

NIH Public Access

Author Manuscript

J Comput Chem. Author manuscript; available in PMC 2010 June 8.

Published in final edited form as:

J Comput Chem. 2008 November 30; 29(15): 2543-2564. doi:10.1002/jcc.21004.

Additive empirical force field for hexopyranose monosaccharides

Olgun Guvench[†], Shannon N. Greene[†], Ganesh Kamath[†], John W. Brady[‡], Richard M. Venable[§], Richard W. Pastor[§], and Alexander D. MacKerell Jr.^{*,†}

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, 20 Penn St., HSF II-629, Baltimore, MD 21201, and Department of Food Science, Cornell University, Ithaca, New York 14853, and Laboratory of Computational Biology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892

Abstract

We present an all-atom additive empirical force field for the hexopyranose monosaccharide form of glucose and its diastereomers allose, altrose, galactose, gulose, idose, mannose, and talose. The model is developed to be consistent with the CHARMM all-atom biomolecular force fields, and the same parameters are used for all diastereomers, including both the α - and β -anomers of each monosaccharide. The force field is developed in a hierarchical manner and reproduces the gas-phase and condensed-phase properties of small-molecule model compounds corresponding to fragments of pyranose monosaccharides. The resultant parameters are transferred to the full pyranose monosaccharides and additional parameter development is done to achieve a complete hexopyranose monosaccharide force field. Parametrization target data include vibrational frequencies, crystal geometries, solute - water interaction energies, molecular volumes, heats of vaporization, and conformational energies, including those for over 1800 monosaccharide conformations at the MP2/ cc-pVTZ//MP2/6-31G(d) level of theory. Though not targeted during parametrization, free energies of aqueous solvation for the model compounds compare favorably with experimental values. Also well-reproduced are monosaccharide crystal unit cell dimensions and ring pucker, densities of concentrated aqueous glucose systems, and the thermodynamic and dynamic properties of the exocyclic torsion in dilute aqueous systems. The new parameter set expands the CHARMM additive force field to allow for simulation of heterogeneous systems that include hexopyranose monosaccharides in addition to proteins, nucleic acids, and lipids.

Keywords

carbohydrate; glucose; galactose; pyranose; force field; molecular dynamics; molecular mechanics; CHARMM

Introduction

Carbohydrates are being increasingly recognized for their important roles in biology and chemistry. The well-known carbohydrate polymers cellulose and starch provide structure and a storehouse of energy in plants.¹ In humans, the highly-branched carbohydrate polymer glycogen serves as a rapidly-available energy reserve, and disturbances in the interconversion between intracellular glycogen and plasma glucose is a hallmark of type 2 diabetes.^{2,3} Protein glycosylation in eukaryotes serves to aid in protein folding, quality control, and localization,

Corresponding author: Phone: 410/706-7442; Fax: 410/706-5017; alex@outerbanks.umaryland.edu.

[†]University of Maryland

[‡]Cornell University

[§]National Institutes of Health

⁴ highlighting the importance of carbohydrates in molecular recognition. In disease-causing prokaryotes, conjugates of carbohydrates with peptides, in the form of peptidoglycans, and with lipids, in the form of lipopolysaccharide, play important roles in pathogenesis,^{5,6} while changes in glycosylation patterns are seen in cancer⁷; these examples reveal the potential impact that carbohydrate research can have on human disease. Finally, interest in carbohydrates in the field of chemistry has seen a surge, as these molecules have applications both as biofuels and as chemical feedstock.^{8,9}

Empirical force field simulations provide a means of probing atomic level-of-detail properties such as structure, dynamics, and thermodynamics that are difficult or impossible to deduce using experimental methods.¹⁰ Such simulations can help to inform and suggest new directions of research in the rapidly-growing fields of glycobiology and biofuels. To perform simulation studies at a level of accuracy representative of the experimental regimen, it is necessary to have a well parametrized force field suitable for simulations in the condensed phase. Significant efforts have already been made toward molecular-mechanics force fields for modeling carbohydrates.^{11–22} Unfortunately, in these models there has been substantial variability in the methods used to develop both the bonded and nonbonded parameters, not to mention differences in force-field functional forms. While such differences are of no consequence when simulating only carbohydrates using a single force field, they become problematic when attempting to model heterogeneous systems using other protein, nucleic acid, and lipid force fields. Differences in functional forms lead to obvious problems in computational implementation. Both differences in functional form and parameterization methods lead to the more subtle, but perhaps more important, problem of the balance of interaction energies caused by a lack of consistency between classes of molecules. We do note that an effort in this direction has been made in the most recent version of the GLYCAM carbohydrate force field so as to make it applicable to all molecular classes,²³ though further work is required to apply and validate it for modeling proteins, nucleic acids, and lipids.

Toward the ultimate aim of a comprehensive molecular mechanics force field for the simulation of carbohydrate polymers and compatible with the existing CHARMM all-atom protein,^{24,25} nucleic acid,^{26–28} and lipid force fields,^{29–34} we present a force field model for glucopyranose and its diastereomers (Figure 1). Glucose in its hexopyranose form (glucopyranose, Figure 1d) is the basic subunit of the polymers cellulose, starch, and glycogen. Glucopyranose and its diastereomers such as galactose (Figure 1c) and mannose (Figure 1g) are the building blocks of a large number of biologically and chemically important carbohydrate polymers. Accordingly, development of a force field for hexopyranoses represents a first major step towards the creation of a comprehensive carbohydrate force field.

Parameter Development Approach

An important goal of the present work was to develop a force-field model that is consistent with existing force fields for other biomolecules, thereby enabling the simulation of heterogeneous systems. We achieve this goal by using the same functional form and parametrization procedure as the other CHARMM all-atom additive biomolecular force fields¹⁰ that have been developed for proteins, nucleic acids, and lipids. Solute-water interactions guide the development of nonbonded parameters, so as to ensure a proper balance between water – water, water – solute, and solute – solute interactions, the so-called "interaction triad." Both partial atomic charge and Lennard-Jones parameter values are adjusted to reproduce scaled quantum mechanical solute – water interaction energies and distances, and are further refined in a self-consistent fashion so as to also reproduce experimental heats of vaporization and molecular volumes for neat liquids. Additionally, 1,4-nonbonded interactions are treated in the same fashion as in the rest of the CHARMM force field in that there is no scaling of either the electrostatic or the Lennard-Jones portion of these interactions. This not

only ensures compatibility with the protein, nucleic acid, and lipid parameters, but also facilitates the reproduction of monosaccharide conformational energetics.³⁵

As is typical in the development of biomolecular force fields,^{24,36–38} the parameters were developed in a hierarchical fashion. A partial parameter set was created using small-molecule model compounds corresponding to fragments of the hexopyranose monosaccharides. The model compounds were: methanol, ethanol, isopropanol, ethylene glycol, cyclohexane, and tetrahydropyran (Figure 2). The model-compound parameters were then applied to the pyranose monosaccharides, missing parameters were developed, and the completed monosaccharide force field was validated for its ability to reproduce calculated quantum mechanical as well as experimental properties. The hierarchical approach makes the parameter development process tractable by minimizing the number of parameters that must be simultaneously optimized while limiting the number of atom types that can be used, thereby reducing the dimensionality of parameter space. The validation step at the end of the parameter-development process ensures that the parameters are indeed transferable from the small-molecule model compounds to the full pyranose monosaccharides.

Reference experimental properties used for development and validation of the model compound parameters were the thermodynamic quantities heat of vaporization (ΔH_{vap}), molecular volume (V_{m}), and free-energy of aqueous solvation (ΔG_{sol}); infrared vibrational frequencies; x-ray crystallographic intramolecular and unit-cell geometries; densities of concentrated aqueous glucose solutions; and the thermodynamics and dynamics of exocyclic-group rotation. The targeted computed quantum mechanical (QM) properties were vibrational frequencies, conformational energies, solute – water interaction energies, and dipole moments.

After the model compounds parameters were finalized, the applicable parameters were used to construct a preliminary monosaccharide force field. The only missing parameters at this stage were a subset of dihedrals. These were fit to over 1800 QM hexopyranose conformational energies. These included rotation of hydroxyl groups, rotation of the exocylic hydroxymethyl, and ring deformations leading to chair-to-boat conversions. The parameters were also tested for their ability to reproduce QM conformational energies for the pyranose mannose, which was not in the dihedral parameter training set. Final testing and parameter adjustment was done to ensure proper monosaccharide behavior in both the crystalline and solution phases, as the ultimate goal was to produce a condensed-phase carbohydrate force field. These tests consisted of molecular dynamics (MD) simulations of infinite crystals, of dilute solutions, and of concentrated solutions, and confirmed the ability of the force field to reproduce crystal geometries including ring pucker, the dynamics and thermodynamics of rotation of the exocylic hydroxymethyl in solution, and the densities of 1 M, 2 M, and 5 M glucose – water solutions.

The remainder of the manuscript describes details of the development of parameters for the various functional groups, starting with the alkane and hydroxyl functionalities and then moving to the ring and ring ether groups. This is followed by a description of parameter development for complete pyranose monosaccharides. The fine details of the various methods, as required to reproduce the computations, are reserved for a "Methods" section located at the end of the manuscript, and additional supporting data is provided as "Supplementary Material." Finally, an Appendix is included showing why the Cremer and Pople ring-pucker parameter, θ , should be computed from an averaged structure, rather than as an ensemble average.

Results and Discussion

Alkane and Hydoxyl Functionalities

Alkanes—The alkane parameters used in the development of the current force field were taken directly from a recent update to the CHARMM additive alkane parameters that accurately

reproduce not only the properties of short linear and branched alkanes $(2 \le n \le 4)$, but also long linear alkanes (n = 7, 10).³⁹ In addition to having been validated in the context of the linear alkanes, these parameters have also been shown to be appropriate when used in the context of ethers.³⁹ Furthermore, the parameters are only minimally different than the prior version of the CHARMM alkane parameters^{31,32,34} and were developed using the same methodology, but employing a larger set of model compounds. Finally, they were designed to be compatible with the highly-optimized CHARMM protein,^{24,25} nucleic acid,^{26–28} and lipid^{29–34} biomolecular force fields. As such, we constrained ourselves not to modify these alkane parameters in the development of the present carbohydrate force field.

Alcohols—To determine force field parameters for the hydroxyl moiety in the pyranose monosaccharides, we started with the alcohol series of methanol, ethanol, and isopropanol (Figure 2b–d). The original CHARMM22 parameters²⁴ for these three alcohols yield $V_{\rm m}$ and $\Delta H_{\rm vap}$ values within a ±2% window of error with respect to experimental results at laboratory temperature and pressure. However, because of the updates to the alkane nonbonded parameters, the hydroxyl nonbonded parameters also required updating. In particular, the oxygen Lennard-Jones σ value was updated from 0.1521 kcal/mol to 0.1921 kcal/mol and the ϵ value from 1.770 Å to 1.765 Å. The partial charges on the hydroxyl oxygen and hydrogen were also updated, from values of -0.66 and 0.43 electrons respectively, to the new values of -0.65 and 0.42 electrons. These small changes reflected the prior small changes to the alkane nonbonded parameters.⁴⁰

Like the CHARMM22 parametrization, the updated model is able to reproduce both the experimental $V_{\rm m}$ and $\Delta H_{\rm vap}$ data for these three alcohols (Table 1). The experimental $\Delta G_{\rm sol}$ values in Table 1 were not directly targeted in the parametrization as they are computationally costly to determine relative to the other two thermodynamic properties. Nonetheless, the values are within 0.5 kcal/mol of experimental values showing that, for this set of compounds, solvation free-energy is well-reproduced by parameters that reproduce $V_{\rm m}$ and $\Delta H_{\rm vap}$.

Scaled HF/6-31G(d) solute – water interaction energies and distances were used to guide the development of hydroxyl nonbonded parameters that yield accurate reproduction of $V_{\rm m}$ and $\Delta H_{\rm vap}$. In order to generate parameters appropriate for a condensed phase force-field, a scale factor of 1.16 needs to be applied to the Hartree-Fock (HF) energies and the HF distances need to be reduced by 0.2 Å.24, 41 The solute – water interaction energies were determined by optimizing the distance between the water molecule and the alcohol at fixed relative orientations, with the TIP3P geometry42 for the water molecule and an MP2/6-31G(d)optimized geometry for the alcohol. Using these geometries, the scaled HF interaction energy is calculated simply as $1.16*(E_{pair} - E_{solute} - E_{water})$ with no basis-set superposition-error correction, so as to be consistent with the CHARMM additive force field development protocol. With the appropriate scaling, these OM distances and energies serve as guides for the development of nonbonded parameters that are capable of reproducing condensed-phase properties and maintain a proper balance in the water – water, water – solute, and solute – solute "interaction triad." The scaling is required as it is not possible to simultaneously reproduce both QM interaction energies and bulk properties in a pairwise-additive molecular mechanics framework.43

The solute – water interaction geometries for methanol use various water orientations to probe this simple alcohol both as a hydrogen bond donor and a hydrogen bond acceptor (Figure 3). The orientation of the water molecule and the internal geometry of the model compound are chosen so as to minimize all other water – solute interactions, especially between the water and aliphatic moieties, since the goal is to probe hydrogen bonding and also because the HF method does a poor job of modeling dispersion interactions that are important for interactions involving aliphatic carbon and hydrogen atoms.⁴⁴ A similar protocol is used to calculate the

molecular mechanics (MM) energies and distances, using a model-compound geometry optimized in the MM representation prior to construction of the solute-water system and optimization of the hydrogen bond distance. The CHARMM22 and current force-field energy and distance values are notable for their similarity both to each other and to the HF target data (Table 2). In particular both force fields overestimate the absolute value of the scaled HF interaction energy by 0.1 to 0.5 kcal/mol and both nearly exactly reproduce the shortened HF interaction distance. These results are consistent with the similar performance of the two force fields with regard to the thermodynamic target data and reflect the relatively small changes that needed to be made to the CHARMM22 hydroxyl nonbonded parameters in order to achieve good thermodynamic properties in the current parameterization using the updated alkane parameters.

The QM conformational energies for the alcohols were calculated at the MP2/cc-pVTZ level45^{,46} by constraining the dihedral angle of interest to a fixed value and fully optimizing (relaxing) all other degrees of freedom. The MP2/cc-pVTZ model chemistry was chosen for these relaxed potential energy scans because it offered a favorable combination of accuracy and computational tractability based on prior studies. In particular, ethylene glycol⁴⁷ and tetrahydropyran conformational energetics48 are well-accounted for at reasonable computational cost as compared to more complicated model chemistries. Figure 4c shows that both the original CHARMM22 and the current optimized parameter set exactly reproduce the rotational energy profile about the central C-C bond in ethanol. Additionally, the new parameters exactly reproduce the surfaces for rotation of the methanol, ethanol, and isopropanol hydroxyl groups whereas the older CHARMM22 parameterization slightly underestimates the MP2/cc-pVTZ barriers for methanol hydroxyl rotation and overestimates one of the two barriers for both ethanol and isopropanol hydroxyl rotation (Figures 4a,b,d).

The vibrational frequencies of the three alcohols were calculated using the MP2/6-31G(d) model chemistry.⁴⁹ The starting conformation for the MP2/6-31G(d) optimization was that of the global energy minimum as determined from the MP2/cc-pVTZ scans, and the calculated frequencies were scaled by 0.9434, as required to reproduce experimental infrared frequencies. 50 Figure 5 shows the experimental infrared, the scaled QM, and the MM frequencies for methanol; the H-C-O angle's force constant was adjusted relative to the CHARMM22 value to obtain the results for the current parameter set. The MM vibrational frequencies for ethanol and isopropanol are in similarly excellent agreement with experiment (the frequencies for all model compounds are listed in Supplementary Material S1). As a final check, the bond lengths and angles of force-field optimized alcohol conformations were confirmed to reproduce the modes of distributions from Cambridge Structural Database⁵¹ surveys. Throughout the optimization of the bond and angle equilibrium parameters, crystal internal geometries were targeted as they include condensed-phase effects.

The parameterization for the alcohols and all other compounds was done in a self-consistent fashion, such that whenever one parameter was changed to better reproduce one of the pieces of target data, all of the other target data were also recomputed so as to ensure consistency of the model.¹⁰ This was done iteratively such that the final parameters were able to accurately reproduce all of the target data. All of the values presented reflect the final bonded and nonbonded parameters.

Ethylene glycol—The model compound ethylene glycol is particularly relevant to the monosaccharides as it is the smallest compound having vicinal hydroxyl groups. There are three pairs of vicinal hydroxyl groups in each of the twelve $C_6H_{12}O_6$ pyranose monosaccharides, and their intramolecular interactions play an important role in the conformational energetics of these molecules. All available parameters from the alcohol series were directly applied to ethylene glycol. These included all of the bond, angle, Lennard-Jones,

and Coulomb parameters, and nearly all of the dihedral parameters. The undetermined dihedral parameters were for the rotation of the hydroxyl moieties and rotation of the O-C-C-O dihedral. The target data for these parameters were, as in the case of the alcohols, conformational energies from MP2/cc-pVTZ optimized structures.

Relaxed QM potential energy scans for the hydroxyl and O-C-C-O rotations in ethylene glycol were started from the all-*trans* conformation. For each of the two scans, the two other dihedrals were constrained to remain *trans* (180°). This eliminated the possibility of internal hydrogen bonding while allowing for relaxation of bond lengths and angles. To account for the other conformations that ethylene glycol can assume, the ten non-symmetry related minimum-energy conformations of the molecule⁵² were also considered as target data.

Three different parameterization strategies were initially pursued for the ethylene glycol dihedral parameters. In one, the hydroxyl and O-C-C-O torsion scans were targeted ("HOCC/OCCO" fit), in another the relative energies of the 10 minimum-energy conformations were targeted with a 10x weighting of the 3 lowest-energy conformations in the fit ("MinE" fit), and in the third both the scans and the weighted minimum-energy conformations were targeted ("combo" fit). The "HOCC/OCCO" and "MinE" approaches both failed because of very inaccurate energies for conformations outside the training set (Figure 6).

The "combo" fit also proved problematic because of the large unfavorable force-field nonbonded energy for geometries with a dihedral angle of near 0° in the O-C-C-O torsion scan. As the CHARMM force field does not scale 1,4-nonbonded interactions, the large negative partial charges on the two oxygens lead to a strong electrostatic repulsion when these two atoms are *cis* or near-*cis* during the scan, as occurs in the O-C-C-O rotation scan as well as in the two highest-energy conformations in the set of 10 minimum-energy conformations. The MP2 results show a substantially lower barrier, presumably due to electronic polarization, which is absent in the force field. So as to ensure compatibility with the rest of the CHARMM biomolecular force field, we did not consider introducing 1,4-electrostatic scaling, and instead focused on fitting the energies of conformations that had a minimal 1,4 O...O electrostatic repulsion, which were those having a *gauche* O-C-C-O conformation and internal hydrogen bonding or a *trans* O-C-C-O conformation.

To account for the overestimated 1.4 O...O electrostatic repulsion, the following target data were used for the final fit ("combo*" fit): the lowest eight of the ten minimum energy conformations, with empirical 10x weighting of the three lowest-energy conformations to better-reproduce their relative energies; the full hydroxyl scan; and only those conformations from the O-C-C-O scan for which the dihedral was greater than or equal to 130°. This approach led to good reproduction of the minimum energy conformations and hydroxyl scan energetics, as well as the energetics for the O-C-C-O scan for values greater than 90° , in particular contrast to the poor performance of the "HOCC/OCCO" and "MinE" fits (Figure 6). The QM local minimum at 70° in the O-C-C-O scan, whose geometry is very similar to the fully-relaxed tGt conformation in the minimum energy series, is shifted to 90° and is 2 kcal/mol too high in energy. However, in the QM representation, this local minimum is nearly 4 kcal/mol above the global minimum (tGg') and thus it is practically irrelevant from a thermodynamic perspective with respect to the gas phase monomer energetics, and is also the case in apolar solvents. Due to 1,4 O...O electrostatic repulsion, the "combo*" fit has a 20.2 kcal/mol barrier at 0° as compared to 12.3 kcal/mol and 12.8 kcal/mol for the QM and "combo" models respectively. However, transitions between the two minima on the O-C-C-O scan take place across the much lower 6 kcal/mol barrier at 120°, and this barrier is very well reproduced by the "combo*" parameters. In polar solvents electrostatic screening will lower this barrier that results from the electrostatic repulsion, serving to lessen the effect of the inherent limitation of the force-field functional form. In support of the present approach of targeting the

thermodynamically relevant conformations instead of introducing 1,4-nonbonded scaling, it has been shown that 1,4-nonbonded scaling hampers the ability of a force field to reproduce the conformational energetics of monosaccharides.³⁵

Using nonbonded parameters transferred from the alcohols, the thermodynamic properties $V_{\rm m}$, $\Delta H_{\rm vap}$, and $\Delta G_{\rm sol}$ for ethylene glycol are in good accord with experimental values (Table 1). The ability of the model to reproduce these bulk properties is encouraging, as they depend primarily on the nonbonded parameters, which in this case have been optimized using only the alcohol series. Using the global-energy minimum tGg' conformation, solute – water interaction energies and distances are in close agreement with the scaled HF target data for geometries in which ethylene glycol is either a hydrogen bond donor or a hydrogen bond acceptor. The average energetic error for these interactions is +0.09 kcal/mol and the average distance error is +0.03 Å (Supplementary Material S2). Like the bulk properties, the accuracy of the solute – water interactions confirms the transferability of the nonbonded parameters from the alcohols to ethylene glycol.

The dipole moment for ethylene glycol shows the same behavior relative to the QM gas-phase value as do the force field models for methanol, ethanol, isopropanol, namely, all are overestimated relative to the QM values by 15% to 30% (Table 3). This overpolarization reflects the well-known phenomenon that fixed-charge additive force field models need to be overpolarized relative to gas-phase in order to correctly capture the energetics of condensed-phase polar and aqueous systems.¹⁰ As expected, the differences between the force field values and the QM values are much larger than between the CHARMM22 values and those using the current parameterization. The current parameter set yields dipole moments for the alcohols that are only 1% smaller than the CHARMM22 results, reflecting the minor change in the hydroxyl charges that resulted from optimizing the hydroxyl moiety nonbonded in the context of the updated alkane nonbonded parameters.

Ring and Ring Ether Functionalities

Model compounds used in the development of ring and ring ether parameters comprise cyclohexane and tetrahydropyran (Figures 2f and 2g). The same updated nonbonded alkane parameters for the carbons and aliphatic hydrogens as tested on the alcohols and ethylene glycol were used directly. As with the alcohols and ethylene glycol, the data considered in the parametrization process included heats of vaporization, molecular volumes, vibrational frequencies, potential energy surfaces, solute – water interaction energies and distances, crystallographic geometries, and dipole moments.

Cylcohexane—With the exception of the C-C-C-C dihedral parameter, all of the bonded and nonbonded parameters required to model cyclohexane were available from the most recent CHARMM alkane parameterization for linear and branched alkanes.³⁹ The C-C-C-C dihedral, which occurs six times in the cyclic molecule, was parameterized to reproduce the relaxed MP2/cc-pVTZ potential energy surface (PES) for rotation of the C-C-C-C dihedral from –100° to +100°, and is distinguished from the linear alkane C-C-C-C parameters by the introduction of atom types specific to aliphatic carbon atoms in six-membered rings. The PES was calculated using a constraint on a single C-C-C-C dihedral, and all other degrees-of-freedom were allowed to relax during energy optimization. However, as all six C-C-C-C dihedral values during the scan was incorporated into the fitting procedure.⁵³ The final parametrization employed only a 3-fold dihedral term, as the introduction of 1- and 2-fold dihedral terms did little to improve the fit. The force field reproduces the QM data over the full 200° span, with energies spanning 25 kcal/mol and geometries that include both the global energy-minimum chair conformation and the local energy-minimum twist-boat conformation (Figure 7).

Using the combination of the linear alkane parameters and the optimized C-C-C-C dihedral parameters for the ring, the thermodynamic properties for cyclohexane are in good agreement with target values (Table 1). Likewise, the bond lengths and angles are all in accord with target geometries, and the vibrational frequencies are well-reproduced (Supplementary Material S1e). Because of the lack of hydrogen bonding functionality on cyclohexane, no solute – water interactions were calculated. The overall performance of the cyclohexane model attests to the transferability of the force field parameters and validates the hierarchical approach to carbohydrate parameter development.

Tetrahydropyran—The parameters for tetrahydropyran (Figure 2g) build on the parameters for cyclohexane. Missing values include the non-bonded parameters for the ring ether oxygen and bonded parameters that involve this atom type. Self-consistent empirical optimization of both the ring ether oxygen charge and oxygen Lennard-Jones parameters along with optimization of the dihedral parameters yielded a parametrization that, like for the other model compounds, well-reproduced the target data.

Potential energy scans of the C-C-C, C-O-C-C, and O-C-C-C dihedrals show that the force field model captures the conformational energetics of the molecule at the MP2/cc-pVTZ level. Like cyclohexane, the scans include local-minimum twist-boat conformations that are higher in energy than the global-minimum chair conformation (Figure 8). Also like cyclohexane, only 3-fold terms were used for the new C-O-C-C and O-C-C-C dihedral terms, while the C-C-C dihedral was transferred directly from cyclohexane. The inclusion of 1- and 2-fold terms did little to improve the fit, which is already excellent using only 3-fold terms.

The optimized ring-ether oxygen non-bonded parameters, developed with guidance from solute-water interaction energies and distances (Table 4, VFigure 9), give excellent reproduction of the experimental m, ΔH_{vap} , and ΔG_{sol} values (Table 1). The THP-water hydrogen bond energy is electrostatic in nature and determined by the ether oxygen partial charge q. Since neat THP is not a hydrogen-bonded liquid, ΔH_{vap} is sensitive mainly to the Lennard-Jones well-depth parameter ϵ . The hydrogen bond interaction distance is largely a function of the R_{\min} Lennard-Jones parameter, which defines the location of the steep repulsive wall in the vdW energy surface. Based on prior work24,41 and as confirmed above for the alcohols, CHARMM force field development applies the general empirical rule that hydrogen bonds needs to be about 0.2 Å shorter than the HF/6-31G(d) optimized distance in order to reproduce condensed phase properties. In the case of the cyclic ether oxygen, we found that the hydrogen bond distance needed to be 0.3 Å shorter in order to properly reproduce the condensed phase property $V_{\rm m}$, which is strongly dependent on $R_{\rm min}$. Thus, the ether oxygen is an exception to the general rule. The scaling rule for the HF energy is however maintained, as the oxygen partial charge developed using a scaling factor of 1.16 for the solute-water interaction energy yields the correct ΔG_{sol} (Table 1). Both the partial charge and Lennard-Jones parameters developed for the ring ether oxygen have also been used in the development of a force-field model for tetrahydrofuran, and are directly transferable to that molecule.³⁹ They are also transferable to linear ethers with minor adjustment of the partial charge, showing that the ether oxygen Lennard-Jones parameters developed for THP are appropriate for both linear and cyclic ethers.³⁹

Monosaccharides

Choice of QM method—Starting with the model-compound parameters, only a subset of dihedral parameters needed to be determined to develop a force-field model for the pyranose monosaccharide form of glucose and its diastereomers, which differ from glucose and one another only in the chirality at the C2, C3, and/or C4 positions (Figure 1). The undetermined dihedral parameters were those involved in hydroxyl rotation, rotation of the exocylic group,

and certain ring deformations. Owing to the computational cost of the MP2/cc-pVTZ model chemistry, relaxed potential energy scans at this level of theory were impractical to obtain for target data. Three alternatives were explored: MP2/6-31G(d), MP2/cc-pVDZ, and MP2/cc-pVTZ single-point calculations using MP2/6-31G(d)-optimized geometries (i.e. the MP2/cc-pVTZ//MP2/6-31G(d) level of theory).

To test the accuracy of the alternative methods, the ΔE between various pyranose monosaccharides was calculated and compared to values at the MP2/cc-pVTZ level. Relative CPU times were also tabulated for the calculations, which included geometry optimization (Table 5). The results show that, while both MP2/6-31G(d) and MP2/cc-pVDZ are an order of magnitude faster than MP2/cc-pVTZ, both also err by 1 or more kcal/mol, which is large given that the MP2/cc-pVTZ ΔE values are ~1 kcal/mol. In contrast, MP2/cc-pVTZ/MP2/6-31G(d), which is also an order of magnitude faster than MP2/cc-pVTZ, quantitatively reproduces the target MP2/cc-pVTZ ΔE values. Consequently, all further QM calculations to generate target data for parametrization of MM conformational energies were done using the MP2/cc-pVTZ// MP2/6-31G(d) model chemistry.

Conformational energies—Though desirable, even at the MP2/cc-pVTZ//MP2/6-31(d) level, potential energy scans of all of the hydroxyl moieties, the exocyclic group, and the ring degrees of freedom to be parameterized for both the α - and β - anomers of the eight diastereomers of D-glucopyranose was too computationally costly. As a compromise, most of the target data were from scans of the α - and β -pyranose forms of glucose (Figure 1d) and of altrose (Figure 1b). This pair of monosaccharides includes the biologically ubiquitous glucose. Altrose was selected because, in combination with glucose, all possible pairs of axial-axial, axial-equatorial, and equatorial vicinal hydroxyls at the C₂, C₃, and C₄ positions are represented: the C₂–C₃ pair in altrose is an axial-axial pairing, the C₃–C₄ pair in altrose is an axial-equatorial pairing, and the C₂–C₃ and C₃–C₄ pairs in glucose are equatorial-equatorial pairings. The inclusion of both α - and β -anomers for both monosaccharides accounts for all possible axial-axial, axial-equatorial, axial-equatorial, and equatorial-equatorial-equatorial hydroxyl pairings at the C₁ and C₂ positions.

In addition to the hydroxyl, exocyclic, and ring deformation scans for glucose and altrose, potential energy scans of the α - and β -galactose ring degrees-of-freedom were also included as target data. These served to improve the MM ring energetics, which initially showed chair-to-boat interconversion in dilute aqueous molecular dynamics simulations of galactose and spuriously-low energy minima at the boat conformation of galactose in the gas phase relative to the QM data. The ring deformation scans were started from ⁴C₁ chair minimum-energy conformations and went 30° in one direction and 150° degrees in the other, with the latter leading to chair-to-boat conversions of the ring. Also included were galactose exocyclic scans and hydroxyl scans from four other hexopyranoses. The final target data in this training set consisted of 1860 MP2/cc-pVTZ//MP2/6-31G(d) conformational energies, 85% of which were of glucose and altrose (Table 6).

The dihedral parameters that required fitting were the hydroxyl torsions $H_1O_1C_1C_2$, $H_1O_1C_1O_5$, $H_2O_2C_2C_1$, $H_2O_2C_2C_3$, $H_3O_3C_3C_2$, $H_3O_3C_3C_4$, $H_4O_4C_4C_3$, $H_4O_4C_4C_5$, $C_5C_6O_6H_6$; the ring deformation torsions $O_5C_1C_2O_2$, $O_1C_1O_5C_5$, $O_5C_5C_4O_4$; and the exocyclic-group torsion $O_5C_5C_6O_6$. Reference MM energies for fitting were obtained by setting the dihedral force constants for these dihedrals to 0, reading in the QM-optimized conformation, placing harmonic restraints with a force constant of 10,000 kcal*mol⁻¹*degree⁻² on the $H_1O_1C_1C_2$, $H_2O_2C_2C_3$, $H_3O_3C_3C_4$, $H_4O_4C_4C_5$, $C_5C_6O_6H_6$, $O_5C_1C_2O_2$, $O_1C_1O_5C_5$, $O_5C_5C_4O_4$, and $O_5C_5C_6O_6$ dihedrals, fully optimizing the system with infinite nonbonded cutoffs, and calculating the energy without the dihedral restraints. During fitting, multiplicities of n = 1, 2, 3 were allowed for each dihedral. All the HOCC parameters

were constrained to be identical so as to prevent over-fitting and loss of parameter transferability. Ten 50,000-step Monte Carlo simulated annealing runs (described in "Methods" and Reference 53) were used to optimize the root mean square error (*RMSE*) between the QM and MM energies. All the parameters to be fit were allowed to vary simultaneously during the fitting, resulting in a 30-dimensional fitting problem (10 dihedrals with n = 1, 2, 3 multiplicities each). Each of the runs was seeded with random parameter values and converged to the same *RMSE*.

Fitting was done using the "HOCC/OCCO," "combo," and "combo*" ethylene glycol O-C-C-O dihedral parameters (Figure 6). The pre-fit RMSE value using "HOCC/OCCO" O-C-C-O parameters and with the parameters to-be-fit force constants set to 0 was 2.97 kcal/mol over all 1860 structures, and was reduced by 0.70 kcal/mol to 2.27 kcal/mol by fitting. With the "combo" O-C-C-O dihedral parameters the pre-fit RMSE of 2.68 kcal/mol was optimized to 1.68 kcal/mol, for a reduction of 1.00 kcal/mol. And using the "combo*" O-C-C-O dihedral parameters the RMSE was reduced by 0.88 kcal/mol from a value of 2.48 kcal/mol to 1.60 kcal/ mol. The ring dihedral geometries using parameters optimized using the "combo" O-C-C-O dihedral parameters were substantially better than the other two sets, having errors of within $\pm 5^{\circ}$ for nearly all ${}^{4}C_{1}$ conformations in the training set, as compared to the "HOCC/OCCO" and "combo*" fits, which had a significant number of conformations with larger errors. The excellent geometries combined with an RMSE only 0.08 kcal/mol larger than for "combo*" led to the use of "combo" O-C-C-O parameters for the pyranose monosaccharides in conjunction with the dihedral parameters developed using the 1860 hexopyranose conformations. Lastly, the 3-fold torsion barrier for rotation of the exocyclic group was empirically lowered by 1.5 kcal/mol to better reproduce the dynamics of exocyclic rotation in solution (see below), and changed the final RMSE of the fitting set from 1.68 to 1.69 kcal/mol.

Figure 10 compares the MM conformational energies to the target QM energies before and after fitting. The overall improvement across the 1860 target monosaccharide conformations is visually apparent. In a few instances, the fit MM data have difficulty in reproducing the QM data. For example, in the rotation of the C2 hydroxyl in α -altrose, a spurious minimum exists at 120° (Figure 11a). However, improvements in agreement with the QM data are more generally the case, as seen in rotation of the C1 hydroxyl in β -glucose (Figure 11b). In the latter instance, the fit MM data agree well with the QM data, including reproduction of the correct global minimum at 180°, as well as the local minimum at -60° . In contrast, prior to fitting the MM scan this surface was qualitatively wrong, with all energies being too low and a single prominent and incorrect minimum at -120° .

As a test of the transferability of the parameters, 20 non-redundant conformations of α -and β mannose each were created by randomizing all of the hydroxyl and exocyclic torsions to -60° , 60° , or 180° , and then used to generate MP2/cc-pVTZ//MP2/6-31G(d) data. Redundant conformations resulting from the QM optimization were discarded, yielding 13 unique α mannose conformations, including all three possible exocyclic-group conformations as well as a ${}^{1}S_{3}$ twist-boat conformation, and 15 unique β -mannose minimum-energy conformations, which included all three possible exocyclic-group conformations. The energies of these 28 conformers were then evaluated in the MM representation using the same restrained geometryoptimization protocol as for monosaccharides in the target data set. The ability of the force field to reproduce the energetics of the biologically-important mannose (Figure 12) are in line with those for the training set (Figure 10). While there are differences in energy between the QM and MM results, the force field does reproduce the QM with an accuracy comparable to the training set results, illustrating the transferability of the dihedral parameters to a molecule outside the training set.

Crystal simulations—Crystal simulations serve as a good test of both the bonded and nonbonded force-field parameters. The force field's ability to reproduce the experimental bond lengths, bond angles, and torsion angles increases confidence in the respective parameters for these degrees of freedom. A further test of the agreement with experiment is provided by the ring puckering parameters, which characterize the ring shape. Likewise, the ability of the force field to reproduce the dimensions of the crystal lattice provides validation of both the bonded and nonbonded parameters since the x, y, and z dimensions of the lattices are dependent on both the size of the monomers and the inter-monomer contact distances. Accordingly, simulations of crystal structures during the parametrization process were used to guide the development of the final parameter set.

Twenty-three crystal structures were found to match the connectivity of glucopyranose in the Cambridge Structural Database⁵¹ (CSD, version 5.28, November 2006). Of these, 2 were excluded because of disorder, 2 because of the presence of ions, 1 because it was a powder structure, 1 because it was cocrystallized with urea, and 1 because it lacked hydrogen coordinates, for a total of 7 excluded structures. The remaining 16 comprised 3 crystals of α -D-galactose (CSD ID codes ADGALA01, ADGALA03, ADGALA10), 1 crystal of α-Dmannose (ADMANN), 2 crystals of α -D-talose (ADTALO01, ADTALO10), 2 crystals of β -D-galactose (BDGLOS01, BDGLOS10), 1 crystal of β-D-allose (COKBIN), 4 crystals of α-D-glucose (GLUCMH11, GLUCSA, GLUCSA03, GLUSA10), and 3 crystals of β-D-glucose (GLUCSE, GLUCSE01, GLUCSE02). 14 of these had all coordinates determined, one molecule of monosaccharide in the reduced unit cell, and no other compounds in the crystal. With respect to the other 2, the single structure of mannose in the database (ADMANN) was included in the analysis because of its biological importance despite the absence of hydrogen on atoms O1, O2, and O4 and the presence of two molecules in the reduced cell. The second of the two, a crystal of α -D-glucopyranose (GLUCMH11), satisfied the same criteria as the other 14 except that it was a monohydrate. It was retained in the analysis because the force field is appropriate for use in heterogeneous systems that include water. Three crystal structures, ADGALA03, GLUCSA03, and GLUCSE02, were obtained at low temperatures (95°, 140°, and 95° K), and were excluded from this study; the best (R factor) structure at 295 K was used for these 3 monosaccharides.

Based on one crystal of each of the 7 unique monosaccharides represented by the 16 crystal structures, a subset of the bond and angle parameters were further optimized beyond the model compound values. In the few instances where this was done, the optimizations were very minor. For example, the equilibrium bond length for ring C-C bonds was decreased by 0.02 Å relative to the pure alkane values from which they were derived. Likewise, the C-O equilibrium bond length for ring hydroxyls at the C_2 , C_3 , and C_4 positions was reduced by 0.01 Å and for the C₁ position by 0.02 Å compared to the alcohol series. The equilibrium bond length for the C-O bonds of the cyclic ether was increased by 0.01 Å, while the C_2 - C_1 - O_5 equilibrium bond angle was reduced by 5.5° and the C₄-C₅-O₅ by 1.5° , and the C-O-H equilibrium bond angles were increased by 3°. To improve agreement with MP2/6-31G(d) vibrational frequencies of low-energy structures of α - and β -D-glucopyranose (Supplementary Material S3) from the set of 684 QM-optimized conformers (Table 6), the force constant for the C-O bond to the C_2 , C_3 , and C_4 hydroxyl groups was reduced relative to that from the model compounds. These minor equilibrium bond length and bond angle changes, along with the single change of a bond force constant, yielded the final monosaccharide bond and angle parameters. It should be noted that the final round MCSA dihedral fitting (see above) included these modifications assuring a self-consistent parametrization. Crystal simulations with these parameters, and using finalized nonbonded and dihedral parameters, yielded excellent reproduction of intramolecular geometries (Table 7).

Another metric of the agreement of simulated crystals with the experimental data is the ring pucker, which was evaluated using two distinct formalisms: that of Cremer and Pople,⁵⁴ and the virtual α torsions described by Rao.⁵⁵ The Cremer and Pople pucker parameters, Q, θ , and ϕ , represent a polar coordinate system, with the ideal ${}^{4}C_{1}$ chair conformation occurring at the $\theta = 0$ pole; φ has large fluctuations near the poles in MD simulations, and that pucker parameter will not be considered further. The ideal ${}^{4}C_{1}$ chair conformation is represented by $\alpha_{1} = \alpha_{2} =$ $\alpha_3 = -35^\circ$ for the virtual α torsions. These torsions represent the respective positions of the C_1 , C_3 , and C_5 ring atoms relative to the plane defined by C_2 , C_4 , and O_5 , and are both easy to calculate and straightforward to interpret. Table 8 compares the pucker values from the crystal structures with those from the simulation, and shows the RMS error based on the differences. These values are computed from an averaged structure, rather than as ensemble averages. Calculation of the ensemble average $\langle \theta \rangle$ gave rise to noticeably larger RMS errors, which were inconsistent with those observed for the virtual α torsions. Further investigation indicated that the cause of the inconsistency was a non-Gaussian distribution of θ making comparison based on ensemble-averaging inappropriate (additional details are provided in the Appendix). The agreement between the computed and experimental ring pucker parameters is very good across all eight monosaccharide crystals that were studied, attesting to the quality as well as the transferability of the dihedral parameters. Finally, although no experimental data is available, the ring pucker was also calculated from the four 20-ns aqueous simulations used to evaluate the exocyclic hydroxymethyl group rotation, as described subsequently. The θ values are: α -galactose 2.2, β -galactose 3.2, α -glucose 2.7, β -glucose 5.9. Compared to the crystal puckers in Table 8, all but α -galactose are 1–2° smaller in the aqueous simulations, suggesting that hexopyranose ring puckers are reduced in going from the crystal phase to aqueous solution.

With the optimized bond and angle parameters, and after refitting of the dihedral parameters to the target QM conformational data, there was a slight but systematic overestimation of the unit-cell volumes in monosaccharide crystal simulations. This overestimation was less than that in prior crystal simulations in this study that used exclusively model-compound derived parameters (not shown), showing that the additional bonded parameter refinement, along with improving monomer geometries in the crystal phase, improved the reproduction of crystal volumes, consistent with the small but systematic shortening of various equilibrium bond lengths. The remaining volume overestimation suggested an investigation of the nonbonded parameters that are involved in inter-molecular crystal contacts.

In all the crystals studied, the intermolecular contacts are exclusively hydrogen bonds involving the hydroxyl groups. During the parameter-optimization process for the alcohol series, both the Lennard-Jones and partial charge parameters for the hydroxyl group were modified relative to the original CHARMM22 values, which are essentially identical to those for TIP3P water. Running the crystal simulations using TIP3P Lennard-Jones and partial charges on the monosaccharide hydroxyl groups lead to no more than 1% improvement in the crystal volume errors, one which was not enough to fix the systematic crystal volume overestimation. Due to the lack of significant improvement and to maintain the internal consistency of the parameter development protocol, the model-compound optimized nonbonded parameters using this final set are listed in Table 9, and show the cell volumes to be overestimated by an average of 5.8%. However, as described below, these parameters give excellent densities for concentrated glucose solutions.

Density and structure of concentrated glucose solutions—The primary milieu for employing the present force field is anticipated to be the aqueous environment, as this is the most important for biomolecular applications, and not the crystalline environment. To further test the developed model in the condensed phase, and in particular, the nonbonded parameters,

we carried out MD simulations at 1 M, 2 M, and 5 M glucose concentrations in water. This was meant to test the transferability of the model compound nonbonded parameters and to determine whether the low densities for monosaccharides were a general problem or particular to the crystal phase.

In each of the three concentrated aqueous solutions, the force field, in combination with TIP3P water molecules, quantitatively reproduces the experimental density (Table 10). Errors in densities are all less than 0.1%, showing that the nonbonded parameters are well-balanced for use in aqueous environments. Because adjusting the nonbonded parameters to increase crystal densities would negatively affect the concentrated aqueous solution densities, we decided to keep the nonbonded parameters as transferred from the model compounds. That is, all the nonbonded parameters from the model compounds were directly transferred unchanged to the full pyranoses. This approach leads to nonbonded parameters that show excellent density behavior in the biologically relevant aqueous environment. In addition, the internal consistency of the force-field with regard to intermolecular interactions is maintained, since the pyranose nonbonded parameters reproduce $V_{\rm m}$, $\Delta H_{\rm vap}$, and $\Delta G_{\rm sol}$ of the model compounds from which they were transferred. This internal consistency is important so as to ensure the proper balance of solute – solute, solute – water, and water – water interaction energies and distances in the liquid phase.

To get a molecular-level view of the interactions in the glucose solutions, we calculated radial distribution functions and number integrals. In the radial distribution functions g(r), there are three types of interactions involving oxygen: $O_{water} - O_{water}$, $O_{glucose} - O_{water}$, and $O_{glucose} - O_{glucose}$. The g(r) for 5 M glucose are representative of all three solutions. $O_{water} - O_{water}$ has the highest peak, followed by $O_{glucose} - O_{water}$, and then $O_{glucose} - O_{glucose}$ (Figure 13a). Peak positions are identical at 1 M and 2 M, and there is very little change in peak heights at these lower concentrations. Also, even at 5 M, the $O_{water} - O_{water}$ peak is nearly identical to that for pure TIP3P water. The smaller peaks for the two other types of interactions reflect the fact that water – water hydrogen-bond networks predominate in these systems. The very small $O_{glucose} - O_{glucose}$ peak implies that, even in very concentrated solutions, the glucose molecules are either interacting with water or water is interacting with itself, to the exclusion of glucose – glucose hydrogen bonding, reflecting the high solubility of glucose in water.

The number integral (NI) at an $O_{water} - O_{water}$ separation distance of 3.3 Å was used to estimate the average number of hydrogen bonds per water molecule⁵⁶ (Figure 13b). In pure TIP3P water, this value is 4.79. For 1M glucose solution, this decreases to 4.52. As the concentration is increased to 2M there is another slight decrease to 4.30. At 5 M, where the concentration of the glucose is near the saturation point, the NI has decreased to 3.72. This decrease is expected as the water molecules participate in interactions with glucose molecules in the system. However, the relatively modest decrease in the NI, along with the pure-water-like $O_{water} - O_{water} g(r)$'s, show that the overall structure of water remains undisturbed by the presence of glucose molecules. These results are in line with neutron scattering experiments under similar conditions,⁵⁷ and provide some evidence suggesting the molecular details of glucose solution in the force field model reflect reality.

Thermodynamics and dynamics of exocylic-group rotation—Prior experiments of aqueous monosaccharide solutions have provided insight into the thermodynamics and dynamics of the exocylic group. The data from these experiments include the relative populations of the exocylic torsion for glucose and galactose, as well as the characteristic time scales on which rotations of the group occur. The developed force field shows proper aqueous behavior for exocyclic rotation in comparison to these experiments, further validating its performance at the molecular level.

The potential of mean force (PMF), calculated with a harmonic biasing potential on the $O_5C_5C_6O_6$ torsion ω , illustrates the free-energy cost associated with rotation of this torsion and is directly related to the relative populations of the local free-energy minima on the torsion reaction coordinate. In the case of glucose, both the α - and β -anomers predominantly populate the minima near $\omega = -60^{\circ}$ (*GG* conformer) and + 60° (*GT*), with these minima having nearly the same free-energy (Figure 14a, b). The third minimum, near 180° (*TG*), is 1.5 kcal/mol higher in free energy and therefore substantially less populated relative to the other two minima. In the case of galactose, the free-energies for *GG* and *TG* are reversed compared to glucose, so that the *GT* and *TG* conformers have the same free-energy while the *GG* state is 1.7 kcal/mol higher in free-energy (Figure 14c, d). Using the relation $-RT\ln(p)$ to calculate the probabilities between each of the free-energy maxima, yields the probabilities associated with each minimum-free-energy well. The probabilities for glucose are in excellent agreement and for galactose in good agreement with recent experimental data⁵⁸ (Table 11).

The conformer populations as calculated directly from 20-ns constant-energy constant-volume (*NVE*) MD agree well with those calculated from the PMFs (Table 11), indicating good sampling of the rotameric conformational space in this time frame. These trajectories show multi-pathway dynamics (Figure 15), as is expected from the PMFs, which all have free-energy barriers to exocyclic rotation of less than 6 kcal/mol (Figure 14). Transitions between the two stable states of glucose, *GT* and *GG*, are mostly direct and go over the barrier at $\omega = 0^{\circ}$, though there are instances of passage through the *TG* intermediate (at 4.7 ns for α -glucose and 6.5 ns for β -glucose in Figure 15). There are numerous *GT* to *TG* transitions, though these quickly return to *GT*, as consistent with the higher free energy for *TG*. Transitions for the two predominant states of galactose, *GT* and *TG*, are mostly direct (over the $\omega = 120^{\circ}$ barrier), though passage through *GG* is also observed (11.8 ns for α -galactose and 3.4 ns for β -galactose).

The force-field dynamics can be compared with experimental measurements of ultrasound relaxation⁵⁹ by assuming a two-state model: *GT* and *GG* for glucose, and *GT* and *TG* for galactose. In this case, the experimental relaxation time, τ is given by

$$\tau = \frac{1}{k_f + k_r},\tag{1}$$

where k_f and k_r are the forward and reverse rate constants from a particular state.⁶⁰ This relaxation time is obtained by simulation using the number correlation function, $C_N(t)^{61}$

$$C_{N}(t) = \frac{\langle \delta N_{A}(0) \delta N_{A}(t) \rangle}{\langle \delta N_{A}^{2} \rangle}, \qquad (2)$$

where

$$\delta N_A(t) = N_A(t) - \langle A \rangle \,. \tag{3}$$

 $N_A(t) = 1$ if the conformer is in state A, and 0 if otherwise, and $\langle A \rangle$ is the average fractional population of state A. For pure 2-state kinetics the number correlation function is a single exponential with the relaxation time given by Eq'n 1. Short-time ("non-phenomenological")

decays do not contribute substantially to the observed relaxation times and are not included when fitting $C_N(t)$.⁶¹ The relaxation time can also be calculated from rates obtained by counting the transitions, though using the correlation function avoids complications associated with inertial effects such as "recoil".⁶²

The decays in $C_N(t)$ for both glucose and galactose are reasonably exponential (Figure 16). As expected from the time series (Figure 15), τ values for glucose are qualitatively longer than for galactose. $C_N(t)$ for passage out of the *GT* glucose conformer includes a short time decay associated with transitions into *TG*; consequently, the first 50 ps was not included in the fitting. Otherwise, correlation functions were fit to a single exponential from $C_N(t) = 1$ to e^{-1} (i.e., one decay time), and τ was averaged for each sugar. This yields $\tau = 600 \pm 100$ ps for glucose and 94 ± 6 ps for galactose. The preceding errors are estimated from $(2\tau/T_{run})^{1/2}$, where T_{run} is the length of the trajectory.⁶³ This formula is also used to estimate the errors in the populations⁶⁴ (Table 11).

Experimentally-derived relaxation times based on fitting to ultrasonic adsorption frequency spectra are 1.94 ± 0.07 ns and 0.57 ± 0.06 ns for glucose and galactose, respectively.⁵⁹ While the trend is correct, simulation underestimates experiment by a factor of 3.2 for glucose and 6.1 for galactose, that is, in the force-field representation transitions occur more often. Much of this error is likely associated with the TIP3P water model, which underestimates the viscosity of water by a factor of $3.^{65}$ Scaling the relaxation times by this factor yields essentially quantitative agreement with experiment for glucose and a 50% underestimate of the transition time for galactose. We note that a 50%, or 2-fold, error in rate corresponds to an error of only

0.4 kcal/mol in the barrier height using the relation $RT\ln(\frac{rate_1}{rate_2})$.66 The simulated values for glucose are also consistent with NMR measurements,⁶⁷ where the spin lattice relaxation times (NT_1) of the exocyclic methylene and the ring carbons are comparable for glucose and a number of other low molecular weight sugars. This implies that the NMR relaxation arises from molecular tumbling (which is approximately 75 ps for glucose68), as opposed to isomerization. If isomerization were on the same time scale or faster than tumbling, NT_1 for the exocyclic methylene would be larger than that of the ring carbons.

Conclusions

The present parameter set for the monosaccharide glucopyranose and its diastereomers extends the CHARMM all-atom additive force field to include these biomolecular building blocks. Importantly, thermodynamic properties of the model compounds used in parameter development, as well as condensed-phase properties of the monosaccharides, show agreement with available experimental data, thereby helping to ensure applicability of the force field in the biologically-relevant condensed phase. Work toward parameters required for modeling the various glycosyl linkages and common chemical modifications of these monosaccharides, as well as for modeling linear sugar alcohols and furanose monosaccharides, is actively underway, with the ultimate goal being development of a consistent and comprehensive parameter set for the modeling of oligosaccharides in conjunction with proteins, nucleic acids, and lipids.

Methods

All QM calculations were done using the Gaussian03 software⁶⁹ and all MM calculations were done with the CHARMM software.⁷⁰ The TIP3P water model⁴² modified for the CHARMM force field⁷¹ was used in all MM calculations.

Molecular Dynamics

MD simulations employed periodic-boundary conditions⁷² with the particle-mesh Ewald method⁷³ for the treatment of Coulomb interactions beyond 12 Å, a force-switching function⁷⁴ to smoothly transition the Lennard-Jones forces and energies to 0 over the range of 10 to 12 Å, and, except in the calculation of solvation free energies, a long-range correction to account for Lennard-Jones interactions beyond the cutoff length.⁷² The long-range correction for solvation free energies was calculated explicitly based on the Lennard-Jones energy difference using the 12 Å cutoff and a 30 Å cutoff. A timestep of 1 fs in conjunction with the "leap-frog" algorithm⁷⁵ was used to integrate the equations of motion, the SHAKE algorithm⁷⁶ was applied to constrain the lengths of covalent bonds to hydrogen atoms to their equilibrium values and to maintain a rigid water geometry by constraining the intramolecular H...H distance, and the Nosé-Hoover thermostat^{77,78} and the Langevin piston barostat⁷⁹ were used to generate the isothermal-isobaric ensemble (NPT) with continuous dynamics. For the constant energy, constant temperature (NVE) simulations, volumes of the systems were chosen to be the average volumes as calculated from NPT simulations at the desired temperature and 1 atm. Equilibration with a heat bath at that temperature was employed prior to the NVE portion of the simulation so that this latter portion had an average temperature equaling the desired temperature.

Automated dihedral fitting

Automated dihedral fitting⁵³ was used to derive dihedral parameters for all model compounds as well as for the hexopyranoses. The quantity to be optimized during automated fitting, the root mean square error (*RMSE*), was defined by a sum over all conformations i

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (w_i (E_i^{QM} - (E_i^{MM, base} + E_i^{MM, dihe}) + c))^2}{\sum_{i=1}^{n} w_i}}$$
(4)

where E_i^{QM} is the QM energy of conformation *i*, $E_i^{\text{MM,base}}$ is the MM energy from the restrained MM optimization of the QM-optimized conformation, $E_i^{\text{MM,dihe}}$ is the dihedral energy of the torsions whose parameters are to be determined, w_i is the weight factor applied to conformation *i*, and *c* is a constant defined by

$$\frac{\partial RMSE}{\partial c} = 0. \tag{5}$$

RMSE was optimized using a Monte Carlo simulated annealing (MCSA) search⁸⁰ in which at each step a random number in the interval of -0.5 to 0.5 kcal/mol was added to each of the dihedral force constants to be optimized, with the constraint that no force constant be less than

-3 kcal/mol or greater than 3 kcal/mol. $E_i^{\text{MM,dihe}}$, which is a function of these force constants, as well as *c*, were then re-computed. These values were then used to calculate *RMSE* at Monte Carlo step *j*, *RMSE_j*. The random change in parameters was accepted or rejected based on the Metropolis criterion⁸¹ with $\Delta E = RMSE_j - RMSE_{j-1}$. The MCSA search was run 10 times with randomized initial values for the dihedral force constants and exponential cooling, and the

value of the best-sampled *RMSE* consistently converged to the same value in all 10 runs (Supplementary Material S4).

The energy V for a particular dihedral is defined by

$$V = \sum_{p=1}^{3} k_p \left(1 + \cos(p * \theta) \right),$$
(6)

where p is the periodicity, k_p is the dihedral force constant associated with that periodicity, and θ is the value of the dihedral angle. No phase angle is included in V; positive values of k_p correspond to a phase angle of 0° and negative values to 180°. Though allowing optimization of the phase angles during MCSA can yield a better fit through the introduction of additional parameters, phase-angle values other than 0° and 180° break the symmetry of the cosine function about 0° and lead to different energies for a molecule and its mirror image, which is non-physical. Each of the k_p 's for each dihedral was allowed to vary independently, with the exception that all the dihedrals involving a hydroxyl rotation were constrained to have the same set of k_p 's. This latter constraint enforced the generation of a general set of hydroxyl rotation parameters that could be applied to all the diastereomers and had the added benefit of simplifying the search through parameter space. Neither MCSA fitting without this constraint nor increasing the number of MCSA steps 2- or 10-fold improved the *RMSE* (not shown), further validating the protocol. Finally, preliminary attempts to fit the monosaccharide conformational energies using the "combo*" O-C-C-O dihedral parameters derived from ethylene glycol suggested that non-uniform weight factors w_i (Eq'n 4) would be required to reproduce the QM target data for ring deformation.⁵³ However, with the "combo" O-C-C-O parameters, which were ultimately used in the final parameterization, uniform weighting of all hexopyranose QM conformational energies was applied, as it not only was the simplest approach, but also yielded very good reproduction of ring deformation conformational energies.

Thermodynamic properties

For the calculation of molecular volumes $V_{\rm m}$ and heats of vaporization $\Delta H_{\rm vap}$, each model compound was placed on the points of an $n \times n \times n$ grid whose dimensions were selected to be greater than two times the cutoff distance at the experimental density of the system. For the larger model compounds n = 6, thus, all heat of vaporization and molecular volume calculations employed at least N = 216 molecules. The heat of vaporization E was calculated as

 $\Delta H_{\text{vap}} = \langle E \rangle_{\text{monomer}} - \frac{\langle E \rangle_{\text{box}}}{N} + RT$, where $\langle E \rangle$ is the average potential energy. $\langle E \rangle_{\text{monomer}}$ was calculated by performing a 20-ps Langevin dynamics^{64,72} gas-phase simulation with an infinite non-bonded cutoff on each of the conformations from the final snapshot of the periodic-box

simulation and averaging all the results. Molecular volumes were calculated as $\frac{\langle V \rangle_{\text{box}}}{N}$ where $\langle V \rangle$ is the average volume.

Solvation free-energies were calculated using a single molecule of solute in a periodic box of 125 water molecules. A previously-developed staging protocol⁸² was used to first "uncharge" the molecule and then to make it "disappear", with 5 ps of equilibration and 50 ps of production MD for each of the 11 stages, with 1 stage each for electrostatic and attractive Lennard-Jones interactions and 9 stages for repulsive Lennard-Jones interactions. Unlike the other *NPT* simulations, the long-range Lennard-Jones contribution was calculated in a *post-hoc* fashion. The solute/solvent system was simulated under the standard cutoff conditions, the Lennard-

Jones energies of the snapshots were re-computed with a long 30-Å cutoff, and the long-range contribution was calculated as the average difference in Lennard-Jones energies using the standard cutoff and the long cutoff; these values agreed quantitatively with an independent earlier study.⁸³ The reported ΔG_{sol} values in Table 1 include this long range correction, though the correction values themselves are included in the Table for informational purposes. The gas-phase leg of the solvation free-energy calculations employed the same staging protocol as the solvated leg.

Monosaccharide – water solution density calculations

Simulations were performed on cubic boxes of 21, 42 and 105 molecules of glucose in 1167 molecules of water corresponding, respectively, to concentrations of 1 M, 2 M, and 5 M. In the simulations for each of the concentrations, a ratio of 1:2 molecules of the α - and β -anomers was used so as to reflect the equilibrium distribution of anomers after mutarotation. For example, in the case of the 1M solution 7 molecules of glucose were of the α type and 14 molecules were of the β type. Additionally, the O₅C₅C₆O₆ dihedral angle was set to a 1:1 mixture of +60° and – 60° to reflect the distribution seen in the exocylic torsion PMF calculations. Each simulation was equilibrated for 600 ps prior to a 5-ns data-collection run.

Monosaccharide crystal calculations

Rectangular-prism unit cells were built using the appropriate transformations to the reduced unit cell. In contrast to other *NPT* simulations in which changes in the system size in response to pressure and temperature were isotropic, the x, y, and z dimensions (a, b, c lattice parameters) could vary independently in response to the target pressure and temperature. The exception to this was α -D-glucose monohydrate (GLUCMH11), which forms a monoclinic crystal, and for which the β angle crystal lattice parameter was allowed to vary along with the a, b, and c edge-length lattice parameters. Analysis of crystalline intramolecular geometries (angles, dihedrals, pucker parameters) was done on time-averaged structures (see Appendix) calculated from the last 4 ns of 5-ns simulations.

Exocylic torsion PMFs

Simulations were carried out using a single molecule of monosaccharide in a truncated octahedron of ~1100 water molecules for an effective concentration of 50 mM. The exocylclic torsion PMF calculation for a monosaccharide involved simulations with a harmonic biasing potential on the dihedral angle ω defined by O₅C₅C₆O₆ and having the form $k(\omega - \omega_0)^2$. The first simulation was started with a monosaccharide conformation having $\omega = -180^{\circ}$ and with ω_0 set to -180°. A 20-ps equilibration involved four 5-ps simulations with k values of 0.003, 0.005, 0.010, and finally 0.025 kcal*mol⁻¹*degree⁻². Data were collected every 0.010 ps from a subsequent 500-ps production simulation with k = 0.025 kcal*mol⁻¹*degree⁻², a value that allowed for routine fluctuations of $\pm 5^{\circ}$ about ω_0 . Using the last snapshot from this simulation, a simulation using the same equilibration and production protocol, except with ω_0 incremented by 10°, was performed. This was continued for a total of 37 simulations (18.5-ns total sampling) with ω_0 ranging from -180° to $+180^\circ$ in 10° -increments for each monosaccharide. The data were unbiased and combined using the Weighted Histogram Analysis Method with a constraint to ensure periodicity of the resultant potential of mean force.^{84,85} Simulations used a truncated octahedron of a size to ensure a minimum of 14 Å between the solute and the nearest edge of the octahedron.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH R01GM070855 (ADM) and F32CA1197712 (OG). The authors wish to thank Dr. Igor Vorobyov and Dr. Victor Anisimov for helpful discussions and to acknowledge a generous grant of computer time from the National Cancer Institute Advanced Biomedical Computing Center.

References

- Lerouxel O, Cavalier DM, Liepman AH, Keegstra K. Curr Opin Plant Biol 2006;9(6):621–630. [PubMed: 17011813]
- 2. DeFronzo RA. Med Clin N Am 2004;88(4):787-835. [PubMed: 15308380]
- 3. Sesti G. Best Pract Res Clin Endoc Metab 2006;20(4):665–679.
- 4. Helenius A, Aebi M. Annu Rev Biochem 2004;73:1019-1049. [PubMed: 15189166]
- 5. Boneca IG. Curr Opin Microbiol 2005;8(1):46-53. [PubMed: 15694856]
- Diks SH, Richel DJ, Peppelenbosch MP. Curr Top Med Chem 2004;4(11):1115–1126. [PubMed: 15279604]
- 7. Dube DH, Bertozzi CR. Nat Rev Drug Discovery 2005;4(6):477-488.
- Hahn-Hagerdal B, Galbe M, Gorwa-Grauslund MF, Liden G, Zacchi G. Trends Biotechnol 2006;24 (12):549–556. [PubMed: 17050014]
- 9. Dietmar P. Biotechnol J 2006;1(7-8):806-814. [PubMed: 16937414]
- 10. MacKerell AD Jr. J Comput Chem 2004;25(13):1584-1604. [PubMed: 15264253]
- Glennon TM, Zheng YJ, Legrand SM, Shutzberg BA, Merz KM. J Comput Chem 1994;15(9):1019– 1040.
- 12. Woods RJ, Dwek RA, Edge CJ, Fraserreid B. J Phys Chem 1995;99(11):3832-3846.
- 13. Ott KH, Meyer B. J Comput Chem 1996;17(8):1068-1084.
- 14. Senderowitz H, Parish C, Still WC. J Am Chem Soc 1996;118(8):2078-2086.
- 15. Durier V, Tristram F, Vergoten G. J Mol Struct: THEOCHEM 1997;395:81-90.
- 16. Momany FA, Willett JL. Carbohydr Res 2000;326(3):194–209. [PubMed: 10903029]
- 17. Momany FA, Willett JL. Carbohydr Res 2000;326(3):210–226. [PubMed: 10903030]
- 18. Kuttel M, Brady JW, Naidoo KJ. J Comput Chem 2002;23(13):1236–1243. [PubMed: 12210149]
- 19. Lii JH, Chen KH, Allinger NL. J Comput Chem 2003;24(12):1504–1513. [PubMed: 12868113]
- 20. Damm W, Frontera A, Tirado-Rives J, Jorgensen WL. J Comput Chem 1997;18(16):1955–1970.
- 21. Kony D, Damm W, Stoll S, van Gunsteren WF. J Comput Chem 2002;23(15):1416–1429. [PubMed: 12370944]
- 22. Reiling S, Schlenkrich M, Brickmann J. J Comput Chem 1996;17(4):450-468.
- Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ. J Comput Chem. 2007 9999(9999), NA.
- 24. 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. J Phys Chem B 1998;102(18):3586–3616.
- 25. MacKerell AD Jr, Feig M, Brooks CL III. J Comput Chem 2004;25(11):1400–1415. [PubMed: 15185334]
- 26. MacKerell AD Jr, Wiórkiewicz-Kuczera J, Karplus M. J Am Chem Soc 1995;117(48):11946–11975.
- 27. Foloppe N, MacKerell AD Jr. J Comput Chem 2000;21(2):86–104.
- 28. MacKerell AD Jr, Banavali NK. J Comput Chem 2000;21(2):105-120.
- Schlenkrich, M.; Brinkman, J.; MacKerell, AD., Jr; Karplus, M. Membrane Structure and Dynamics. Merz, KM.; Roux, B., editors. Birkhauser; Boston: 1996. p. 31-81.
- Feller SE, Yin DX, Pastor RW, MacKerell AD Jr. Biophys J 1997;73(5):2269–2279. [PubMed: 9370424]
- 31. Yin DX, MacKerell AD Jr. J Comput Chem 1998;19(3):334-348.
- 32. Feller SE, MacKerell AD Jr. J Phys Chem B 2000;104(31):7510-7515.

- Feller SE, Gawrisch K, MacKerell AD Jr. J Am Chem Soc 2002;124(2):318–326. [PubMed: 11782184]
- Klauda JB, Brooks BR, MacKerell AD Jr, Venable RM, Pastor RW. J Phys Chem B 2005;109(11): 5300–5311. [PubMed: 16863197]
- 35. Kirschner KN, Woods RJ. Proc Natl Acad Sci U S A 2001;98(19):10541–10545. [PubMed: 11526221]
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. J Am Chem Soc 1995;117(19):5179–5197.
- 37. Jorgensen WL, Maxwell DS, Tirado-Rives J. J Am Chem Soc 1996;118(45):11225–11236.
- Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. J Comput Chem 2004;25(13):1656–1676. [PubMed: 15264259]
- Vorobyov I, Anisimov VM, Greene S, Venable RM, Moser A, Pastor RW, MacKerell AD Jr. J Chem Theory Comput 2007;3(3):1120–1133.
- 40. Vorobyov IV, Anisimov VM, MacKerell AD. J Phys Chem B 2005;109(40):18988–18999. [PubMed: 16853445]
- 41. MacKerell AD Jr, Karplus M. J Phys Chem 1991;95(26):10559-10560.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. J Chem Phys 1983;79(2):926– 935.
- 43. Jorgensen WL, Tirado-Rives J. Proc Natl Acad Sci U S A 2005;102(19):6665–6670. [PubMed: 15870211]
- 44. Rappé AK, Bernstein ER. J Phys Chem A 2000;104(26):6117-6128.
- 45. Møller C, Plesset MS. Phys Rev 1934;46:618–622.
- 46. Dunning TH. J Chem Phys 1989;90(2):1007-1023.
- 47. Guvench O, MacKerell AD Jr. J Phys Chem A 2006;110(32):9934–9939. [PubMed: 16898697]
- Woodcock HL, Moran D, Pastor RW, MacKerell AD Jr, Brooks BR. Biophys J 2007;93(1):1–10. [PubMed: 17554075]
- 49. Hariharan PC, Pople JA. Theor Chim Acta 1973;28:213-222.
- 50. Scott AP, Radom L. J Phys Chem 1996;100(41):16502–16513.
- 51. Allen FH. Acta Crystallogr Sect B-Struct Sci 2002;58:380–388.
- 52. van Alsenoy C, van den Enden L, Schäfer L. J Mol Struct: THEOCHEM 1984;108(1-2):121-128.
- 53. Guvench O, MacKerell AD Jr. J Mol Model. in press.
- 54. Cremer D, Pople JA. J Am Chem Soc 1975;97(6):1354–1358.
- 55. Rao, VSR.; Qasba, PK.; Balaji, PV.; Chandrasekaran, R. Conformation of Carbohydrates. Harwood Academic Publishers; Amsterdam: 1998.
- 56. Chialvo AA, Cummings PT. J Phys Chem 1996;100(4):1309–1316.
- Mason PE, Neilson GW, Barnes AC, Enderby JE, Brady JW, Saboungi ML. J Chem Phys 2003;119 (6):3347–3353.
- Thibaudeau C, Stenutz R, Hertz B, Klepach T, Zhao S, Wu QQ, Carmichael I, Serianni AS. J Am Chem Soc 2004;126(48):15668–15685. [PubMed: 15571389]
- 59. Stenger J, Cowman M, Eggers F, Eyring EM, Kaatze U, Petrucci S. J Phys Chem B 2000;104(19): 4782–4790.
- 60. Andreae JH, Lamb J. Proc Phys Soc B 1951;64(12):1021-1031.
- Chandler, D. Introduction to Modern Statistical Mechanics. Oxford University Press; New York: 1987.
- 62. Loncharich RJ, Brooks BR, Pastor RW. Biopolymers 1992;32(5):523-535. [PubMed: 1515543]
- 63. Zwanzig R, Ailawadi NK. Phys Rev 1969;182(1):280-283.
- 64. Pastor, RW. The Molecular Dynamics of Liquid Crystals. Luckhurst, GR.; Veracini, CA., editors. Kluwer Academic Publishers; The Netherlands: 1994. p. 85-138.
- Feller SE, Pastor RW, Rojnuckarin A, Bogusz S, Brooks BR. J Phys Chem 1996;100(42):17011– 17020.

- 66. Straub, JE. Computational Biochemistry and Biophysics. Becker, OM.; MacKerell, AD.; Roux, B.; Watanabe, M., editors. Marcel Deckker, Inc; New York: 2001. p. 199-220.
- 67. Benesi AJ, Brant DA. Macromolecules 1985;18(6):1109-1116.
- 68. Bogusz S, Venable RM, Pastor RW. J Phys Chem B 2001;105(35):8312-8321.
- 69. 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, Inc. Pittsburgh, PA: 2003.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. J Comput Chem 1983;4(2):187–217.
- 71. Durell SR, Brooks BR, Ben-Naim A. J Phys Chem 1994;98(8):2198-2202.
- 72. Allen, MP.; Tildesley, DJ. Computer Simulation of Liquids. Oxford University Press; Oxford: 1987.
- 73. Darden T, York D, Pedersen L. J Chem Phys 1993;98(12):10089-10092.
- 74. Steinbach PJ, Brooks BR. J Comput Chem 1994;15(7):667–683.
- 75. Hockney, RW. Methods in Computational Physics. Alder, B.; Fernbach, S.; Rotenberg, M., editors. Academic Press; New York: 1970. p. 136-211.
- 76. Ryckaert JP, Ciccotti G, Berendsen HJC. J Comput Phys 1977;23(3):327-341.
- 77. Nosé S. Mol Phys 1984;52(2):255-268.
- 78. Hoover WG. Phys Rev A 1985;31(3):1695-1697. [PubMed: 9895674]
- 79. Feller SE, Zhang YH, Pastor RW, Brooks BR. J Chem Phys 1995;103(11):4613-4621.
- 80. Kirkpatrick S, Gelatt CD, Vecchi MP. Science 1983;220(4598):671-680. [PubMed: 17813860]
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. J Chem Phys 1953;21(6):1087– 1092.
- 82. Deng YQ, Roux B. J Phys Chem B 2004;108(42):16567-16576.
- 83. Shirts MR, Pitera JW, Swope WC, Pande VS. J Chem Phys 2003;119(11):5740-5761.
- Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. J Comput Chem 1992;13(8):1011– 1021.
- 85. Grossfield, A. WHAM: an implementation of the Weighted Histogram Analysis Method. 2003. http://dasher.wustl.edu/alan/wham
- 86. Lide, DR. CRC Handbook of Chemistry and Physics. CRC Press; Boca Raton, FL: 2003.
- 87. Kelly CP, Cramer CJ, Truhlar DG. J Chem Theory Comput 2005;1(6):1133–1152.
- 88. NIST Chemistry WebBook. 2005. http://webbook.nist.gov/chemistry
- 89. Fuchs K, Kaatze U. J Phys Chem B 2001;105(10):2036-2042.
- Computational Chemistry Comparison and Benchmark Database. 2005. http://srdata.nist.gov/cccbdb



Figure 1.

Chemical structures of the diastereomers of the pyranose form of glucose, α -D-glucopyranose: allose (a), altrose (b), galactose (c), glucose (d), gulose (e), idose (f), mannose (g), and talose (h). The β anomer of each of the eight diasteromers differs from the respective α anomer only in the chirality at C₁ (allose carbons have been numbered).





 β -D-glucopyranose (a) and model compounds corresponding to pyranose monosaccharide fragments used in the development of the pyranose monosaccharide parameters: methanol (b), ethanol (c), isopropanol (d), ethylene glycol (e), cyclohexane (f), and tetrahydropyran (g).



Figure 3.

Methanol – water interaction geometries. Hydrogen-bond distances (dashed lines) are those after HF/6-31G(d)-optimization with all other intra- and inter-molecular degrees of freedom constrained, including enforced colinearity of the O-H...O hydrogen bond donor...acceptor angle. The O-H bond vector is constrained to lie on the bisector of the H-O-H angle when methanol is the donor (ai, aii) and on the bisector of the H-O-C angle when water is the donor (bi, bii, biii). The constrained H-O...O-C inter-molecular dihedral angle is rotated as shown.



Figure 4.

Relaxed QM and MM potential energy scans. Scans for the methanol H-C-O-H dihedral (a), the ethanol C-C-O-H dihedral (b), the ethanol H-C-C-O dihedral, and the isopropanol H-C-O-H dihedral (d) were done using the MP2/cc-pVTZ model chemistry ("QM," crosses), the original CHARMM22 parameters²⁴ ("C22," long dashes), and the current parameterization ("carb," short dashes).



vibrational mode

Figure 5.

Vibrational frequencies for methanol in the MP2/6-31G(d) (QM), CHARMM22 (C22), and current parameterization (carb) representations, and the experimental gas-phase infrared frequencies (IR).⁹⁰



Figure 6.

Ethylene glycol QM and MM relative energies. QM energies are using the MP2/cc-pVTZ model chemistry. MM energies are calculated using dihedral parameters fit to only the hydroxyl and O-C-C-O scans ("HOCC/OCCO" fit), to only the 10 minimum energy conformations with 10x weighting for the three lowest energy conformations ("MinE" fit), to both the weighted minimum energy conformations and the scans ("combo" fit), and to the 8-lowest minimum energy conformations with 10x weighting of the lowest three, the entire hydroxyl scan, and the portion of the O-C-C-O scan with dihedral values greater than or equal to 130° ("combo*" fit). The naming convention for the minimum energy conformations refers to the geometries of the H-O-C-C, O-C-O, and C-C-O-H dihedrals. For example, the *tGg'* conformation has these three dihedrals set to *trans, gauche*, and *gauche prime* respectively. All points are relative to the *tGg'* global minimum.



Figure 7.

Relaxed potential energy scan of the C-C-C-C dihedral in cyclohexane at the MP2/cc-pVTZ level (QM) and using the current parameter set (carb).

NIH-PA Author Manuscript



Figure 8.

Relaxed potential energy scans of the C-C-C-C (a), C-O-C-C (b), and O-C-C-C (c) dihedrals in tetrahydropyran at the MP2/cc-pVTZ level (QM) and using the current parameter set (carb).



Figure 9.

Tetrahydropyran – water interaction geometries. The water molecule is constrained to lie in the plane orthogonal to the plane of the tetrahydropyran C-O-C angle, and the O-H...O angle of the hydrogen bond is constrained to 0° . Relative to a virtual particle V lying on the C-O-C angle bisector, the H...O-V angles are 120° (a), 0° (b), and 120° (c).



Figure 10.

Monosaccharide energies in the MP2/cc-pVTZ//MP2/6-31G(d) (red crosses), MM before torsion fitting (dashed blue lines), and MM after torsion fitting (dashed green lines) representations. Both sets of MM data have been RMS aligned to the QM data. Supplementary Material S5 contains the difference energies between the QM and MM data before and after fitting.



Figure 11.

 α -altrose (a) and β -glucose (b) hydroxyl scan energies in the MP2/cc-pVTZ//MP2/6-31G(d) (red crosses), MM before torsion fitting (dashed blue lines), and MM after torsion fitting (dashed green lines) representations. The vertical alignment of the MM data reflects RMS alignment of the entire MM data sets to the QM data set of 1860 monosaccharide conformations (per Figure 10).



Figure 12.

 α -mannose (a) and β -mannose (b) conformational energies in the MP2/cc-pVTZ//MP2/6-31G (d) representation (crosses and solid lines), and the MM representation after torsion fitting (x's and dashed lines).



Figure 13.

Oxygen – oxygen radial distribution functions g(r) (a) and number integrals (b) in concentrated glucose solutions with TIP3P water. Full g(r) data are shown for $O_{water} - O_{water}$ (water-water), $O_{glucose} - O_{water}$ (carb-water), and $O_{glucose} - O_{glucose}$ (carb-carb) in 5 M aqueous glucose, as well as for $O_{water} - O_{water}$ in pure water, along with peak heights in 1 M and 2 M solutions. Number integrals are for $O_{water} - O_{water} - O_{water}$ interactions at the three concentrations and for pure water.



Figure 14.

 α -glucose (a), β -glucose (b), α -galactose (c), and β -galactose (d) free-energy profiles for rotation of the $O_5C_5C_6O_6$ exocyclic torsion calculated in explicit water. Data are at 298 K and 1 atm.



Figure 15.

Time series of the $O_5C_5C_6O_6$ exocyclic torsion ω for the α - and β -anomers of glucose (top two panels) and galactose (bottom two panels). The minima for the *GT* state are shown with dotted lines. The minima for the *GT* state are shown with dotted lines. Supplementary Material S6 plots the data on a [0:360] y-axis.



Figure 16.

Number correlation functions for the *GT* (solid) and *GG* (dashed) conformers of glucose, and the *GT* (solid) and *TG* (dotted) conformers of galactose. Correlation functions were averaged from those of the α - and β -anomers using the weighting 36:64. The negative values at long times for the galactose number correlation functions arise from statistical error.

Thermodynamic properties of the model compounds.

		$V_{\rm m}^{ m a}$		$\Delta H_{\mathrm{vap}}^{a}a$		$\Delta G_{ m sol}{}^{a}~({ m lrc}{}^{b})$	
		Å3	% error ^c	kcal/mol	% error ^c	kcal/mol	error ^c
methanol	expd	67.23		8.95		-5.11	
	C22 ^e	68.78	+2.30	9.12	+1.83	-5.71 (-0.16)	-0.60
	carb^{e}	68.04	+1.20	8.81	-1.61	-5.06 (-0.17)	+0.05
ethanol	exp	96.92		10.12		-5.01	
	C22	98.50	+1.63	10.19	+0.60	-5.22 (-0.26)	-0.21
	carb	95.47	-1.50	10.20	+0.76	-4.93 (-0.25)	+0.08
isopropanol	exp	127.79		10.86		-4.76	
	C22	128.75	+0.75	10.80	-0.58	-4.58 (-0.34)	+0.18
	carb	125.18	-2.05	10.83	-0.27	-4.33 (-0.33)	+0.43
ethylene glycol	exp	92.56		15.70f		-9.30	
	carb	95.64	+3.33	15.79	+0.59	-10.00 (-0.30)	-0.70
cyclohexane	exp	180.598		7.88		1.23	
	carb	182.658	+1.14	7.70	-2.38	0.09 (-0.48)	-1.14
tetrahydro-pyran	exp	162.28		8.26		-3.12	
	carb	163.78	+0.92	8.36	+1.25	-3.48 (-0.44)	-0.36

J Comput Chem. Author manuscript; available in PMC 2010 June 8.

 $^{a}V_{\rm III}$ are at 293.15 K, and $\Delta H_{\rm Vap}$ and $\Delta G_{\rm Sol}$ data are at 298.15 K.

bLong-range Lennard-Jones contribution.

c Errors are calculated as (simulation – experimental) and percent errors (% error) as 100^{*} error/experimental.

 d Experimental (exp) $V_{\rm m}$ and $\Delta H_{\rm Vap}$ are from Reference 86 , and $\Delta G_{\rm sol}$ is from Reference 87 .

 e C22 parameters are from Reference 24 and carb parameters are from the present work.

 $f_{\text{Data from Reference 88}}$.

^gMeasured at 298.15 K.

NIH-PA Author Manuscript

series.
7
ų
3
al
the
I
f
lces
ar
ist
q
pu
g
les
ිවි
Je
G
uc
Ξ
ac
ter
II.
Ы
ato
8
ė
lu
So
-

		energy (kcal/mol)					distance (Å)				
	geometry ^a	$1.16^{*}\mathrm{HF}^{b}$	C22 ^c	error	carb ^c	error	HF -0.20 b	C22	error	carb	error
methanol	ai	-5.73	-6.22	-0.48	-5.88	-0.14	1.83	1.82	-0.01	1.85	0.02
	aii	-5.95	-6.32	-0.37	-5.97	-0.03	1.81	1.82	0.00	1.84	0.03
	bi	-5.70	-5.75	-0.05	-5.59	0.11	1.84	1.85	0.00	1.87	0.03
	bii	-5.93	-6.10	-0.17	-5.91	0.02	1.82	1.83	0.01	1.86	0.04
	biii	-5.46	-6.18	-0.72	-5.97	-0.51	1.85	1.83	-0.02	1.86	0.01
	average			-0.36		-0.11			0.00		0.02
	standard deviation			0.26		0.24			0.01		0.01
ethanol	ai	-5.66	-6.31	-0.64	-6.05	-0.38	1.83	1.82	-0.01	1.84	0.01
	aii	-5.87	-6.42	-0.55	-6.16	-0.28	1.81	1.82	0.00	1.84	0.02
	bi	-5.76	-6.04	-0.28	-5.87	-0.11	1.85	1.85	0.00	1.87	0.02
	bii	-6.03	-6.47	-0.44	-6.27	-0.24	1.83	1.84	0.01	1.86	0.03
	biii	-5.62	-6.66	-1.03	-6.45	-0.83	1.86	1.83	-0.03	1.85	-0.01
	average			-0.59		-0.37			0.00		0.02
	standard deviation			0.28		0.27			0.02		0.02
isopropanol	ai	-5.54	-6.14	-0.61	-5.91	-0.37	1.85	1.85	-0.01	1.87	0.01
	aii	-5.89	-6.27	-0.38	-6.04	-0.15	1.84	1.84	0.01	1.86	0.03
	bi	-5.87	-6.12	-0.25	-5.97	-0.10	1.84	1.85	0.00	1.87	0.02
	bii	-6.12	-6.41	-0.29	-6.23	-0.11	1.82	1.84	0.01	1.86	0.04
	biii	-5.71	-6.65	-0.94	-6.45	-0.75	1.85	1.83	-0.03	1.85	0.00
	average			-0.49		-0.30			0.00		0.02
	standard deviation			0.28		0.27			0.02		0.02

J Comput Chem. Author manuscript; available in PMC 2010 June 8.

 a Molecular geometries are as illustrated in Figure 3.

 $^b\mathrm{HF}$ target energies have been scaled by 1.16 and distances have been shortened by 0.20 Å.

 $^{\rm c}{\rm C22}$ parameters are from Reference 24 and carb parameters are from the present work.

Dipole moments for the global energy minimum conformations of the polar model compounds.

	MP2/6-31G(d)	C22 ^{<i>a</i>}	carb ^a
methanol	1.95 ^b	2.38	2.35
ethanol	1.81	2.36	2.33
isopropanol	1.80	2.39	2.37
ethylene glycol	2.90	n/a	3.37
tetrahydropyran	1.69	n/a	2.03

 a C22 parameters are from Reference 24 and carb parameters are from the present work.

^bDipole moments are in Debye.

Tetrahydropyran solute-water interaction energies and distances.

geometry ^a	energy (kcal/mol)			distance (Å)		
	1.16*HF ^b	carb ^c	error	HF-0.2b	carb ^c	error
a	-5.25	-6.14	-0.89	1.85	1.73	-0.12
þ	-6.07	-6.21	-0.14	1.80	1.71	-0.09
c	-6.36	-6.43	-0.07	1.80	1.72	-0.08
average			-0.37			-0.10
standard deviation			0.45			0.02

 a Molecular geometries are as illustrated in Figure 9.

 $^b\mathrm{HF}$ target energies have been scaled by 1.16 and distances have been decreased by 0.20 Å.

 $^{\ensuremath{c}}$ carb parameters are from the present work.

Optimized pyranose monosaccharide energies and relative CPU times as a function of QM method.

	MP2/cc-pVTZ	MP2/cc-pVDZ	MP2/6-31G(d)	MP2/cc-pVTZ//MP2/6-31G(d)
⊿E (kcal/mol)				
$\alpha glc^a - \beta glc^b$	-1.43	-2.35	-2.45	-1.41
$\beta glc^a - \beta alt^c$	-1.25	0.04	0.17	-1.30
Relative job time ^d				
αglc	16.2	1.5	1.0	1.3 ^e
βglc	29.6	3.7	1.5	1.8 ^e
βalt	19.6	1.9	1.0	1.4 e

^{*a*}α-D-glucose.

 $^{b}_{\beta-D-glucose.}$

 $^{c}\beta$ -D-altrose.

 $^d\mathrm{All}$ jobs used 8 processors in parallel and 20 GB of memory.

 e Job time is the total for MP2/6-31G(d) geometry optimization followed by MP2/cc-pVTZ single-point energy calculation.

Pyranose monosaccharide conformational energies at the MP2/cc-pVTZ//MP2/6-31G(d) level used as the training set for dihedral fitting.

monosaccharide	type of conformational scan ^a	number of conformations
α-altrose	all	450
β-altrose	all	450
α-glucose	all	354
β-glucose	all	330
α-galactose	exocyclic + ring	90
β-galactose	exocyclic + ring	90
β-mannose	C ₂ hydroxyl	24
α-talose	C ₃ hydroxyl	24
β-gulose	C ₄ hydroxyl	24
β-idose	C ₃ hydroxyl	24
total		1860

a "exocyclic" = O5-C5-C6-O6 exocyclic torsion. "ring" = O1-C1-O5-C5, O2-C2-C1-O5, and O4-C4-C5-O5 ring torsions. "all" = "exocyclic" + "ring" + C1, C2, C3, C4, and C6 hydroxyls.

Table 7

_
SS.
Ē
me
eo G
al al
nn 2
nte
•=
Ъ.
٠Ĕ
Ja
Ċ
ဥ
S
B
ē
Ξ
ē
Ξ.
ita
5
J
\sim

average errorb 0.4 -0.1 0.1 average errorb 0.9 0.4 0.1 0.1 standard deviation of errorsb 0.9 0.4 0.7 0.7 dihedral $C_1C_2C_3C_4$ $C_2C_3C_4C_5$ $C_3C_4C_5O_5$ 0.3 average errorb 0.5 -0.9 0.3 0.3	05C1C2 C	$c_1 c_2 c_3$	$C_2C_3C_4$	$C_3C_4C_5$	$C_4C_5O_5$	$C_5O_5C_1$	$C_5 C_6 O_6$
standard deviation of errors 0.9 0.4 0.7 dihedral $C_1C_2C_3C_4$ $C_2C_3C_4C_5$ $C_3C_4C_5O_5$ $C_3C_4C_5O_5$ average error 0.5 -0.9 0.3	0.4	-0.1	0.1	0.0	1.0	6.0-	0.5
dihedral $C_1C_2C_3C_4$ $C_2C_3C_4C_5$ $C_3C_4C_5O_5$ $C_3C_4C_5O_5$ $C_3C_4C_5O_5$ average errorb 0.5 -0.9 0.3 0.3	0.9 0.9	0.4	0.7	0.6	0.8	0.7	0.5
average error b 0.5 -0.9 0.3	$C_1C_2C_3C_4$ C_2C_4	$C_3C_4C_5$	$C_3C_4C_5O_5$	$C_4C_5O_5C_1$	$C_5O_5C_1C_2$	$O_5C_1C_2C_3$	$O_5C_5C_6O_6$
	0.5	6.0-	0.3	-0.2	0.0	0.4	9.0
standard deviation of errors ^{b} 1.7 1.0 1.5	rors <i>b</i> 1.7	1.0	1.3	1.6	1.3	1.6	5.1

^d Data are from simulations of a-allose (Cambridge Structural Database ID: COKBIN), a-galactose (ADGALA01), b-galactose (BDGLOS01), a-glucose (GLUCSA10), a-glucose monohydrate (GLUCMH11), β -glucose (GLUCSE01), α -mannose (ADMANN), and α -talose (ADTAL001). All experimental data and simulations were at 295 K and 1 atm. R-factors for all crystals were ≤ 7 .

 b In degrees.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 8

Monosaccharide o	rrystal r	ing pu	icker p	arameto	ers.
Molecule ^a	Qb	θ	α1	a2	a3
α-allose experiment	0.611	2.9	-38.0	-33.2	-32.9
simulation ^c	0.601	4.4	-38.7	-34.1	-31.4
difference	0.010	-1.4	0.7	0.9	-1.5
α-galactose	0.582	2.1	-34.4	-33.7	-32.3
	0.577	2.0	-35.5	-31.7	-33.3
	0.005	0.2	1.2	-2.0	1.0
α-glucose	0.567	3.6	-32.7	-29.4	-35.5
	0.560	5.4	-32.4	-28.1	-36.6
	0.006	-1.8	-0.3	-1.3	1.1
α-glucose.H ₂ O	0.561	4.9	-31.3	-28.8	-36.3
	0.552	3.6	-32.9	-28.4	-35.3
	0.009	1.3	1.6	-0.4	-1.1
α-mannose	0.575	5.3	-35.4	-32.7	-27.9
	0.566	4.4	-36.4	-32.0	-30.1
	0.008	0.9	1.0	-0.6	2.2
	0.566	2.3	-32.8	-33.1	-33.1
	0.569	3.7	-35.0	-34.0	-34.0
	-0.003	-1.4	2.2	0.8	0.8
α-talose	0.588	3.0	-31.6	-35.0	-35.2
	0.584	2.8	-32.6	-35.3	-34.1
	0.004	0.1	1.0	0.2	-1.2
β-galactose	0.591	5.2	-33.6	-30.8	-38.6
	0.583	5.3	-33.5	-29.9	-38.4
	0.008	-0.1	-0.1	-1.0	-0.2
ß-glucose	0.584	6.9	-32.8	-28.1	-38.8

Molecule ^a	\mathbf{Q}^{h}	θ	α_1	a ₂	a3
	0.553	7.4	-31.9	-26.4	-38.0
	0.031	-0.5	-0.9	-I.6	-0.8
RMS error	0.014	1.1	1.2	1.1	1.4

^a Cambridge Stuctural Datbase IDs are: a-allose, COKBIN; a-galactose, ADGALA01; aglucose, GLUCSA10; a-glucose monohydrate, GLUCMH11; a-mannose, ADMANN; a-talose, ADTALO01; β-galactose, BDGLOS01; and β-glucose, GLUCSE01.

^bQ is in Å, all other values degrees, and the variation between molecules in a unit cell was in the final significant digit shown, i.e. standard deviations no larger than 0.001 Å or 0.2°.

cSimulation values are averaged over the monosaccharide molecules in the unit cell. All have four molecules except α -mannose (8) and α -glucose monohydrate (2).

NIH-PA Author Manuscript

		a (Å)	% error	b (Å)	% error	c (Å)	% error	volume (\mathring{A}^3)	% error
a-glucose	experiment (GLUCSA10 b)	10.37		14.85		4.98		765.9	
	simulationb	10.27	-0.95	14.58	-1.81	5.44	9.42	814.6	6.36
ß-glucose	(GLUCSE01)	9.21		12.64		6.65		774.2	
		9.13	-0.81	12.93	2.26	6.81	2.34	802.8	3.69
α-galactose	(ADGALA01)	5.94		7.87		15.80		738.6	
		6.20	4.34	8.08	2.63	15.85	0.28	7.29T	7.33
β-galactose	(BDGLOS01)	12.66		LT.T		7.70		757.3	
		11.93	-5.74	8.21	5.69	8.04	4.34	786.5	3.85
α-mannose	(ADMANN)	23.45		9.46		6.89		1528.5	
		24.82	5.87	9.65	2.03	6.82	-1.10	1632.5	6.81
β-allose	(COKBIN)	4.92		11.93		12.81		751.0	
		5.19	5.57	12.19	2.19	12.66	-1.17	800.5	6.59
α-talose	(ADTALO01)	8.10		12.13		7.66		752.0	
		8.36	3.22	12.65	4.29	7.49	-2.21	2.067	5.12
α -glucose.H ₂ O	(GLUCMH11)	8.80		5.09		9.71		430.7	
		8.78	-0.27	5.41	6.39	9.88	1.75	460.5	6.93
average			1.40		2.96		1.71		5.83
standard deviation			4.03		2.56		3.79		1.43

J Comput Chem. Author manuscript; available in PMC 2010 June 8.

 a Experimental and simulation data are at 295 K and 1 atm.

^b Cambridge Crystallographic Database ID code. Crystals have four molecules per unit cell, except ADMANN, which has eight.

Table 10

Concentrated glucose solution densities.

concentration (M)	temperature (K)	density (g/cc)		% error
		experiment	simulation	
1	298	1.065 ⁸⁹	1.0664	0.09
2	298	1.132 ⁸⁹	1.1326	0.07
5	293	1.20957	1.2094	0.03

Fractional populations of the conformers of the exocyclic torsion.

		gluco	se		galact	ose
conformer	PMF a	$q \mathrm{MD} p$	experiment ^c	PMF	MD	experiment
$GG(\omega = -60^{\circ})$	0.446^{d} $0.363 \ d$	$0.472 \\ 0.337$	0.40 0.31	0.027 0.026	$0.044 \\ 0.063$	0.03
$GT(\omega = 60^{\circ})$	$\begin{array}{c} 0.514 \\ 0.580 \end{array}$	$0.446 \\ 0.590$	$\begin{array}{c} 0.53\\ 0.61\end{array}$	$0.542 \\ 0.479$	$0.460 \\ 0.486$	$\begin{array}{c} 0.74 \\ 0.72 \end{array}$
$TG\left(\omega = 180^{\circ}\right)$	$0.041 \\ 0.056$	0.063 0.073	0.07 0.08	$0.432 \\ 0.495$	$0.496 \\ 0.452$	$\begin{array}{c} 0.23 \\ 0.25 \end{array}$

 a Data from potential of mean force. Standard deviations in the population estimates from three consecutive 166-ps blocks are < 1.3%.

b Data from unrestrained 20-ns constant energy, constant volume (*NVE*) MD simulations. Statistical error estimates from the relaxation time of the number correlation function are 17% for glucose and 7% for galactose.

 c Experimental data for the 1-methoxy derivatives of the α - and β -anomers are from Reference 58.

 d The first row of data for each conformer is for the α anomer and the second row for the β anomer.