which is present in QM and captured by the fit-MM energy landscapes (Figure 2g,h). The QM energy varies by less than 0.2 kcal/mol in the range $-105 < \Psi < -90$. Also, the difference in MM energy of the structures optimized in the MM and QM representations (Ψ -106.6 vs -95.7) is only 0.06 kcal/mol. Thus, the discrepancy in the location of the Ψ minimum is primarily due to the flatness of the energy landscape rather than due to a significant limitation in the fit parameters. Finally, the appropriateness of transferred partial charge and Lennard-Jones parameters for the O_{link} atom was evaluated by pair interaction energy calculations with water. Table 2 shows that for all minimum-energy conformations of the anomers, the interaction energies as well as distances are well reproduced by the transferred non-bonded parameters.

Model Compound 2—Model compound 2 (Figure 1e) was included to cover linkages at positions 1 or 6 of the hexofuranose ring. Pyranose-furanose or furanose-furanose linkages at these positions involve three rotatable bonds. As a result, a potential energy scan involving two backbone rings (THP-THP or THP-THF) would be computationally prohibitive. Thus, a model compound involving only THF was used. As the α and β anomers

are structurally equivalent, only one was needed. For the α anomer, two sets of 2D QM MP2/cc-pVTZ//MP2/6-31G(d) scans about $^5\text{O}_{\text{ring}}\text{-}^5\text{C}_2\text{-}^5\text{C}_1\text{-O}_{\text{methyl}}/^5\text{C}_2\text{-}^5\text{C}_1\text{-O}_{\text{methyl}}\text{-C}_{\text{methyl}}$ (Φ/Ψ) were obtained with the THF ring constrained to either the South or North conformation during the 2D scan; ring dihedral parameters specially optimized for THF in the context of compound 2 were used to reproduce the relative energies of the South and North conformations (Supplementary materials, Figure S3). Valence angles $^5\text{O}_{\text{ring}}\text{-}^5\text{C}_2\text{-}^5\text{C}_1$ and $^5\text{C}_3\text{-}^5\text{C}_2\text{-}^5\text{C}_1$ were transferred from an analogous THP-based model compound previously employed in hexopyranose disaccharide parametrization consisting of the same exocyclic substituent linked to THP at the C₅ hexopyranose position (see model compound 11 in Ref 15). An initial set of parameters for the Φ/Ψ dihedrals were also transferred from this previous THP-based model compound and were used for an initial geometry optimization of model compound 2.

Overestimation of the $^5\text{C}_2\text{-}^5\text{C}_1$ bond length was corrected by shortening its equilibrium value to 1.500 Å. To correct errors in the valence angle vibrational frequency for $^5\text{O}_{\text{ring}}\text{-}^5\text{C}_2\text{-}^5\text{C}_1$ and geometry for $^5\text{C}_3\text{-}^5\text{C}_2\text{-}^5\text{C}_1$, the equilibrium value of $^5\text{C}_3\text{-}^5\text{C}_2\text{-}^5\text{C}_1$ was reduced by 0.5° to 113° and the force constant of $^5\text{O}_{\text{ring}}\text{-}^5\text{C}_2\text{-}^5\text{C}_1$ was increased from the transferred value of 45 kcal/mol/rad² to 60 kcal/mol/rad². Following these changes, the MCSA procedure was applied for dihedral fitting. It was observed that using the newly fit parameters resulted in nearly the same RMSE as the transferred parameters. This is likely due to the similarity of chemical environment of the exocyclic substituent $\text{-CH}_2\text{-O}_{\text{link}}\text{-CH}_3$ in THP (attached to C₅) and THF (attached to C₂). As the fit dihedral parameters do not provide any advantage over those that were transferred, in the interest of generality, the latter were selected for use. Using these transferred dihedral parameters yields an RMSE value of 0.72, as compared to 2.46 kcal/mol with the force constants for these parameters set to 0. Figure 3 shows that for model compound 2, the MM parameters reproduce even fine details of the QM energy surface. Table 3 shows that unrestrained optimized geometries in the QM and MM representations are very similar. The partial charge on the O_{link} atom was validated by water interaction calculation. Water interaction energies and distances shown in Table 2 are well reproduced by the transferred partial charge of -0.36. Finally, empirical vibrational frequencies compare favorably to reference QM values (Table S2 Supporting Information).

Model Compound 3—Model compound 3 (Figure 1f,g,h,i) was included to cover P1 → F3, P1 → F4 linkages. QM MP2/cc-pVTZ//MP2/6-31G(d) 2D dihedral scans of $^6\text{O}_{\text{ring}}\text{-}^6\text{C}_1\text{-O}_{\text{link}}\text{-}^5\text{C}_3$ (Φ) / $^6\text{C}_1\text{-O}_{\text{link}}\text{-}^5\text{C}_3\text{-}^5\text{C}_2$ (Ψ) were computed for the four possible stereoisomers ($\alpha\alpha$, $\alpha\beta$, $\beta\alpha$ and $\beta\beta$) in two frozen THF conformations as described above with 15° increments, yielding $4 \times 2 \times 24 \times 24 = 4608$ conformations. Parameters for the O-linkage at the C₃ position of THF were not available. Therefore, model compound 6 (Figure S1b) was created for the purpose of generating $\text{O}_{\text{ring}}\text{-C}_2\text{-C}_3\text{-O}_{\text{methyl}}$ pucker dihedral parameters for transfer to the full model compound 3; toward this goal, the $\text{C}_{2/4}\text{-C}_3\text{-O}_{\text{methyl}}\text{-C}_{\text{methyl}}$ dihedrals in model compound 6, also lacking parameters, were also parametrized (see Supporting Information, Figure S4). The optimized pucker parameters were then transferred to model compound 3. A first round of MCSA optimization was undertaken yielding Φ/Ψ dihedral parameters. Using these, all 4 stereoisomers were minimized starting from the QM Φ/Ψ surface minimum energy geometries in both QM and MM representations and compared. A discrepancy in the $^6\text{C}_1\text{-O}_{\text{link}}\text{-}^5\text{C}_3$ valence angle led to adjustment of the associated equilibrium parameter to 109.5° (decreased by 2° from a value of 111.5° in the transferred pyranose disaccharide parameter). Associated bond angle constants were validated by vibrational frequency analysis on model compound 3 (see Supporting Information Table S3). Following the parameter update, a second round of MCSA fitting with separate RMS alignment for each stereoisomer was undertaken and Φ/Ψ dihedral parameters refit. Figure 4 shows that the fit MM parameters reproduce the QM energy

surface well for all stereoisomers for both ring pucker geometries. The RMSE improves from 1.38 to 1.06 kcal/mol after fitting. Table 4 shows that the optimized parameters reproduce the QM geometries well. In the case of $\alpha\beta$ and $\beta\alpha$ isomers, an error of about 15° is seen in the ring-dihedral minimum angle. However, the same dihedral geometry is reproduced to very high accuracy in the $\alpha\alpha$ and $\beta\beta$. Importantly, the error in ring minima geometries was not considered to be problematic because in the case of full sugars the furanose ring geometry is governed by parameters that have been explicitly optimized in the context of hydroxyls. Table 2 shows that for the QM optimized conformation, the interaction energies as well as distances are well reproduced by the transferred partial charge of -0.36.

Model Compound 4—Model compound 4 (Figure 1j,k) was included to cover the P2/3/4 $\rightarrow$ F2/5 linkages. QM MP2/cc-pVTZ//MP2/6-31G(d) 2D dihedral scans of ${}^6\text{C}_5\text{-}{}^6\text{C}_1\text{-O}_{\text{link}}\text{-}{}^5\text{C}_2$ (Φ) / ${}^6\text{C}_1\text{-O}_{\text{link}}\text{-}{}^5\text{C}_3\text{-}{}^5\text{C}_2$ (Ψ) were computed for only two of the four possible stereoisomers ($\alpha\beta$, $\beta\alpha$) in two frozen THF conformations as described above with 15° increments, yielding $2 \times 2 \times 24 \times 24 = 2304$ conformations. Including only two isomers suffices as cyclohexane is a symmetric molecule about the glycosidic bond. As a result, the dihedral scans for $\alpha\alpha$ and $\beta\beta$ anomers would be symmetric with respect to $\alpha\beta$ and $\beta\alpha$, respectively. The optimized pucker parameters from model compound 5 were transferred and the Φ/Ψ dihedral parameters were fit using the MCSA procedure with separate RMS alignment for each isomer (Figure 5). RMSE improves, but only marginally (1.56 from 1.70 kcal/mol) with the fitting. Table 5 shows that the optimized parameters reproduce well the QM optimized geometry for the two anomers. For the $\alpha\beta$ anomer a discrepancy of 9.3° is observed in the Φ dihedral, which can again be explained by the relatively flat energy landscape that is observed in the interval $-150 < \Phi < -60$. This feature of the QM landscape is captured by MM-fit parameters as seen in Figure 5c,d. Table 2 shows that for the QM optimized conformations of the two anomers, the interaction energies as well as distances are well reproduced by the transferred partial charge of -0.36.

Validation

Molecular dynamics (MD) simulations both in crystalline and aqueous phases were performed to validate the parameters optimized using QM target data. The parameters involved in the glycosidic linkages were transferred from model compounds 1-4. The remaining parameters involved in the component monosaccharide hexopyranoses¹³ and fructofuranoses¹⁴ in the di/poly-saccharides were used without modification. The furanose ring pucker dihedral parameters that involve the glycosidic oxygen (e.g. the $\text{O}_{\text{ring}}\text{-C}_2\text{-C}_3\text{-O}_{\text{methyl}}$ dihedral parameters in compound 6) were transferred directly from the analogous hydroxyl parameter for use in the full disaccharides. In addition, a second parallel set of validation calculations was attempted with the pucker parameters being transferred from model compounds 5 and 6 (results not shown). Both sets gave nearly equally good agreement with crystal data, but the former reproduced NMR *J*-coupling constants more accurately than the latter. This finding was not surprising and can be explained by the following two reasons. First, as the pucker dihedral parameters used in the former set were optimized in the context of full sugars including the presence of all hydroxyl substituents in furanoses (as opposed to model compounds 2, 5 and 6), it is expected that they will better capture the conformational properties in full disaccharides. Second, the target data used in the former were QM optimized geometries of tens of conformations obtained by rotation around the ring dihedrals whereas the latter used frozen ring geometry in only two conformations (South and North). In the following discussion, we therefore only present results obtained from the former set of parameters.

Crystalline intramolecular geometries and unit cell parameters—Crystal simulations were used to check the ability of the parameters to reproduce intramolecular

geometries, thus testing bond, angle and dihedral parameters. Additionally, the reproduction of the crystal cell volumes and unit cell parameters by the force field provided a way to validate the non-bonded parameters. The choice of di, tri and tetra-saccharides for crystal simulations was mainly driven by their availability in the Cambridge Structural Database (CSD) of small molecules⁴⁷ and the desire to test all distinct linkage types parametrized. Thus, any sugar crystals containing any of the parametrized linkages was selected from the CSD. A total of eight compounds were simulated as infinite crystals using MD at constant temperature and pressure. Table 6 lists them, the linkages they cover and the numbers of sugar and water molecules in the simulation unit cell. The majority of sugar crystals containing the parametrized linkages were composed of α -D-glucopyranose and β -D-fructofuranose with the exception of raffinose, which contains α -D-galactopyranose in addition. This limited selection reflects partly the limited kinds of monosaccharides in pyranose-furanose and furanose-furanose linked sugars found in nature. Nonetheless, the test set contains all the different types of linkages parameterized using model compounds 1-3. Most simulated unit cells contained four independent sugar molecules and many contained water molecules, which further tested the robustness of the force field.

Table 7 shows the computed unit cell parameters A , B , C and volumes averaged over the 2 ns simulation time. Overall the parameters reproduce crystal geometries to a satisfactory level of accuracy, with the average volume error being 5.5 % and the average errors in the unit cell parameters A , B and C being 2.4, 0.6 and 2.5 %, respectively. The systematic overestimation of unit cell volumes has been observed also in prior works^{13,48}. Since the hydroxyl nonbonded parameters have been optimized from simulations of neat liquid alcohols, it leads to difficulty in representing the highly directional hydrogen bonding interactions seen in the crystalline phase. Therefore, as expected, there is much better agreement in the case of aqueous solution densities, as described in the next section. The unit cell angle parameter β in the sucrose crystal was not equal to 90° and allowed to vary during the simulation. This parameter was very well preserved (MD average and crystal values being 102.84 and 102.95, respectively).

The comparison of the MD-averaged intramolecular geometric terms provides a very precise way to judge the accuracy of the force field. For each crystal simulated, the values of each bond, angle and dihedral was averaged over all independent molecules and over the simulation time of 2ns. The difference of this value and the corresponding crystal value yields the error. A first round of test crystal simulations revealed the valence angle 5C_2 - O_{link} - $^5C_{1/6}$ present in KESTOS, CELGIJ and PEKHES crystals to be about 2.5° too small as compared to the crystal geometry. To remedy this, the equilibrium value of the associated angle parameter was increased to 112.2° and properties re-computed. Table 8 shows the average and the standard deviation of errors for bonds, angles and dihedrals associated with the glycosidic linkages. Average errors in bond lengths, angles and dihedrals are less than 0.01 Å, 1.5° and 5°, respectively. Overall, the optimized parameters reproduce well the intramolecular geometries of the eight di/poly-saccharides in their crystal environments.

Due to the absence of crystals in the CSD involving any furanose sugar other than fructofuranose in the context of the linkages covered in this work, crystal validation could not be performed for di/poly-saccharides containing aldopentofuranoses (example: arabinoxylan). However, the parameters were optimized using “backbone” model compounds and thus, due to the hierarchical nature of the present parametrization procedure, it is expected that the optimized parameters will cover sugars containing aldopentofuranoses as well.

Solution densities—As CHARMM is a biomolecular force field²³, the parameters are typically used in the aqueous phase, making this type of validation important. Aqueous

solution density data was available for sucrose, isomaltulose and lactulose, which involve P1 $\leftrightarrow$ F2, P1 $\rightarrow$ F6 and P1 $\rightarrow$ F4 linkages, respectively. Such data provides the opportunity to validate parameters optimized using model compounds 1-3. MD simulations of aqueous solutions of various concentrations of the three disaccharides were performed under the conditions of 298 K and 1 atm. The densities calculated over a production simulation time of 5ns according to equations 2 and 3 compare favorably with the experimental values. Table 9 shows that all errors are within 1.52 % and the average error is 1.05 %. The parameters reproduce well the experiment across a range of solution densities that vary by nearly an order of magnitude. As observed in previous studies^{13,48}, the errors in reproducing aqueous densities are significantly less than the errors in crystal unit cell volumes.

Conformational properties in aqueous solution—Karplus-type equations have previously been developed for ^{13}C - ^1H spin couplings and are used here to relate the experimentally measured three-bond coupling constant values 3J to the glycosidic linkage dihedral angles (see Methods). We use these equations to compute 3J values corresponding to the $^6\text{H}_1$ - $^6\text{C}_1$ - O_{link} - $^5\text{C}_2$ glycosidic linkage dihedral in sucrose and kestose. $^3J_{\text{HCCH}}$ values were computed for both sugars that correspond to the $^5\text{H}_3$ - $^5\text{C}_3$ - $^5\text{C}_4$ - $^5\text{H}_4$ and $^5\text{H}_4$ - $^5\text{C}_4$ - $^5\text{C}_5$ - $^5\text{H}_5$ ring dihedrals in the fructofuranose moieties, which provide information on the pucker geometry. Additionally, we test the ability of the parameters to reproduce the North/South pucker conformational equilibrium by comparing the simulation results of a model arabinose disaccharide to results obtained from PSEUROT⁴⁹ analysis of NMR data.

The choice of compounds was mainly driven by the availability of NMR data. In particular, previously-published J -coupling constants were found to be available for sucrose² and 1-kestose^{45,50}. For these two sugars, the measured 3J value pertaining to the glycosidic linkage is $J_{\text{H-1g-C-2f}}$ (corresponding to $^6\text{H}_1$ - $^6\text{C}_1$ - O_{link} - $^5\text{C}_2$). Table 10 compares the experimental and calculated values, showing the MD simulation results obtained with 50mM solute concentration to agree well with the experiment. The experimental concentration of sucrose (60mM)² was very close to the value used in the simulation. However, as the experimental concentration of kestose was much higher at 300mM⁴⁵, a 20ns long MD simulation was performed at that concentration to check for any dependence of conformational properties on concentration. Table 10 shows no significant change in the J -value associated with the glycosidic linkage dihedral with a six-fold increase in concentration, which suggests the absence of any significant solute-solute interactions even at the higher concentration.

The conformation of the fructofuranose ring in sucrose has also received significant attention^{2,51}. The interpretation of NMR data confirms that the ring conformation is confined to the NE quadrant with the pseudorotation angle value between 0 and 90°^{3,4}. Figure S5 (Supporting Information) shows the distribution of the pucker phase calculated over the 20 ns simulation time, where it is indeed confined to this region. One precise method for assessing the solution conformation of furanose rings involves measurement of three bond ^1H - ^1H coupling constants of the ring hydrogens and subsequent usage of the program PSEUROT⁴⁹. PSEUROT assumes a model in which two conformations, one each in South and North pucker angle hemispheres, respectively, exist in equilibrium. However, this analysis cannot be applied to sucrose (or kestose) as the pseudorotation angle is confined to only the NE quadrant. Therefore, we compare directly the ^1H - ^1H coupling constants from $^5\text{H}_3$ - $^5\text{C}_3$ - $^5\text{C}_4$ - $^5\text{H}_4$ and $^5\text{H}_4$ - $^5\text{C}_4$ - $^5\text{C}_5$ - $^5\text{H}_5$ dihedral angles (Table 10). Results from the MD simulations systematically overestimate the experimental values. For kestose, the same behavior is observed at the higher concentration. While potentially problematic, it should be noted that the Haasnoot-Altona equation⁴⁴ used to obtain the computational coupling constants did not include furanoid systems in its parametrization, therefore limiting

the interpretation of the extent of agreement between calculation and experiment. In addition, it has been observed that the glycosidic conformational distribution is sensitive to pucker⁴. The good agreement obtained with the J -coupling constant of for ${}^6\text{H}_1$ - ${}^6\text{C}_1$ - O_{link} - ${}^5\text{C}_2$ therefore suggests that pucker distribution is sampled correctly. Based on these observations and the quality of the force field in reproducing the sugar pucker parameters in the crystals, as indicated by the quality of the agreement for dihedrals ${}^5\text{O}_{\text{ring}}$ - ${}^5\text{C}_2$ - ${}^5\text{C}_3$ - ${}^5\text{C}_4$ and ${}^5\text{C}_2$ - ${}^5\text{C}_3$ - ${}^5\text{C}_4$ - ${}^5\text{C}_5$ in Table 8, further optimization of the ring-associated parameters was deemed unnecessary.

Having shown the agreement with crystal and the available NMR data, we tested the sensitivity of the parameters to the environment using sucrose as a test case. We compared the conformational ensembles populated by sucrose in vacuum and in solution. 400 ns-long replica exchange simulations (see Methods) in vacuum were performed with a single sucrose molecule. For the aqueous phase, the 40 ns simulation of a single sucrose molecule in a box of 1100 water molecules used to compute the J -coupling constants above was used. Figure 6 shows the Φ/Ψ glycosidic dihedral distribution obtained during the course of the two sets of simulations. Firstly, as observed in previous *in silico* investigations of sucrose⁴, the Φ dihedral is limited to values between 0° and 180° . In vacuum, the sugar predominantly populates a single conformational basin centered near Φ/Ψ values 90° - 30° . Φ/Ψ values from most of the known crystal structures (Figure 6, black triangles) of sucrose (References from Venable et. al⁴) lie in this basin, supporting our observations of this being the preferred conformation in the absence of aqueous solvation. Conformations near Φ/Ψ 60° - 60° and 90° - 60° are visited, but have very low probabilities. However, in solution, the two conformational basins 90° - 30° and 90° - 60° are comparably populated. More than 20 transitions occur between these two conformational basins indicating the convergence of data. Not surprisingly, the J value computed from conformations sampled in vacuo (5.0 Hz) differs from the experiment (4.2 Hz) significantly as opposed to the quantitative agreement from the solution simulation (4.2 Hz). Taken together, these results show that the hierarchically developed parameters are robust in that, the same set of parameters is able to capture subtle differences seen in different environments.

PSEUROT⁴⁹ analysis has been applied to a model disaccharide with an F2→F6 linkage (F1→F5 according to IUPAC naming of arabinose carbons) containing D-arabinofuranose⁵². We compared the populations of North (N) and South (S) conformations, phase angle and amplitudes obtained from MD simulations of model compound 9 (αAraf -(1→5) α -1-Ome-Araf) used in Houseknecht et al.⁵² (Table 11). These data show that the probabilities of North and South conformations, the average pseudorotation angle and phase agree reasonably with the prediction of PSEUROT applied to NMR data. Houseknecht et al. note that introduction of a glycosidic linkage between two arabinofuranose residues does not change the ring conformational preference as compared to methyl- α -arabinofuranose. Our results also suggest the same when a comparison is made with previous MD simulation results of methyl- α -arabinofuranose¹⁴. Thus, the parameters are able to reproduce condensed phase properties of di/poly-saccharides composed of not only of fructofuranose, but also aldopentofuranoses.

Conclusions—In this work, we have developed and validated additive all-atom CHARMM force field parameters for glycosidic linkages involving furanose sugars. These parameters cover $1 \leftrightarrow 2$, $1 \rightarrow 3$, $1 \rightarrow 4$, $1 \rightarrow 6$ pyranose-furanose linkages and $2 \rightarrow 1$ and $2 \rightarrow 6$ furanose-furanose linkages. Building on existing parameters for pyranose and furanose sugars, several simple model compounds that contained the key ring atoms of disaccharides linked by different configurations were selected. Target data for optimization of the bond, angle and dihedral parameters associated with the glycosidic linkage included 2D scans of the Φ/Ψ glycosidic dihedral angles in the model compounds computed by QM MP2/cc-

pVTZ single point energies on MP2/6-31G(d) optimized structures with the five-membered ring restrained to the favorable North and South conformations of furanose sugars. All possible chiralities of the model compounds at the linkage carbon were considered, yielding over 12600 conformations. The optimized parameters are able to reproduce accurately the crystal intramolecular geometries on a test set of eight di/polysaccharide crystals containing the different linkages, with slight systematic overestimation of unit cell dimensions, a phenomenon that was observed for other CHARMM carbohydrate^{13-15,48,53} as well GLYCAM carbohydrate force field parameters⁵³. Parameters were validated in the aqueous condensed phase by comparison to experimental solution densities, with all errors 1.5% or less. Good agreement was obtained with the available NMR data for three compounds, validating the ability of the hierarchically developed parameters to describe the conformational behavior of the target compounds. However, it should be emphasized that the present optimization did not address the relative energies of the anomeric states of the individual disaccharides and, accordingly, it is anticipated that these energies may not be accurately treated by the present force field. The newly developed parameters, presented in Tables S5 and S6 of the supporting information, allow for the modeling of linear, branched, and cyclic pyranose/furanose polysaccharides both alone and in heterogeneous systems including proteins, nucleic acids and/or lipids when combined with existing additive CHARMM biomolecular force fields.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grant GM070855. The authors acknowledge computer time and resources from the Department of Defense High Performance Computing, and the Pittsburgh Supercomputing Center. The authors thank Wenbo Yu for performing consistency checks of the data reported in this paper.

References

1. Burton BA, Brant DA. Biopolymers. 1983; 22:1769.
2. Penhoat, CHd; Imberty, A.; Roques, N.; Michon, V.; Mentech, J. J Am Chem Soc. 1991; 113:3720.
3. Freedberg DI. J Am Chem Soc. 2002; 124:2358. [PubMed: 11878992]
4. Venable RM, Delaglio F, Norris SE, Freedberg DI. Carbohydrate Research. 2005; 340:863. [PubMed: 15780252]
5. Rehm, BHA., editor. Microbial Production of Biopolymers and Polymer Precursors: Applications and Perspectives. Caister Academic Press; Norfolk, UK: 2009.
6. Calazans GMT, Lopes CE, Lima RMOC, França Fpd. Biotechnology Letters. 1997; 19:19.
7. Wong, CH., editor. Carbohydrate-Based Drug Discovery. Wiley-VCH Verlag GmbH & Co.; KGaA: 2003.
8. Glennon TM, Zheng YJ, Legrand SM, Shutzberg BA, Merz KM. Journal of Computational Chemistry. 1994; 15:1019.
9. Kowizjer MLCE, Grootenhuis PDJ. J Phys Chem. 1995
10. Dauchez M, Derreumaux P, Lagant P, Vergoten G. Journal of Computational Chemistry. 1995; 16:188.
11. Clark M III, C RD, Opdenbosch NV. Journal of Computational Chemistry. 1989; 10:982.
12. Seo M, Castillo N, Ganzynkiewicz R, Daniels CR, Woods RJ, Lowary TL, Roy PN. J Chem Theory Comput. 2008; 4:184.
13. Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, MacKerell AD Jr. Journal of Computational Chemistry. 2008; 29:2543. [PubMed: 18470966]

14. Hatcher E, Guvench O, MacKerell Alexander D J. *J Phys Chem B*. 2009; 113:12466. [PubMed: 19694450]
15. Guvench O, Hatcher ER, Venable RM, Pastor RW, MacKerell AD. *Journal of Chemical Theory and Computation*. 2009; 5:2353. [PubMed: 20161005]
16. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. *Journal of Physical Chemistry B*. 1998; 102:3586.
17. MacKerell AD Jr, Wiórkiewicz-Kuczera J, Karplus M. *Journal of the American Chemical Society*. 1995; 117:11946.
18. Feller SE, Gawrisch K, MacKerell AD Jr. *Journal Of The American Chemical Society*. 2002; 124:318. [PubMed: 11782184]
19. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes PEM, Vorobyov I, MacKerell Alexander D J. *Journal of Computational Chemistry*. 2009; 31:671. [PubMed: 19575467]
20. Arjona O, Gómez AM, López JC, Plumet J. *Chemical Reviews*. 2007; 107:1919. [PubMed: 17488060]
21. Altona C, Sundaralingam M. *Journal of the American Chemical Society*. 1972; 94:8205. [PubMed: 5079964]
22. Altona C, Sundaralingam M. *Journal of the American Chemical Society*. 1973; 95:2333. [PubMed: 4709237]
23. Brooks BR, Brooks CL III, MacKerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RV, Woodcock HL, Wu X, Yang W, York DM, Karplus M. *J Comp Chem*. 2009; 30:1545. [PubMed: 19444816]
24. Klauda JB, Venable RM, Freites A, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, Vorobyov I, MacKerell Alexander D J, Pastor RW. *J Phys Chem B*. 2010; 114:7830. [PubMed: 20496934]
25. Allen, MP.; Tildesley, DJ. *Computer Simulation of Liquids*. Oxford University Press; Oxford: 1987.
26. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *Journal of Chemical Physics*. 1983; 79:926.
27. Durell SR, Brooks BR, Ben-Naim A. *Journal of Physical Chemistry*. 1994; 98:2198.
28. Ryckaert JP, Ciccotti G, Berendsen HJC. *Journal of Computational Physics*. 1977; 23:327.
29. Darden T, York D, Pedersen L. *Journal of Chemical Physics*. 1993; 98:10089.
30. Hockney, RW. The potential calculation and some applications. In: Alder, B.; Fernbach, S.; Rotenberg, M., editors. *Methods in Computational Physics*. Vol. 9. Academic Press; New York: 1970. p. 136
31. Nosé S. *Molecular Physics*. 1984; 52:255.
32. Hoover WG. *Physical Review A*. 1985; 31:1695. [PubMed: 9895674]
33. Feller SE, Zhang YH, Pastor RW, Brooks BR. *Journal of Chemical Physics*. 1995; 103:4613.
34. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA.; Vreven, T., Jr; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, K.; Kitao, O.; Nakai, H.; Klene, M.; Li, TW.; Knox, JE.; Hratchian, HP.; Cross, JB.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Ayala, PY.; Morokuma, K.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Zakrzewski, VG.; Dapprich, S.; Daniels, AD.; Strain, MC.; Farkas, O.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Ortiz, JV.; Cui, Q.; Baboul, AG.; Clifford, S.; Cioslowski, J.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Challacombe, M.; Gill, PMW.; Johnson, B.;

- Chen, W.; Wong, MW.; Gonzalez, C.; Pople, JA. Gaussian 03. Revision B.04. Gaussian, Inc.; Pittsburgh, PA: 2003.
35. Møller C, Plesset MS. Physical Review. 1934; 46:618.
36. Hariharan PC, Pople JA. Theoretica Chimica Acta. 1973; 28:213.
37. Scott AP, Radom L. Journal Of Physical Chemistry. 1996; 100:16502.
38. Pulay P, Fogarasi G, Pang F, Boggs JE. Journal of the American Chemical Society. 1979; 101:2550.
39. Dunning TH. Journal of Chemical Physics. 1989; 90:1007.
40. MacKerell AD Jr. Journal Of Computational Chemistry. 2004; 25:1584. [PubMed: 15264253]
41. MacKerell AD Jr, Karplus M. Journal Of Physical Chemistry. 1991; 95:10559.
42. Guvench O, MacKerell AD. Journal of Molecular Modeling. 2008; 14:667. [PubMed: 18458967]
43. Tvaroska I, Hricovíni M, Petráková E. Carbohydrate Research. 1989; 189:359.
44. Haasnoot CAG, Leeuw FAAMd, Altona C. Tetrahedron. 1980; 19
45. Calub TM, Waterhouse AL, Chatterton NJ. Carbohydrate Research. 1990; 199:11.
46. Sugita Y, Okamoto Y. Chemical Physics Letters. 1999; 314:141.
47. Allen FH. Acta Crystallographica Section B-Structural Science. 2002; 58:380.
48. Hatcher ER, Guvench O, MacKerell AD. Journal of Chemical Theory and Computation. 2009; 5:1315. [PubMed: 20160980]
49. Leeuw FAAMD, Altona C. Journal of Computational Chemistry. 1983; 4:428.
50. Waterhouse AL, Calub TM, French AD. Carbohydrate Research. 1991; 217:29. [PubMed: 1797404]
51. Engelsen S, Penhoat CHd, Perez S. J Phys Chem. 1995; 99:13334.
52. Houseknecht JB, Altona C, Hadad CM, Lowary TL. Journal of Organic Chemistry. 2002; 67:4647. [PubMed: 12098271]
53. Tschampel SM, Kennerty MR, Woods RJ. Journal of Chemical Theory and Computation. 2007; 3:1721.

Figure 1.
Model compounds used in this study

Figure 2.

Φ/Ψ potential energy surfaces in the QM and MM (fit) representations for the 4 anomers of model compound 1 with the five membered ring frozen to South (left) and North (right) conformations, respectively. $\alpha\alpha$ South (a,b), $\alpha\alpha$ North (c,d), $\alpha\beta$ South (e,f), $\alpha\beta$ North (g,h), $\beta\alpha$ South (i,j), $\beta\alpha$ North (k,l), $\beta\beta$ South (m,n), $\beta\beta$ North (o,p). Energies are in kcal/mol with contours every 1 kcal/mol.

Figure 3. Φ/Ψ potential energy surfaces in the QM and MM (fit) representations for model compound 2 α anomer with the five membered ring frozen to South (a, b) and North (c, d) conformations, respectively. Energies are in kcal/mol with contours every 1 kcal/mol.

Figure 4.

Φ/Ψ potential energy surfaces in the QM and MM (fit) representations for the 4 stereoisomers of model compound 3 with the five membered ring frozen to South (left) and North (right) conformations, respectively. $\alpha\alpha$ South (a,b), $\alpha\alpha$ North (c,d), $\alpha\beta$ South (e,f), $\alpha\beta$ North (g,h), $\beta\alpha$ South (i,j), $\beta\alpha$ North (k,l), $\beta\beta$ South (m,n), $\beta\beta$ North (o,p). Energies are in kcal/mol with contours every 1 kcal/mol. Although the linkages are not between two anomeric centers, stereoisomers are denoted by the α/β nomenclature, to be consistent with other model compounds.

Figure 5.

Φ/Ψ potential energy surfaces in the QM and MM (fit) representations for the 2 stereoisomers of model compound 4 with the five membered ring frozen to south (left) and north (right) conformations, respectively. $\alpha\beta$ South (a,b), $\alpha\beta$ North (c,d), $\beta\alpha$ South (e,f), $\beta\alpha$ North (g,h). Energies are in kcal/mol with contours every 1 kcal/mol. Although the linkages are not between two anomeric centers, stereoisomers are denoted by the α/β nomenclature, to be consistent with other model compounds.

Figure 6. Distribution of Φ/Ψ dihedral values of sucrose in vacuum (left) and in solution (right). Black triangles show the values observed for sucrose in crystal media. Two representative conformations of sucrose corresponding to the two conformational basins populated are shown.

Table 1

Optimized geometries in QM and MM representations for all anomers of model compound 1. Valence angles and dihedral angles are in degrees, bond lengths are in Å

Anomer	Bond length/ valence angle/ dihedral angle	QM	MM	MM-QM
1 $\alpha\alpha$	⁶ C ₁ -O _{link}	1.422	1.416	-0.006
	⁵ C ₂ -O _{link}	1.426	1.424	-0.002
	⁶ O _{ring} - ⁶ C ₁ -O _{link}	111.2	109.6	-1.6
	⁶ C ₂ - ⁶ C ₁ -O _{link}	106.8	108.7	1.9
	⁵ O _{ring} - ⁵ C ₂ -O _{link}	111.4	111.0	-0.4
	⁵ C ₃ - ⁵ C ₂ -O _{link}	106.6	106.2	-0.4
	⁶ C ₁ -O _{link} - ⁵ C ₂	112.6	112.6	0.0
	⁶ O _{ring} - ⁶ C ₁ -O _{link} - ⁵ C ₂ (Φ)	67.5	70.6	3.1
	⁶ C ₁ -O _{link} - ⁵ C ₂ - ⁵ O _{ring} (Ψ)	66.3	71.1	4.8
	⁵ O _{ring} - ⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄	41.3	34.5	-6.8
	⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄ - ⁵ C ₅	-33.6	-36.4	-2.8
1 $\alpha\beta$	⁶ C ₁ -O _{link}	1.424	1.418	-0.006
	⁵ C ₂ -O _{link}	1.426	1.428	0.002
	⁶ O _{ring} - ⁶ C ₁ -O _{link}	112.7	110.6	-2.0
	⁶ C ₂ - ⁶ C ₁ -O _{link}	106.4	108.8	2.4
	⁵ O _{ring} - ⁵ C ₂ -O _{link}	111.9	111.4	-0.5
	⁵ C ₃ - ⁵ C ₂ -O _{link}	107.4	106.4	-1.0
	⁶ C ₁ -O _{link} - ⁵ C ₂	114.3	114.7	0.4
	⁶ O _{ring} - ⁶ C ₁ -O _{link} - ⁵ C ₂ (Φ)	57.2	57.5	0.3
	⁶ C ₁ -O _{link} - ⁵ C ₂ - ⁵ O _{ring} (Ψ)	-95.7	-106.6	-10.8
	⁵ O _{ring} - ⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄	-41.2	-35.0	6.2
	⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄ - ⁵ C ₅	36.6	35.9	-0.7
1 $\beta\alpha$	⁶ C ₁ -O _{link}	1.400	1.416	0.016
	⁵ C ₂ -O _{link}	1.433	1.427	-0.006
	⁶ O _{ring} - ⁶ C ₁ -O _{link}	107.1	109.4	2.3
	⁶ C ₂ - ⁶ C ₁ -O _{link}	108.9	106.8	-2.1
	⁵ O _{ring} - ⁵ C ₂ -O _{link}	113.2	112.7	-0.5
	⁵ C ₃ - ⁵ C ₂ -O _{link}	106.0	106.5	0.5
	⁶ C ₁ -O _{link} - ⁵ C ₂	114.6	114.7	0.1
	⁶ O _{ring} - ⁶ C ₁ -O _{link} - ⁵ C ₂ (Φ)	-92.1	-92.2	-0.1
	⁶ C ₁ -O _{link} - ⁵ C ₂ - ⁵ O _{ring} (Ψ)	43.0	42.7	-0.3
	⁵ O _{ring} - ⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄	41.9	38.0	-3.9

Anomer	Bond length/ valence angle/ dihedral angle	QM	MM	MM-QM
	$^5\text{C}_2\text{-}^5\text{C}_3\text{-}^5\text{C}_4\text{-}^5\text{C}_5$	-33.1	-32.6	0.5
1 β	$^6\text{C}_1\text{-O}_{\text{link}}$	1.402	1.413	0.011
	$^5\text{C}_2\text{-O}_{\text{link}}$	1.429	1.424	-0.005
	$^6\text{O}_{\text{ring}}\text{-}^6\text{C}_1\text{-O}_{\text{link}}$	107.5	109.2	1.7
	$^6\text{C}_2\text{-}^6\text{C}_1\text{-O}_{\text{link}}$	108.5	106.9	-1.5
	$^5\text{O}_{\text{ring}}\text{-}^5\text{C}_2\text{-O}_{\text{link}}$	111.5	111.1	-0.4
	$^5\text{C}_3\text{-}^5\text{C}_2\text{-O}_{\text{link}}$	106.4	106.2	-0.2
	$^6\text{C}_1\text{-O}_{\text{link}}\text{-}^5\text{C}_2$	113.1	112.4	-0.7
	$^6\text{O}_{\text{ring}}\text{-}^6\text{C}_1\text{-O}_{\text{link}}\text{-}^5\text{C}_2$ (Φ)	-68.0	-68.9	-0.9
	$^6\text{C}_1\text{-O}_{\text{link}}\text{-}^5\text{C}_2\text{-}^5\text{O}_{\text{ring}}$ (Ψ)	-64.2	-68.1	-3.9
	$^5\text{O}_{\text{ring}}\text{-}^5\text{C}_2\text{-}^5\text{C}_3\text{-}^5\text{C}_4$	-41.1	-34.4	6.7
	$^5\text{C}_2\text{-}^5\text{C}_3\text{-}^5\text{C}_4\text{-}^5\text{C}_5$	33.6	36.5	2.9

Table 2

Water pair interaction energies for model compounds.

Model compound	Water orientation ^a	Energy (kcal/mol)	MM	MM-QM	Distance (Å)	MM	MM-QM
		1.16*HF^b					
1 $\alpha\alpha$	90	-3.32	-3.52	-0.20	2.11	2.15	0.04
	-90	-3.58	-3.96	-0.38	2.03	2.12	0.09
1 $\alpha\beta$	0	-4.64	-3.31	1.33	1.97	2.24	0.27
	90	-3.28	-4.40	-1.12	2.19	2.08	-0.11
	270	-3.84	-3.97	-0.13	2.05	2.11	0.06
1 $\beta\alpha$	90	-5.96	-4.87	1.09	1.84	1.81	-0.03
	180	-5.05	-4.44	0.61	1.93	1.93	0.00
	270	-6.26	-6.33	-0.07	1.83	1.77	-0.06
1 $\beta\beta$	0	-5.57	-4.77	0.8	1.89	1.84	-0.05
	90	-5.27	-5.26	0.01	1.87	1.80	-0.07
	270	-6.17	-6.15	0.02	1.83	1.79	-0.04
			average	0.18			0.01
			standard deviation	0.49			0.01
2 α	90	-5.87	-5.31	0.56	1.87	1.82	-0.05
	180	-5.83	-5.56	0.27	1.82	1.77	-0.05
	270	-5.05	-4.73	0.32	1.93	1.83	-0.10
			average	0.38			-0.07
			standard deviation	0.15			0.03
3 $\alpha\beta$	90	-1.93	-2.87	-0.94	2.63	2.50	-0.13
	270	-1.83	-2.66	-0.83	2.48	2.49	0.01
3 $\beta\alpha$	0	-3.79	-3.61	0.18	2.00	2.02	0.02
	90	-5.38	-3.50	1.88	1.85	2.01	0.16
	180	-5.07	-4.16	0.91	1.95	2.06	0.11
	270	-5.98	-5.17	0.81	1.85	1.96	0.11
3 $\beta\beta$	0	-5.47	-4.82	0.65	1.94	1.89	-0.05

Model compound	Water orientation ^a	Energy (kcal/mol)	MM	MM-QM	Distance (Å)	MM	MM-QM
		1.16*HF^b					
	90	-4.89	-5.23	-0.34	1.95	1.84	-0.11
	180	-1.68	-1.69	-0.01	2.50	2.68	0.18
	270	-5.00	-5.86	-0.86	1.90	1.82	-0.08
			average	-0.03			0.02
			standard deviation	0.84			0.09
4αβ	0	-4.58	-4.35	0.23	2.02	2.11	0.09
4βα	90	-6.46	-5.87	0.59	1.82	1.78	-0.04
	180	-4.97	-5.12	-0.15	1.97	1.89	-0.08
	270	-5.93	-6.13	-0.2	1.84	1.77	-0.07
			average	0.12			-0.03
			standard deviation	0.37			0.08

^a Dihedral angle (degrees) defined by C1/2-Olink...Owater-H2 where H2 is the non-interacting water hydrogen atom. In all cases, the water H1-Owater bond vector is in the C-Olink-C angle plane and along the C-Olink-C bisector.

^b HF target energies have been scaled by 1.16 and distances have been shortened by 0.20 Å.

Table 3

Optimized geometries in QM and MM representations for model compound 2. Valence angles and dihedral angles are in degrees, bond lengths are in Å.

Anomer	Bond length/ valence angle/ dihedral angle	QM	MM	MM-QM
2α	⁵ C ₂ - ⁵ C ₁	1.520	1.524	0.004
	⁵ C ₁ -O _{link}	1.422	1.428	0.006
	O _{link} -C _M	1.426	1.422	-0.004
	⁵ O _{ring} - ⁵ C ₂ - ⁵ C ₁	109.7	111.0	1.3
	⁵ C ₃ - ⁵ C ₂ - ⁵ C ₁	113.6	114.7	1.0
	⁵ C ₂ - ⁵ C ₁ -O _{link}	112.0	112.2	0.2
	⁵ C ₁ -O _{link} -C _M	112.9	112.0	-0.9
	⁵ O _{ring} - ⁵ C ₂ - ⁵ C ₁ -O _{link} (Φ)	65.0	67.3	2.3
	⁵ C ₂ - ⁵ C ₁ -O _{link} -C _M (Ψ)	-95.4	-96.3	-0.9
	⁵ O _{ring} - ⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄	22.2	12.5	-9.6
	⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄ - ⁵ C ₅	-36.6	-29.8	6.8

Table 4

Optimized geometries in QM and MM representations for all stereoisomers of model compound 3. Valence angles and dihedral angles are in degrees, bond lengths are in Å

Stereoisomer	Bond length/ valence angle/ dihedral angle	QM	MM	MM-QM
3 $\alpha\alpha$	⁶ C ₁ -O _{link}	1.416	1.401	-0.015
	⁵ C ₃ -O _{link}	1.438	1.433	-0.005
	⁶ O _{ring} - ⁶ C ₁ -O _{link}	111.6	110.1	-1.6
	⁶ C ₂ - ⁶ C ₁ -O _{link}	106.7	108.6	1.9
	⁵ C ₂ - ⁵ C ₃ -O _{link}	110.9	111.7	0.8
	⁵ C ₄ - ⁵ C ₃ -O _{link}	106.2	105.9	-0.3
	⁶ C ₁ -O _{link} - ⁵ C ₃	113.5	114.0	0.56
	⁶ O _{ring} - ⁶ C ₁ -O _{link} - ⁵ C ₃ (Φ)	63.4	67.4	4.0
	⁶ C ₁ -O _{link} - ⁵ C ₃ - ⁵ C ₂ (Ψ)	85.7	85.4	-0.3
	⁵ O _{ring} - ⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄	-24.2	-24.2	-0.1
	⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄ - ⁵ C ₅	38.0	36.5	-1.5
3 $\alpha\beta$	⁶ C ₁ -O _{link}	1.416	1.403	-0.013
	⁵ C ₃ -O _{link}	1.431	1.435	0.004
	⁶ O _{ring} - ⁶ C ₁ -O _{link}	112.9	110.6	-2.3
	⁶ C ₂ - ⁶ C ₁ -O _{link}	106.3	108.5	2.2
	⁵ C ₂ - ⁵ C ₃ -O _{link}	113.3	113.2	-0.2
	⁵ C ₄ - ⁵ C ₃ -O _{link}	105.5	105.4	0.0
	⁶ C ₁ -O _{link} - ⁵ C ₃	116.1	115.4	-0.7
	⁶ O _{ring} - ⁶ C ₁ -O _{link} - ⁵ C ₃ (Φ)	74.0	76.7	2.6
	⁶ C ₁ -O _{link} - ⁵ C ₃ - ⁵ C ₂ (Ψ)	-77.4	-86.2	-8.9
	⁵ O _{ring} - ⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄	40.7	25.8	-14.9
	⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄ - ⁵ C ₅	-39.4	-37.4	2.0
3 $\beta\alpha$	⁶ C ₁ -O _{link}	1.394	1.399	0.005
	⁵ C ₃ -O _{link}	1.438	1.433	-0.005
	⁶ O _{ring} - ⁶ C ₁ -O _{link}	108.4	110.1	1.7
	⁶ C ₂ - ⁶ C ₁ -O _{link}	108.6	106.7	-1.9
	⁵ C ₂ - ⁵ C ₃ -O _{link}	111.1	110.6	-0.5
	⁵ C ₄ - ⁵ C ₃ -O _{link}	107.5	106.2	-1.3
	⁶ C ₁ -O _{link} - ⁵ C ₃	113.3	113.8	0.5
	⁶ O _{ring} - ⁶ C ₁ -O _{link} - ⁵ C ₃ (Φ)	-59.6	-58.3	1.3
	⁶ C ₁ -O _{link} - ⁵ C ₃ - ⁵ C ₂ (Ψ)	93.4	99.1	5.7
	⁵ O _{ring} - ⁵ C ₂ - ⁵ C ₃ - ⁵ C ₄	-3.6	-20.0	-16.5

Stereoisomer	Bond length/ valence angle/ dihedral angle	QM	MM	MM-QM
	$^5\text{C}_2\text{-}^5\text{C}_3\text{-}^5\text{C}_4\text{-}^5\text{C}_5$	27.3	35.1	7.8
3 $\beta\beta$	$^6\text{C}_1\text{-O}_{\text{link}}$	1.396	1.399	0.003
	$^5\text{C}_3\text{-O}_{\text{link}}$	1.440	1.433	-0.007
	$^6\text{O}_{\text{ring}}\text{-}^6\text{C}_1\text{-O}_{\text{link}}$	107.9	109.7	1.8
	$^6\text{C}_2\text{-}^6\text{C}_1\text{-O}_{\text{link}}$	108.3	106.6	-1.8
	$^5\text{C}_2\text{-}^5\text{C}_3\text{-O}_{\text{link}}$	111.0	112.2	1.1
	$^5\text{C}_4\text{-}^5\text{C}_3\text{-O}_{\text{link}}$	105.9	105.9	0.0
	$^6\text{C}_1\text{-O}_{\text{link}}\text{-}^5\text{C}_3$	113.8	114.3	0.5
	$^6\text{O}_{\text{ring}}\text{-}^6\text{C}_1\text{-O}_{\text{link}}\text{-}^5\text{C}_3$ (Φ)	-66.5	-64.9	1.6
	$^6\text{C}_1\text{-O}_{\text{link}}\text{-}^5\text{C}_3\text{-}^5\text{C}_2$ (Ψ)	-84.8	-82.1	2.7
	$^5\text{O}_{\text{ring}}\text{-}^5\text{C}_2\text{-}^5\text{C}_3\text{-}^5\text{C}_4$	24.5	24.9	0.4
	$^5\text{C}_2\text{-}^5\text{C}_3\text{-}^5\text{C}_4\text{-}^5\text{C}_5$	-38.0	-36.7	1.3

Table 5

Optimized geometries in QM and MM representations for the two stereoisomers of model compound 4. Valence angles and dihedral angles are in degrees, bond lengths are in Å

Stereoisomer	Bond length/ valence angle/ dihedral angle	QM	MM	MM-QM
4 $\alpha\beta$	$^6\text{C}_2\text{-O}_{\text{link}}$	1.444	1.441	-0.003
	$^5\text{C}_2\text{-O}_{\text{link}}$	1.411	1.398	-0.013
	$^6\text{C}_1\text{-}^6\text{C}_2\text{-O}_{\text{link}}$	106.2	107.7	1.5
	$^6\text{C}_3\text{-}^6\text{C}_2\text{-O}_{\text{link}}$	110.4	108.6	-1.7
	$^5\text{O}_{\text{ring}}\text{-}^5\text{C}_2\text{-O}_{\text{link}}$	113.2	112.0	-1.2
	$^5\text{C}_3\text{-}^5\text{C}_2\text{-O}_{\text{link}}$	106.9	106.0	-0.9
	$^6\text{C}_2\text{-O}_{\text{link}}\text{-}^5\text{C}_2$	114.4	113.7	-0.7
	$^6\text{C}_1\text{-}^6\text{C}_2\text{-O}_{\text{link}}\text{-}^5\text{C}_2$ (Φ)	-144.3	-135.0	9.3
	$^6\text{C}_2\text{-O}_{\text{link}}\text{-}^5\text{C}_2\text{-}^5\text{O}_{\text{ring}}$ (Ψ)	-64.4	-61.8	2.6
	$^5\text{O}_{\text{ring}}\text{-}^5\text{C}_2\text{-}^5\text{C}_3\text{-}^5\text{C}_4$	-41.6	-35.7	5.9
	$^5\text{C}_2\text{-}^5\text{C}_3\text{-}^5\text{C}_4\text{-}^5\text{C}_5$	33.6	36.4	2.8
4 $\beta\alpha$	$^6\text{C}_2\text{-O}_{\text{link}}$	1.439	1.440	0.001
	$^5\text{C}_2\text{-O}_{\text{link}}$	1.411	1.399	-0.013
	$^6\text{C}_1\text{-}^6\text{C}_2\text{-O}_{\text{link}}$	107.1	107.8	0.7
	$^6\text{C}_3\text{-}^6\text{C}_2\text{-O}_{\text{link}}$	110.7	108.7	-2.0
	$^5\text{O}_{\text{ring}}\text{-}^5\text{C}_2\text{-O}_{\text{link}}$	113.0	111.9	-1.1
	$^5\text{C}_3\text{-}^5\text{C}_2\text{-O}_{\text{link}}$	106.9	106.1	-0.8
	$^6\text{C}_2\text{-O}_{\text{link}}\text{-}^5\text{C}_2$	114.2	113.4	-0.8
	$^6\text{C}_1\text{-}^6\text{C}_2\text{-O}_{\text{link}}\text{-}^5\text{C}_2$ (Φ)	140.7	135.3	-5.4
	$^6\text{C}_2\text{-O}_{\text{link}}\text{-}^5\text{C}_2\text{-}^5\text{O}_{\text{ring}}$ (Ψ)	62.6	62.7	0.1
	$^5\text{O}_{\text{ring}}\text{-}^5\text{C}_2\text{-}^5\text{C}_3\text{-}^5\text{C}_4$	41.5	35.5	-6.0
	$^5\text{C}_2\text{-}^5\text{C}_3\text{-}^5\text{C}_4\text{-}^5\text{C}_5$	-34.5	-36.3	-1.8

Table 6

Sugar crystals used for force field validation. In the linkage type notation, arrows point from the anomeric center, and double-headed arrows indicate a linkage between two anomeric centers.

CSD accession code	Common name	# of monosaccharides	Component monosaccharides	Linkages covered	# of sugar molecules	# of water molecules
SUCROS04	Sucrose	2	α -D-glucopyranose, β -D-fructofuranose	P1 $\leftrightarrow$ F2	2	0
RAFINOS	Raffinose	3	α -D-galactopyranose, α -D-glucopyranose, β -D-fructofuranose	P1 $\leftrightarrow$ F2	4	20
MELTZT	Melzitose	3	α -D-glucopyranose (2), β -D-fructofuranose	P1 $\rightarrow$ F3, P1 $\leftrightarrow$ F2	4	4
KESTOS	1-Kestose	3	α -D-glucopyranose, β -D-fructofuranose (2)	F2 $\rightarrow$ F1, P1 $\leftrightarrow$ F2	4	0
CELGII	6-Kestose	3	α -D-glucopyranose, β -D-fructofuranose (2)	F2 $\rightarrow$ F6, P1 $\leftrightarrow$ F2	4	0
PLANTE10	Planteose	3	α -D-glucopyranose (2), β -D-fructofuranose,	P1 $\rightarrow$ F6, P1 $\leftrightarrow$ F2	4	8
PEKHES	Nystose	4	α -D-glucopyranose, β -D-fructofuranose (3)	F2 $\rightarrow$ F1, P1 $\leftrightarrow$ F2	4	16
IMATUL	Isomaltulose	2	α -D-glucopyranose, β -D-fructofuranose	P1 $\rightarrow$ F6	4	4

Table 7

Crystalline unit cell geometries and volumes.

	A (Å)		B (Å)		C (Å)		volume (Å³)		
	expt	MD	% error	expt	MD	% error	expt	MD	% error
SUCROS04	10.86	10.91	0.45	8.71	8.79	0.99	715.04	753.86	5.43
RAFINO	8.97	9.20	2.61	12.33	12.35	0.15	2634.56	2767.46	5.04
MELEZT	14.95	15.40	2.99	13.86	14.25	2.8	2240.81	2393.33	6.81
KESTOS	7.94	7.91	-0.33	9.99	10.49	4.91	2117.29	2276.71	7.53
CELGII	8.72	9.16	5.08	11.32	10.74	-5.09	2307.30	2447.27	6.07
IMATUL	9.05	9.26	2.26	12.16	12.15	-0.02	1554.35	1598.93	2.87
PLANTEI0	32.43	33.85	4.37	8.15	8.11	-0.53	2302.92	2418.94	5.04
PEKHES	13.58	13.83	1.84	23.39	23.68	1.27	3239.05	3405.36	5.13
average error			2.41			0.56			2.51
									5.49

Table 8

Errors in reproduction of crystalline internal geometries. Data are from simulations of crystals listed in Table 7; MD data for a particular crystal were averaged over the independent disaccharide molecules in the crystal as well as over the MD trajectory; bonds are in Å and valence and dihedral angles are in degrees. The different types of bonds, angles and dihedrals present in the diverse linkages covered are tabulated. For each type, average and standard deviation of error over all crystals containing that type of bond/angle dihedral was computed.

	Average error	Standard dev. of errors
Bonds		
${}^6\text{C}_1\text{-O}_{\text{link}}$	0.009	0.007
${}^5\text{C}_{2/3}\text{-O}_{\text{link}}$	0.005	0.014
${}^5\text{C}_{2/5}\text{-}{}^5\text{C}_{1/6}$	0.003	0.016
${}^5\text{C}_{1/6}\text{-O}_{\text{link}}$	0.003	0.008
Angles		
${}^6\text{O}_{\text{ring}}\text{-}{}^6\text{C}_1\text{-O}_{\text{link}}$	0.5	0.4
${}^6\text{C}_2\text{-}{}^6\text{C}_1\text{-O}_{\text{link}}$	0.6	1.1
${}^5\text{O}_{\text{ring}}\text{-}{}^5\text{C}_2\text{-O}_{\text{link}}$	1.0	1.3
${}^5\text{C}_3\text{-}{}^5\text{C}_2\text{-O}_{\text{link}}$	0.6	1.5
${}^6\text{C}_1\text{-O}_{\text{link}}\text{-}{}^5\text{C}_2$	-1.0	1.6
${}^5\text{O}_{\text{ring}}\text{-}{}^5\text{C}_{2/5}\text{-}{}^5\text{C}_{1/6}$	-0.4	2.5
${}^5\text{C}_{3/4}\text{-}{}^5\text{C}_{2/5}\text{-}{}^5\text{C}_{1/6}$	0.0	2.0
${}^5\text{C}_{2/5}\text{-}{}^5\text{C}_{1/6}\text{-O}_{\text{link}}$	1.4	2.0
$\text{C-O}_{\text{link}}\text{-}{}^5\text{C}_{1/6}$	-1.0	0.9
Dihedrals		
${}^6\text{O}_{\text{ring}}\text{-}{}^6\text{C}_1\text{-O}_{\text{link}}\text{-}{}^5\text{C}_2$	-2.4	4.2
${}^6\text{C}_1\text{-O}_{\text{link}}\text{-}{}^5\text{C}_2\text{-}{}^5\text{O}_{\text{ring}}$	1.5	3.8
${}^5\text{O}_{\text{ring}}\text{-}{}^5\text{C}_{2/5}\text{-}{}^5\text{C}_{1/6}\text{-O}_{\text{link}}$	-2.5	9.0
${}^5\text{C}_{2/5}\text{-}{}^5\text{C}_{1/6}\text{-O}_{\text{link}}\text{-}{}^5\text{C}_{2/5}$	0.6	10.0
${}^5\text{C}_{1/6}\text{-O}_{\text{link}}\text{-}{}^5\text{C}_{2/5}\text{-}{}^5\text{O}_{\text{ring}}$	3.4	4.9
${}^5\text{O}_{\text{ring}}\text{-}{}^5\text{C}_2\text{-}{}^5\text{C}_3\text{-}{}^5\text{C}_4$	4.7	4.8
${}^5\text{C}_2\text{-}{}^5\text{C}_3\text{-}{}^5\text{C}_4\text{-}{}^5\text{C}_5$	-1.0	2.1

Table 9

Comparison of the calculated and experimental solution densities for different Molal concentrations of Sucrose, Isomaltulose and Lactulose in 1100 TIP3P waters at $T = 298\text{ K}$ and $P = 1\text{ atm}$.

Sugar	Molality	N_{solute}	Expt.	MD	% error
Sucrose	0.22	4	1.025	1.035	0.97
	0.44	9	1.050	1.059	0.94
	0.56	11	1.062	1.068	0.59
	0.59	12	1.066	1.073	0.68
	0.90	18	1.095	1.097	0.2
Lactulose	0.12	2	1.012	1.025	1.23
	0.15	3	1.016	1.030	1.38
	0.20	4	1.023	1.035	1.21
	0.26	5	1.030	1.040	0.97
Isomaltulose	0.12	2	1.011	1.025	1.32
	0.15	3	1.014	1.030	1.52
	0.20	4	1.021	1.035	1.38
	0.25	5	1.026	1.040	1.32

Table 10

J-coupling constants computed from MD simulations of Sucrose and Kestose compared to experimental values. The second set of MD calculated values (at 50mM concentration) in parentheses are from an independent simulation to verify convergence. For kestose, MD results are also shown for 300mM concentration.

Sugar	Dihedral angle	Exp	MD (50 mM)	MD (300 mM)
Sucrose ^a	⁶ C ₁ -O _{link} - ⁵ C ₂ - ⁵ H ₂	4.2 ²	4.24 (4.30) ^a	-
Kestose ^a	⁶ C ₁ -O _{link} - ⁵ C ₂ - ⁵ H ₁	4.3 ⁴⁵	4.25 (4.18) ^a	4.15 ^a
Sucrose ^b	⁵ H ₃ - ⁵ C ₃ - ⁵ C ₄ - ⁵ H ₄	8.8 ²	9.77 (9.77) ^b	-
Sucrose ^c	⁵ H ₄ - ⁵ C ₄ - ⁵ C ₅ - ⁵ H ₅	8.3 ²	9.09 (9.10) ^c	-
Kestose ^b	⁵ H ₃ - ⁵ C ₃ - ⁵ C ₄ - ⁵ H ₄	8.0 ⁵⁰	9.80 (9.80) ^b	9.78 ^b
Kestose ^c	⁵ H ₄ - ⁵ C ₄ - ⁵ C ₅ - ⁵ H ₅	8.5 ⁵⁰	9.21 (9.18) ^c	9.17 ^c
Kestose ^{b,d}	⁵ H ₃ - ⁵ C ₃ - ⁵ C ₄ - ⁵ H ₄	8.0 ⁵⁰	9.82 (9.82) ^{b,d}	9.77 ^{b,d}
Kestose ^{c,d}	⁵ H ₄ - ⁵ C ₄ - ⁵ C ₅ - ⁵ H ₅	7.5 ⁵⁰	8.98 (8.97) ^{c,d}	8.95 ^{c,d}

Computed using ^a Eqn 4 ^b Eqn 5 ^c Eqn 6.

^d The second fructofuranose ring in kestose.

Comparison of ring pucker properties for α Araf-(1 $\rightarrow$ 5) α -1-Ome-Araf obtained from PSEUROT⁵² and from MD simulations in aqueous solution. The second calculated set of values in parentheses are from an independent simulation to verify convergence.

Table 11

Sugar	Expt. P _N	Calc P _N	Expt. P _S	Calc P _S	Expt %N	Calc %N	Expt %S	Calc %S	Avg. ϕ_m
AARB1	45	68 (67)	126	125 (126)	40	32 (28)	60	68 (72)	40 (40)
AARB2	44	68 (68)	121	126 (125)	39	31 (30)	61	69 (70)	41 (41)