



Published in final edited form as:

*J Phys Chem B*. 2009 September 17; 113(37): 12466–12476. doi:10.1021/jp905496e.

## CHARMM Additive All-Atom Force Field for Aldopentofuranoses, Methyl-Aldopentofuranosides and Fructofuranose

Elizabeth Hatcher, Olgun Guvench, and Alexander D. MacKerell Jr.\*

Department of Pharmaceutical Sciences, 20 Penn Street HSF II, University of Maryland, Baltimore, Maryland 21201

### Abstract

An additive all-atom empirical force field for aldopentofuranoses, methyl-aldopentofuranosides (Me-aldopentofuranosides) and fructofuranose carbohydrates, compatible with existing CHARMM carbohydrate parameters, is presented. Building on existing parameters transferred from cyclic ethers and hexopyranoses, parameters were further developed using target data for complete furanose carbohydrates as well as *O*-methyl tetrahydrofuran. The bond and angle equilibrium parameters were adjusted to reproduce target geometries from a survey of furanose crystal structures, and dihedral parameters were fit to over 1700 quantum mechanical (QM) MP2/cc-pVTZ//MP2/6-31G(d) conformational energies. The conformational energies were for a variety of complete furanose monosaccharides, and included two-dimensional ring pucker energy surfaces. Bonded parameter optimization led to the correct description of the ring pucker for a large set of furanose compounds, while furanose-water interaction energies and distances reproduced QM HF/6-31G(d) results for a number of furanose monosaccharides, thereby validating the nonbonded parameters. Crystal lattice unit cell parameters and volumes, aqueous-phase densities, and aqueous NMR ring pucker and exocyclic data were used to validate the parameters in condensed-phase environments. Conformational sampling analysis of the ring pucker and exocyclic group showed excellent agreement with experimental NMR data, demonstrating that the conformational energetics in aqueous solution are accurately described by the optimized force field. Overall, the parameters reproduce available experimental data well and are anticipated to be of utility in future computational studies of carbohydrates, including in the context of proteins, nucleic acids and/or lipids when combined with existing CHARMM biomolecular force fields.

### Keywords

furanose; furanoside; aldopentose; carbohydrates; ribose; arabinose; fructose; empirical force field

### INTRODUCTION

Furanoses are carbohydrates containing five-membered rings produced from the cyclization of linear aldopentoses or ketohexoses. In solution, furanose carbohydrates exist in equilibrium with the isomeric pyranose form, which is the six-membered ring, with a linear intermediate.<sup>1,2</sup> Due to ring strain, furanoses are typically thermodynamically less favorable in aqueous solution than their pyranose isomers.<sup>3,4</sup> However, furanose moieties still exist in many biological and natural products including DNA, RNA and bacterial cell walls.<sup>5–9</sup> An example

\*Corresponding author: alex@outerbanks.umaryland.edu.

#### Supporting Information Available

The supporting information includes 8 tables, including the CHARMM topology and parameter files and 5 figures. This material is available free of charge via the Internet at <http://pubs.acs.org>

is the bacterial cell wall in *Mycobacterium tuberculosis*, which contains two types of polysaccharide chains, arabinogalactan (AG) and lipoarabinomannan (LAM), both composed of arabinofuranose carbohydrates.<sup>6,9–11</sup> Owing to their biological significance, it is important to understand the structure and dynamics of this class of carbohydrates. Computational methods based on empirical force fields<sup>12</sup> can contribute to this understanding.

One of the major complexities in understanding the structural and dynamic properties of furanose-containing carbohydrates is the inherent flexibility of the five-membered ring.<sup>7,13–16</sup> In pyranose carbohydrates, the most stable ring conformation is a chair conformation, which is in equilibrium with the less favorable boat conformation.<sup>13,17</sup> However, due to the large activation energy for this transition and due to thermodynamic favorability, the pyranose carbohydrates are predominantly in the chair conformation.<sup>13,17</sup> This is not the case for furanose carbohydrates. In furanoses the ring is able to adopt many different conformations due to low energy barriers and small energy differences between ring pucker minima.<sup>7</sup> Moreover, intramolecular hydrogen bonding between hydroxyls on the ring can influence the stabilization/destabilization of different pucker conformations.<sup>14,18</sup> Altona and Sundaralingam developed a pseudorotational model, involving two parameters, the pseudorotation angle  $P$  and amplitude  $\Phi_m$ , which characterizes furanose ring pucker conformations.<sup>19,20</sup> In this model, the furanose is in equilibrium between two ring pucker conformations, a North and South conformation described by the pseudorotational wheel (Figure 1).

In the present manuscript the parametrization of the CHARMM all-atom additive force field for aldopentofuranoses, methyl-aldopentofuranoside (Me-aldopentofuranoside) and fructofuranose is reported. Only the D-series of the sugar compounds are studied, however, it is anticipated that this force field will also be applicable to the L-series sugars. The furanose compounds investigated in this study are shown in Figure 2. The parameters presented are an extension of the recently developed CHARMM carbohydrate force field<sup>21–23</sup> and are compatible with the protein<sup>24–26</sup>, lipid<sup>27,28</sup> and nucleic acid<sup>29,30</sup> force fields. All bonded and nonbonded parameters were initially transferred from the hexopyranose<sup>21</sup> and cyclic ether<sup>31</sup> force fields. The nonbonded parameters proved to be completely transferable, and only slight modifications of the internal bond and angle parameters were necessary to reproduce crystallographic intramolecular geometric data. To properly capture the subtle and complicated conformational energetics of the furanose sugars, torsional parameters were fit to unrestrained quantum mechanical (QM) MP2/cc-pVTZ//MP2/6-31G(d) potential energy surfaces of complete furanose sugar molecules, with a total of 1779 different conformations (this includes the aldopentofuranoses, deoxy-aldopentofuranoses and fructofuranose), using a Monte Carlo simulated annealing (MCSA) procedure.<sup>32</sup> Due to the extent of experimental data for the methyl furanosides, the methylated compounds were also investigated, using *O*-methyl-tetrahydrofuran (*O*-methyl-THF) as a model compound for the optimization of the parameters. *O*-methyl-THF parameters were transferred from the cyclic ether force field<sup>31</sup> and *O*-methyl-tetrahydropyran (THP)<sup>23</sup> and only the parameters involving the methyl functional group required optimization. Final validation of the optimized parameters for the furanoses and methyl furanosides included comparison to QM HF/6-31G(d) furanose-water pair interaction energies and distances, infinite crystal lattice parameters, and aqueous-phase densities and NMR conformational properties.

In the remainder of this paper, details of the parametrization are presented, including a discussion of the challenges encountered during the parametrization process involving the ring pucker.

## METHODS

Empirical force fields calculations were performed using the program CHARMM<sup>33</sup>. Quantum mechanical calculations were performed using Gaussian03<sup>34</sup>. MCSA calculations used were performed using the “fit\_dihedral.py” software<sup>32</sup> available on the MacKerell laboratory web site ([www.mackereell.umaryland.edu/CHARMM\\_ff\\_params.html](http://www.mackereell.umaryland.edu/CHARMM_ff_params.html)). In the MCSA method the selected dihedral parameters were fit simultaneously to minimize the root mean squared error (RMSE),

$$\text{RMSE} = \sqrt{\frac{\sum_i w_i (E_i^{\text{QM}} - E_i^{\text{MM}} + c)^2}{\sum_i w_i}}, \quad (1)$$

where  $E_i^{\text{QM}}$  and  $E_i^{\text{MM}}$  are the QM and MM energies of conformation  $i$ ,  $w_i$  is a weighting factor for conformation  $i$  and  $c$  is a constant that aligns the MM data to the QM data in order to minimize the sum of the squares.

The potential energy function used for the CHARMM<sup>35,36</sup> additive force field is:

$$U(r) = \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{Urey-Bradley}} K_{UB} (S - S_0)^2 + \sum_{\text{dihedrals}} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{\text{impropers}} K_{imp} (\varphi - \varphi_0) + \sum_{\text{nonbonded}} \varepsilon_{ij} \left[ \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}} \quad (2)$$

In Equation 1,  $K_b$ ,  $K_\theta$ ,  $K_{UB}$ ,  $K_\chi$  and  $K_{imp}$  are bond, valence angle, Urey-Bradley, dihedral angle, and improper dihedral angle force constants, respectively, while  $b$ ,  $\theta$ ,  $S$ ,  $\chi$  and  $\varphi$  are the bond distance, valence angle, Urey-Bradley 1,3-distance, dihedral angle and improper dihedral angle, where the subscript zero represents the equilibrium value. In the dihedral potential energy term,  $n$  is the multiplicity and  $\delta$  is the phase angle as in a Fourier series. The nonbonded interaction energy between atoms  $i$  and  $j$  is separated into two terms, the Lennard-Jones (LJ) 6–12 term and the Coulomb term. For the nonbonded terms,  $\varepsilon_{ij}$  is the LJ well depth,  $R_{\text{min},ij}$  is the distance at the LJ energy minimum,  $q_i$  and  $q_j$  are the partial atomic charges, and  $r_{ij}$  is the distance between atoms  $i$  and  $j$ . For the LJ parameters, the Lorentz-Berthelot combining rules are applied.<sup>37</sup>

The nonbonded parameters were transferred from the hexopyranose monosaccharides<sup>21</sup> and cyclic ether<sup>31</sup> force fields and were used to compare the empirical water-furanose pair interaction energies and distances to HF/6-31G(d) QM results. This standard protocol of water-solute interactions was used in order to maintain the transferability with other CHARMM biomolecular force fields.<sup>12</sup> In these calculations, all degrees of freedom were constrained except for the  $\text{H}_{\text{donor}} \dots \text{O}_{\text{acceptor}}$  distance. For comparison, the QM energies are scaled by a factor of 1.16 and the QM distances are offset by  $-0.2 \text{ \AA}$ . These empirical corrections to the QM values have been developed to account for the many-body effects in the condensed-phase, and have consistently been applied in the development of all of the additive all-atom CHARMM biomolecular force fields.<sup>12,26,36,38</sup>

Examples of the water-furanose pairs are shown in Figure 3. In pairs 1 through 4, the furanose monomer geometry was optimized at the MP2/6-31G(d) level, starting from the methyl furanoside crystal structures (methyl at the O1 position on the ring), obtained from the Cambridge Structural Database (CSD)<sup>39</sup>, where the methyl group is replaced with a hydrogen atom. Note that the methyl furanoside crystal structure was used as the Cambridge Structural Database (CSD)<sup>39</sup> only contains methyl furanosides. After geometry optimization of the furanose monosaccharide, TIP3P water<sup>40</sup>-furanose pairs are built in order to calculate the water interaction energies and distances. In both the QM and MM, the water-furanose distance is varied while all other degrees of freedom are constrained. In pairs 1 and 2, one hydrogen atom of the water molecule is in the C-O<sub>ring</sub>-C plane and the other hydrogen is located above and below the plane, respectively. For pairs 3 and 4, the hydrogen of the water molecule is 120 degrees above and below, respectively, the C-O<sub>ring</sub>-C plane. In pairs 5 through 8, the hydroxyl hydrogens in the furanose monomer were constrained with a dihedral of 180° during the QM geometry optimization in order to include alternative hydroxyl conformations in the analysis of the furanose-water interactions. Similar to pairs 1 through 4, pairs 5 through 8 target both the in-plane interactions and the lone pair interactions of the ether oxygen of the furanose with water.

The model compound *O*-methyl-THF was used to develop nonbonded parameters for the methyl functional group of the methyl furanosides. Water interaction energies and distances of *O*-methyl-THF were investigated using nonbonded parameters transferred from the cyclic ether force field and *O*-methyl-THP. A schematic of the water-*O*-methyl-THF interaction pairs, targeting the *O*-methyl oxygen, is shown in Figure S1 of the supplemental material.

Condensed-phase molecular dynamics (MD) simulations were performed in the constant pressure – constant temperature (*NPT*) ensemble using a Nosé-Hoover thermostat<sup>41,42</sup> and a Langevin piston barostat<sup>43</sup>. The reference pressure was set to 1 atm and the reference temperature was set to 298K except as otherwise specified. All of the condensed-phase simulations were performed using a leapfrog integrator<sup>37,44</sup> with a 1 fs timestep. The long-range electrostatic interactions were treated with a Particle Mesh Ewald (PME)<sup>45</sup> method with a real space cutoff of 12 Å. The long-range Lennard-Jones interactions were treated with a force-switched smoothing function<sup>46</sup> in the 10–12 Å range and an isotropic long-range pressure correction<sup>37</sup> was applied beyond the truncation distance. The SHAKE algorithm<sup>47</sup> was used to constrain all hydrogen atom bonds to their equilibrium length and to maintain rigid TIP3P water geometry by constraining a virtual bond connecting the two hydrogen atoms as well as the O-H bonds.

MD simulations of infinite crystals were used for comparison to experimental crystal lattice parameters and intramolecular geometries. The crystal simulations were performed using complete crystal unit cells retrieved from the CSD<sup>39</sup> as the starting structures at the experimental temperature. Only methyl furanoside crystal unit cells were found in the CSD; therefore, simulations were performed on the methyl furanoside compounds. The periodic boundary conditions of the simulation box were determined from the crystal unit cell type and dimensions. Initially, the crystals were minimized for 5000 steps using the steepest descent method and then for 500 steps using an adopted basis Newton-Raphson method<sup>33</sup> to remove any bad contacts. After minimization, the crystals were equilibrated for 100 ps and then a production run was performed for 2 ns.

Aqueous-phase MD simulations were performed to compare to solution experimental densities at varying concentrations and NMR data.<sup>7,48–51</sup> For calculation of the solution densities, simulations were performed on a 50:50 mixture of the  $\alpha$  and  $\beta$  anomers of the furanose in a cubic water box of 1100 TIP3P waters. Each system was first minimized and then equilibrated for 500 ps using harmonic restraints with a force constant of 1\*(particle mass)

\*kcal\*mol<sup>-1</sup>\*Å<sup>-2</sup>\*amu<sup>-1</sup> on only the solute molecules. The production run was performed for 5 ns with all restraints removed. For calculation of the NMR observables, simulations were performed on only the  $\alpha$  or  $\beta$  anomers of the methyl furanoside compounds in a cubic water box of 1100 TIP3P waters. These simulations were performed similarly to those described above with the exception that the production run was performed for 20 ns. The 20 ns simulation length was necessary to reach convergence of the calculated NMR data.

Solution densities  $\rho$  for various concentrations were calculated from the aqueous-phase MD simulations using Equations 3a and 3b.

$$\rho = \frac{N}{\langle V \rangle} \quad (3a)$$

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

In these equations,  $\langle V \rangle$  is the average volume of the solution,  $N_{\text{water}}$  and  $N_{\text{solute}}$  are the number of water and solute molecules, respectively,  $N_{\text{Avogadro}}$  is Avogadro's number and  $\langle MW \rangle$  is the average molecular weight of the solution.

North/South ring pucker conformations and their probability distributions were calculated from the 20 ns aqueous-phase simulations for all of the methyl-aldopentofuranosides investigated in this study. Each of the ring dihedrals were determined every 1 ps and used to calculate the pseudorotation angle  $P$  and amplitude  $\Phi_m$  from Equations 4a and 4b.<sup>19,20</sup>

$$\tan P = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2\nu_2(\sin 36^\circ + \sin 72^\circ)} \quad (4a)$$

$$\Phi_m = \frac{\nu_2}{\cos P} \quad (4b)$$

Here  $\nu_0 = \text{C4-O4-C1-C2}$ ,  $\nu_1 = \text{O4-C1-C2-C3}$ ,  $\nu_2 = \text{C1-C2-C3-C4}$ ,  $\nu_3 = \text{C2-C3-C4-O4}$ , and  $\nu_4 = \text{C3-C4-O4-C1}$ . North and South conformations were calculated from the average values of the time-dependent pseudorotation angle corresponding to the appropriate region of the pseudorotation wheel (Figure 1). The North conformation is any angle smaller than  $90^\circ$  and greater than  $270^\circ$  and the South conformation is any angle between  $90$  and  $270^\circ$ . The percent distribution of the North and South conformations were calculated by creating a pseudorotation histogram with  $2^\circ$  bin widths.

The exocyclic C3-C4-C5-O5 rotamer populations were also obtained from the 20 ns aqueous-phase simulation. During the simulation, the C3-C4-C5-O5 dihedral was monitored every 1 ps and calculated as three probability distributions, corresponding to the *gg* (gauche, gauche), *gt* (gauche, trans) and *tg* (trans, gauche) rotamers (Figure 4). The rotamer probability distributions were calculated from histograms with a bin width of  $2^\circ$ .

## RESULTS AND DISCUSSION

Furanose compounds, used for parameter optimization and validation are shown in Figure 2. As shown, both the hemiacetals and their methyl glycoside derivatives were studied based on the availability of experimental data.

### Parameter Optimization

**Bonds and Angles**—Initial bond and angle parameters for the furanose monosaccharides were transferred from existing hexopyranose monosaccharide and cyclic ether parameters. Performing crystal simulations revealed that only slight modifications of some of the bonds and angles were necessary to achieve accurate geometries. The equilibrium bond length of the CO of the hydroxyl oxygen was shortened by 0.005 Å and the bond length of the CC of the ring carbons was shortened by 0.01 Å. The OCC equilibrium angle containing the ring oxygen and exocyclic carbon was decreased by 1°. These modifications in the equilibrium values led to much better reproduction of the crystal bond lengths and angles. A table containing the experimental crystal bond lengths and angles compared to the calculated results is given in the supplemental materials section (Table S1).

Bond and angle parameters for model compound *O*-methyl-THF were transferred from the cyclic ether force field and *O*-methyl-THP. After performing a CSD crystal database search on methyl furanoside compounds, equilibrium angle parameters were corrected to reproduce the crystal data. The CCO(-methyl) angle was decreased by 2.5° and the OCO(-methyl) angle was increased by 2°. Subsequently, these modifications of the bond and angle parameters of *O*-methyl-THF were transferred to the methyl functional group for the methyl furanoside compounds. However, after performing crystal simulations on the methyl furanosides, another modification was made. Crystal simulations showed that the COC<sub>methyl</sub> angle was too small; therefore, this angle parameter was increased by 1.3° to give an equilibrium angle of 111°. The results of intramolecular geometries calculated from crystal simulations after these modifications are given in Table S1 of the supplemental materials.

Vibrational analysis was performed to validate the force constants of the bonded parameters. The results are given in Table S2 of the supplemental materials. Overall, the empirical frequencies and contributions to the potential energy distribution match well with the QM results; therefore, no further optimization of the frequencies was undertaken. Moreover, the dihedral parameters associated low frequency torsional modes were fit to the QM potential energy scans as described below.

**Dihedrals**—Arabinofuranose and ribofuranose compounds were used as the training set for the dihedral fitting procedure and xylofuranose and lyxofuranose were used to validate the optimized parameters as test set cases. For the deoxyribofuranose and fructofuranose compounds, only the torsions containing the C2 (which for fructofuranose is the anomeric carbon) atom were fit separately, while all other dihedrals were transferred from aldopentofuranose. The dihedrals containing the methyl functional group of the methyl furanoside compounds were transferred from the optimized aldopentofuranose torsional parameters and *O*-methyl-THF, which was also optimized in this study. All optimized torsional parameters were fit to the relaxed QM potential energy scans at the MP2/cc-pVTZ//MP2/6-31G(d) level using the MCSA approach<sup>32</sup>.

Initially, dihedral parameters for the furanose monosaccharides were transferred from the cyclic ethers and hexopyranose monosaccharide force fields. However, due to the flexible nature of the five-membered rings as well as the electrostatics of the hydroxyls, it was necessary to optimize most of the dihedral parameters in the furanose compounds. For arabinofuranose and ribofuranose, C<sub>R</sub>-C<sub>R</sub>-C<sub>R</sub>-C<sub>R</sub>, C<sub>R</sub>-C<sub>R</sub>-C<sub>R</sub>-O<sub>H</sub>, C<sub>R</sub>-C<sub>R</sub>-C<sub>R</sub>-O<sub>R</sub>, C<sub>R</sub>C<sub>R</sub>-O<sub>H</sub>-H, C<sub>R</sub>-C<sub>E</sub>-O<sub>H</sub>-H,

$C_R-O_R-C_R-C_R$ ,  $C_R-O_R-C_R-O_H$ ,  $O_R-C_R-C_R-O_H$  and  $O_R-C_R-O_H-H$  dihedral parameters were fit simultaneously. In the above notation, subscripts R, E and H represent ring, exocyclic and hydroxyl atoms, respectively. After the first optimization of the above dihedrals, further optimizations of the ring dihedrals,  $C_R-C_R-C_R-C_R$ ,  $C_R-C_R-C_R-O_R$  and  $C_R-O_R-C_R-C_R$  were performed, while simultaneously including the  $O_H-C_R-C_R-O_H$  dihedrals in order to improve the description of the ring pucker. A table of all of the aldopentofuranose dihedrals targeted in the optimization procedure is given in Table S3 of the supplemental material. Only conformations that have a relative QM energy below 12 kcal/mol, which includes 1079 different conformations, were used in the fitting procedure. Applying an energy cutoff is acceptable because energies above 12 kcal/mol were associated with ring deformations that are physically unreasonable. A comparison of the QM (black) and MM (blue: unoptimized, red: optimized parameters) conformational energies is given in the upper panel of Figure 5 for all of the 1079 conformations. The bottom panel presents the difference in conformational energies between the QM and the unoptimized (blue) and optimized (red) MM. The *RMSE* calculated from the unoptimized and optimized parameter for all 1079 conformations was 2.04 and 1.15 kcal/mol, respectively. Therefore, the optimization of the dihedral parameters improved the reproduction of the QM conformational energies by approximately 1 kcal/mol, yielding qualitatively better results (Figure 5).

The transferability of the torsional parameters was validated on a collection of conformations of xylofuranose and lyxofuranose. Using the parameters optimized for arabinofuranose and ribofuranose, the *RMSE* of the xylofuranose and lyxofuranose conformational energies was calculated to be 1.82 kcal/mol applying a QM energy cutoff of 12 kcal/mol, indicating that transferring these parameters yields satisfactory results. The QM and MM relative energies for different conformations of xylofuranose and lyxofuranose are shown in Figure S2 of the supplemental materials.

The dihedrals involving atom C2 in deoxyribofuranose were optimized separately from arabinofuranose and ribofuranose. The C2 atom type in deoxyribofuranose is different than that in the aldopentofuranoses, allowing for selective optimization of the C2 dihedral parameters. All other torsional parameters were transferred from the aldopentofuranose compounds. The dihedral parameters were fit to QM MP2/cc-pVTZ//MP2/6-31G(d) conformational energies of the  $\alpha$  and  $\beta$  anomers of deoxyribofuranose, using a relative energy cutoff of 12 kcal/mol (444 conformations). The QM (black) and MM potential energies (blue: unoptimized, red: optimized parameters) are shown in Figure 6. Evident from Figure 6, the optimized parameters significantly improve the agreement of the MM model with the QM data. The *RMSE* improved from a value of 2.07 kcal/mol for the unoptimized parameters to 0.92 kcal/mol following the optimization.

Unlike the aldopentofuranoses, fructofuranose is a ketohexose furanose. Therefore, the IUPAC naming of the atoms change from those in the aldopentofuranose form. In fructofuranose, the C2 atom is the anomeric carbon. The C2 atom in fructofuranose, like that of deoxyribofuranose, has a different atom type allowing the dihedral parameters involving the C2 atom in fructofuranose to be fit separately. All other parameters not involving the C2 atom were transferred from the aldopentofuranoses. The results of the optimization of C2 dihedral parameters are shown in Figure 7. In Figure 7, the QM potential energies (black) for 254 different conformations are plotted along with the MM energies (blue: unoptimized, red: optimized parameters). Without C2 dihedral parameter optimization, the *RMSE* was calculated to be 1.76 kcal/mol as compared to an *RMSE* of 0.92 kcal/mol following optimization. The improvement of the *RMSE* indicated the importance of the optimization of the dihedral parameters involving the C2 atom in fructofuranose.

The dihedral parameters in the *O*-methyl-THF model compound were transferred from the cyclic ether and hexopyranose force fields.  $C_R-C_R-O-C_{\text{methyl}}$  and  $O_R-C_R-O-C_{\text{methyl}}$  dihedral parameters were optimized using MCSA to reproduce the QM conformational energies at the MP2/cc-pVTZ//MP2/6-31G(d) level of theory. In Figure S3 of the supplemental information, a comparison of the QM versus MM conformational energies is given. These optimized dihedral parameters were used for the methyl functional group of the methyl furanosides.

In furanose carbohydrates, the ring can assume a wide range of conformations referred to as puckers. Ring flexibility is a result of low energy barriers between ring conformations and small differences in energy between the ring pucker minima. Owing to the ring flexibility, correctly defining the ring pucker dihedrals is challenging, yet very important. Hence, as previously described above, a second round of optimization was necessary to achieve the correct conformational energy landscape of the ring pucker. In Figure 8, three-dimensional plots of the QM (left) and MM (right) energy surfaces for ring dihedrals  $\phi=O4-C1-C2-C3$  and  $\psi=C1-C2-C3-C4$  for  $\alpha$  (top) and  $\beta$  (bottom) arabinofuranose are shown. Note that in Figure 8 both the QM and MM energies have been zeroed for comparison purposes. In these plots, the MM reproduces the QM ring pucker values and difference in energy between the two lowest energy minima in the North and South regions of the pseudorotation wheel (Figure 1). For  $\alpha$ -arabinofuranose ( $\alpha$ -Ara), the two lowest QM and MM energy minima have  $\phi/\psi$  values of 40/−30 in the North conformation with an energy difference of 0.00 kcal/mol and −30/40 in the South conformation with an energy difference of 0.38 versus 0.65 kcal/mol for the QM and MM models, respectively. Like  $\alpha$ -arabinofuranose, a similar trend was obtained for all of the aldopentofuranoses investigated (See Table S4 in supplemental materials).

### Parameter validation

Validation of the optimized furanose parameters was performed by comparing MM water interaction energies and distances to QM results, comparing results from crystal simulations to experimental crystal lattice parameters, and comparing results from aqueous-phase simulations to furanose solution experimental densities and NMR data. Importantly, analyses of the conformational sampling of the ring pucker and exocyclic dihedrals in solution were performed.

**Water interactions**—Furanose-water interacting pairs, in the orientations shown in Figure 3, were used to test the accuracy of the nonbonded parameters by comparing the MM interaction energies and distances to QM HF/6-31G(d) results. The comparison of these results is given in Table 1. As evident from Table 1, interaction pairs 1 through 4 are less favorable as compared to interaction pairs 5 through 8. Moreover, the MM results for pairs 1 through 4 tend to underestimate the interaction energies. Further examination of the conformations for pairs 1 through 4 (Figure 3, left) showed that the hydroxyl hydrogens on the monosaccharide were pointed towards the ether oxygen, forming weak intramolecular hydrogen bonds. These intramolecular hydrogen bonds compete with the intermolecular hydrogen bonds between the monosaccharide and water. This competition between intra- and intermolecular hydrogen bonding leads to the less favorable QM interaction energies for pairs 1 through 4 as well as leading to a trend where the MM interaction energies are less favorable than the QM values. Therefore, in order to account for the apparent bias associated with the hydroxyl hydrogen orientations, different furanose conformations with the hydroxyl hydrogens in orientations trans to the ether oxygen (pairs 5 through 8) were sampled. MP2 calculations showed these orientations to be local minima. The resulting interactions with water were typically more favorable than for pairs 1 through 4, indicating the importance of accounting for the orientation of the furanose hydroxyls when performing the analysis with water. Using this collection of water-furanose interacting pairs for several different furanoses, the average error in the interaction energies was calculated to be 0.04 kcal/mol and the average error in the interaction



distance to be  $-0.07 \text{ \AA}$ . The *RMSE* in the interaction energy and distance was calculated to be  $0.68 \text{ kcal/mol}$  and  $0.12 \text{ \AA}$ , respectively. The small average error and small *RMSE* in the interaction energies and distances provides evidence that the nonbonded parameters correctly reproduce the hydrogen bond interactions for many different furanose compounds in many different conformations.

Water interactions targeting the O(-methyl) for the methyl furanoside compounds were performed using the model compound *O*-methyl-THF. As mentioned previously, the nonbonded parameters for *O*-methyl-THF were transferred from the cyclic ether force field and *O*-methyl-THP. The *O*-methyl-THF-water pairs were setup using a similar approach to the furanose-water pairs in Figure 3. The water interaction pairs are shown in Figure S1 of the supplemental material. Eight different *O*-methyl-THF-water interaction pairs were used in order to provide a thorough investigation of the nonbonded parameters. A comparison of the QM to the MM *O*-methyl-THF-water interaction energies and distances is given in Table S5 of the supplemental material. The nonbonded parameters for the *O*-methyl oxygen reproduce the QM energies and distance very well with respective average errors of  $0.04 \text{ kcal/mol}$  and  $-0.09 \text{ \AA}$  and *RMSE* values of  $0.37 \text{ kcal/mol}$  and  $0.26 \text{ \AA}$ .

**Crystal Simulations**—Crystal simulations were performed using the methyl furanosides found in the CSD (Table 2). The methyl derivatives of these monosaccharides were chosen because the CSD did not contain the forms of the sugars with a hydroxyl at the anomeric position. The crystal simulations of these compounds were used to test the accuracy of both nonbonded parameters, through comparison of the crystal lattice parameters, and bonded parameters, by examining the conformational sampling of the bonds, angles and dihedrals.

In Table 2, the crystal simulation results are compared to the experimental crystal lattice parameters. The total average percent error in the crystal volume was calculated to be 3.8%, which shows a systematic overestimation. This overestimation has been seen previously in the hexopyranoses<sup>21</sup> and polyols<sup>22</sup>. The total average error in A, B and C unit cell parameters were calculated to be 4.7%,  $-0.7\%$  and  $0.6\%$ , respectively. Although the results show a slight overestimation of the crystal volume and unit cell length A, the optimized parameters reproduce the experimental crystal parameters in a satisfactory fashion.

After a slight modification of the bond and angle parameters, discussed in the parameter optimization section above, the values were monitored and compared to the experimental crystal data. Note that the dihedral parameters were not adjusted based on the crystal simulations because they were optimized using the fitting procedure discussed above. In Table S1 of the supplemental materials, a few representative bond length, valence angle, and dihedral angle values are compared to the experimental data. All of the average bond lengths are within  $0.01 \text{ \AA}$  of the experimental values, all of the average angles are within 1.3 degrees of the experimental values and all of the average dihedrals are within 6.4 degrees. In addition, comparison of the pseudorotation angles and amplitudes, calculated from equations 4a and 4b, from the MD simulations shows them to be in excellent agreement with the experimental values (Table S6 in supplemental material). These results speak to the quality of the force field with respect to reproducing the intramolecular geometries of the furanoses.

**Aqueous-Phase Solution Densities**—Solution densities of monosaccharide solute molecules in 1100 TIP3P waters, at varying concentrations, were calculated from the aqueous-phase simulations using Equations 3a and 3b. In Table 3, the simulation results are compared to the experimental results at 298K and 1atm.<sup>48</sup> For all compounds and concentrations studied, the largest percent error is within 1.4% of the experimental density values and the total average error for all compounds is 1.2%. While there is a tendency for the calculated densities to be systematically larger than the experimental data, the differences being in the vicinity of 1%

may be considered satisfactory. Also, the percent error decreases with increasing concentration, suggesting that the error in the densities may decrease further with more concentrated solutions.

### Conformational Sampling in Aqueous Solution

**Ring pucker**—The pseudorotation angles and amplitudes for all of the aldopentofuranoses were calculated as a function of time from the aqueous phase simulations using equations 4a and 4b. A plot of the time series of the pseudorotation angle for one molecule of  $\alpha$ -arabinofuranose is shown in Figure S4 of the supplemental material. Apparent in Figure S4, transitions between North and South ring pucker conformations occur rapidly. Moreover, this occurs with all of the aldopentofuranoses (not shown), consistent with the low barrier pathway connecting the two pucker conformations as seen in the QM studies (Figure 8). With regard to thermodynamics, pseudorotation angle histograms were calculated to obtain the North and South conformation values and populations. The average pseudorotation angle and North versus South conformer populations for each half of the trajectory were calculated to test the convergence of the simulations. Both the angles and the populations for each of the halves were similar to those of the full simulation indicating that adequate convergence was achieved on the 20 ns timescale.

Presented in Table 4 are the average pseudorotation angles and amplitudes (top) and the North versus South pseudorotation angle populations (bottom) calculated from the aqueous phase simulations of all methyl aldopentofuranosides, along with the corresponding experimental data. The experimental values were obtained from the combination of the PSEUROT program and experimental NMR couplings.<sup>14,49,50</sup> The calculated amplitudes,  $\Phi_m$ , showed little variability having values between  $38^\circ$  and  $41^\circ$  when calculated for all furanoses. This is in the range of the physically reasonable puckers as judged by the CSD structures in Table 2, where the amplitudes ranged from 38 to 44 (Table S6 of supplemental materials). Concerning average pseudorotation angles for the North and South conformers (Table 4), the calculated values are often in satisfactory agreement with the experimental estimates (ie. within  $25^\circ$ ), but still some significant differences exist.

In many cases the poor agreement occurs with the weakly populated puckers (bottom section of Table 4), such as with the North ring puckers of Me- $\alpha$ -ribofuranoside and Me- $\alpha$ -xylofuranoside and with the South ring puckers of Me- $\beta$ -ribofuranoside, Me- $\alpha$ -lyxofuranoside, Me- $\beta$ -lyxofuranoside, and Me- $\beta$ -xylofuranoside. In some of these cases the reported experimental pseudorotation angles are in the opposite hemisphere (e.g. a value of  $119^\circ$  was reported for North conformer of Me- $\alpha$ -ribofuranoside) or in the unfavorable West region at  $\sim 270^\circ$ . These results suggest that the experimental estimates, which involve inputting 3-bond J-coupling constants as a function of temperature into the PSEUROT program, may be poorly defined in some cases.

In practice there may be a number of problems that limit comparison of the calculated and experimental pseudorotation angles and probabilities. During the simulations, the calculated pseudorotation angles take on a range of values (Figure S4) beyond sampling only the accepted North and South puckers, behavior that may be representative of the experimental regimen. This is due to interconversion between the minimum energy ring puckers and the sampling of low-energy intermediates. In fact, in some cases such as Me- $\alpha$ -arabinofuranoside, Me- $\alpha$ -lyxofuranoside and Me- $\beta$ -xylofuranoside a single distribution of puckers sampling both the North and South conformations is obtained (Figure S5); similar distributions have been observed with the GLYCAM force field.<sup>51,52</sup> As the North versus South distributions are not well defined in such situations, calculation of both the pseudorotation angle average and populations based on the ranges specified in the methods section may be a poor approximation. If such distributions are indeed present it may adversely affect the pseudorotation values calculated from the PSEUROT approach. PSEUROT assumes a two-state equilibrium between

the North and South conformations,<sup>50</sup> a situation that the present calculations indicate is not occurring in all cases. Another possible limitation of the PSEUROT program is an inability to correctly define the minor conformation when it is only weakly populated. These problems may limit the ability to quantitatively compare the simulated and experimental pseudorotation data, though the simulations qualitatively reproduce a majority of the experimental data.

In contrast, results of the pseudorotation angle and amplitude calculated from the crystal simulation are in very good agreement with the values calculated from the experimental crystal structure. In Table S6 in the supplemental materials, all of the computed pseudorotation angles are within 20° of the experimental values. Also, the amplitudes calculated from the crystal simulations are in excellent agreement with the experimental values with absolute deviations no larger than 1. In sum, these results provide evidence that the optimized parameters adequately reproduce the experimental pseudorotation angles and amplitudes in both solution and in the crystals.

Included in the bottom portion of Table 4 are the experimental and calculated North and South pseudorotation angle populations. In all cases except Me- $\alpha$ -xylofuranoside the calculated relative populations of North versus South are qualitatively in good agreement with the experimental estimates. Quantitatively, both over and undersampling of the south conformation is observed, suggesting that a satisfactory balance in the sugar puckering conformational energetics has been achieved. The only qualitative exception is Me- $\alpha$ -xylofuranoside where the model predicts that the North hemisphere will dominate in contrast to the experimental estimate. Despite this discrepancy, the overall ability of the force field in reproducing the North versus South experimental pucker populations indicates its quality with respect to the treatment of furanose ring conformational sampling in aqueous solution.

**Exocyclic group**—From the aqueous-phase simulations of Me- $\alpha$ - and Me- $\beta$ -arabinofuranoside, the C3-C4-C5-O5 torsions were monitored every 1 ps and a histogram of the dihedral was computed. From the histogram of each Me-arabinofuranoside, the probability distribution was calculated; the results are presented in Table 5 and compared to the experimental estimates calculated from NMR data.<sup>7,11,51,52</sup> The rotamer distributions calculated from the simulations reproduce the experimental results very well, including the trend of the rotamer populations. Comparable to experimental results, the calculated MM results show that the *gg* rotamer for  $\alpha$ -Me-arabinofuranoside is more populated than the *gt* rotamer, which is more populated than the *tg* rotamer. Moreover, for the  $\beta$ -Me-arabinofuranoside epimer, the calculated MM results match those of the experiment, showing that the *gt* rotamer is more populated than the *gg* and that *gt* is the more weakly populated rotamer. Although the MM results reproduce the trends in the ordering of the population well, the results show a tendency to underestimate the population of the *tg* rotamer. This underestimation is compensated in the *gt* rotamer; however, as the *tg* rotamer is only weakly populated this deficiency in the model was not considered significant. In previous studies, the rotameric distributions about the C4-C5 bond for the two epimers of Me-arabinofuranoside were calculated using the GLYCAM carbohydrate force field.<sup>51,52</sup> While they were able to calculate the correct trend in the populations for the alpha epimer, they were not able to achieve the correct qualitative results for the beta epimer using the TIP3P water model. However, using a more sophisticated water model they were able to obtain the correct qualitative results. Unlike the GLYCAM force field, our optimized parameter set for furanose reproduces the qualitative trend in the rotameric distribution for both epimers of Me-arabinofuranoside in the TIP3P water model.

## CONCLUSIONS

In this investigation the additive all-atom CHARMM force field for aldopentofuranose and fructofuranose is presented. Initially, parameters for the furanose compounds were transferred from the cyclic ether and hexopyranose force fields. Subsequently, the bonded parameters were optimized using crystal data and QM data at the MP2/cc-pVTZ//MP2/6-31G(d) level of theory. Torsion parameters were fit to QM data using an automated MCSA procedure developed in this group. After validating the nonbonded parameters by comparing MM to QM monosaccharide-water interaction energies and distances, no further optimization of the nonbonded parameters was deemed necessary. In order to test the accuracy of the optimized parameters results from condensed-phase simulations were compared to experimental crystal data, aqueous-phase densities and NMR data. Pseudorotation angles and North/South conformation populations calculated from the aqueous-phase simulations satisfactorily reproduce the experimental results calculated from a combination of the PSEUROT program and NMR coupling. Notably, distributions of the pseudorotation values from the simulations indicate that assumptions used in the PSEUROT program may limit its ability to derive quantitative estimates of the pseudorotation distributions from the NMR data. From the simulations, exocyclic rotamer populations were computed and compared to experimental rotamer population calculated from NMR data. The calculated rotamer populations agree quite well with the experimental results, indicating the quality of the developed force field in treating the conformational properties of the flexible furanose ring and its exocyclic substituent.

One of the difficulties faced in the parametrization of these five-membered ring compounds was the inherent flexible nature of the ring. Due to low energy barriers between the minima in the ring pucker dihedral surfaces (eg. Figure 8), the furanose rings can take on different conformations (i.e. North and South conformations) as well as make rapid transitions between those conformations (See Figure S4). In order to account for this flexibility, the parameters must adequately reproduce the relative energies of all the accessible ring pucker conformations. Moreover, the ring pucker torsions are not independent of the other torsions in these compounds. Fitting to a large number of conformations for a range of different compounds while retaining the transferability of the parameters can be very challenging. This was overcome in the present study using an MCSA method, which allowed the simultaneous fitting of large numbers of conformations. This gave a general set of parameters that capture the conformational energetics of the furanose carbohydrates well.

Overall, the optimized parameters reproduce the structural and dynamical properties of the furanose monosaccharide compounds, as evident from the condensed-phase results. Conformational properties are in very good agreement because the fitting procedure was performed on a large number of conformations of many different compounds to reproduce high-level QM MP2/cc-pVTZ//MP2/6-31G(d) energies, a QM level of theory that has been shown to give excellent results for these types of systems.<sup>21,23,53,54</sup> Since the parameter optimization was done using a procedure compatible with that used for the other CHARMM additive biomolecular force fields, the carbohydrate force field, including the furanoses, can be used in conjunction with the protein, nucleic acid and lipid force fields. This will facilitate the simulation of heterogeneous biological systems involving carbohydrates, including glycoproteins and glycolipids and help to provide atomistic-detail information regarding their structural, dynamic, thermodynamic, and energetic properties.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

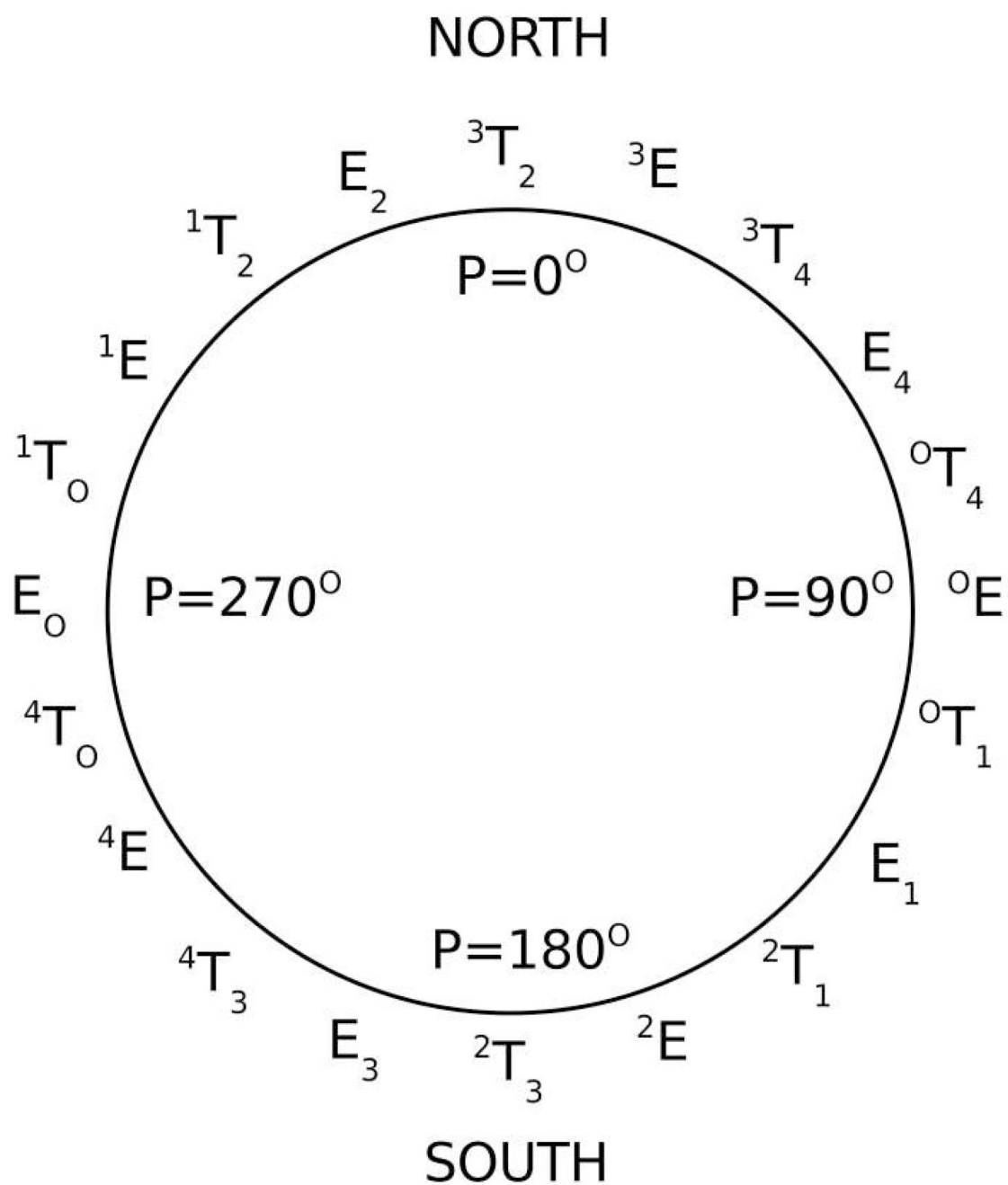
## Acknowledgements

This work was supported by NIH GM070855 (ADM) and F32CA119771 (OG). The authors acknowledge computer time and resources from the National Cancer Institute Advanced Biomedical Computing Center and Department of Defense High Performance Computing.

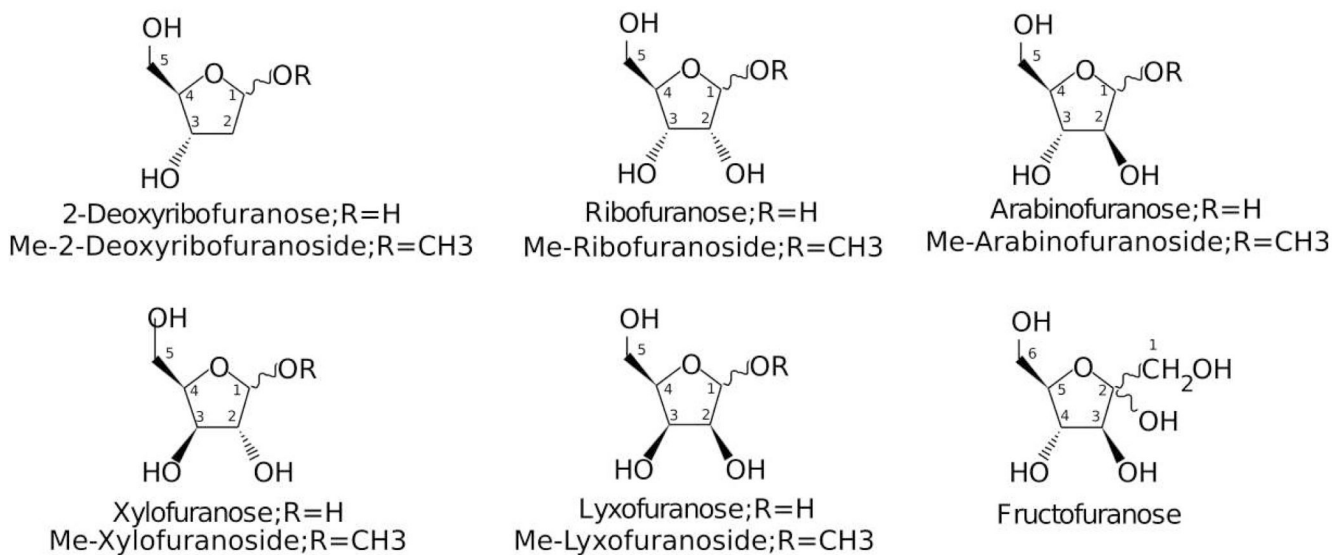
## References

1. Harmon RE. *Asymmetry in Carbohydrates* New York. 1979
2. Guler LP, Yu Y-Q, Kenttämä HI. *J. Phys. Chem. A* 2002;106:6754.
3. Goldberg RN, Tewari YB. *J. Phys. Chem. Ref. Data* 1989;18:809.
4. Mackie W, Perlin AS. *Can. J. Chem* 1966;44:2039.
5. Saenger, W. *Principles of Nucleic Acid Structure*. Berlin: Springer-Verlag; 1988.
6. Brennan PJ. *Rev. Infect. Dis* 1989;2:5420.
7. Arjona O, Gómez AM, López JC, Plumet J. *Chem. Rev* 2007;107:1919. [PubMed: 17488060]
8. Pigman, W., editor. *The Carbohydrates. Chemistry and Biochemistry*. Vol. Vol. 1A. New York: Academic Press; 1972.
9. Lowary TL. *Curr Opin Chem Biol* 2003;7:749. [PubMed: 14644185]
10. Crick DC, Mahapatra S, Brennan PJ. *Glycobiology* 2001;11:107R.
11. D'Souza FW, Ayers JD, McCarren PR, Lowary T. *J. Am. Chem. Soc* 2000;122:1251.
12. MacKerell AD Jr. *J. Comput. Chem* 2004;25:1584. [PubMed: 15264253]
13. Rao, VSR.; Qasba, PK.; Balaji, PV.; Chandrasekaran, R. *Conformation of Carbohydrates*. Amsterdam: Harwood Academic Publishers; 1998.
14. Gordon MT, Lowary TL, Hadad CM. *J Org Chem* 2000;65:4954. [PubMed: 10956478]
15. Levitt M, Warshel A. *J. Am. Chem. Soc* 1978;100:2607.
16. Sundaralingam M. *J. Am. Chem. Soc* 1965;87:599. [PubMed: 14263384]
17. Angyal SJ. *Angew. Chem* 1969;8:157.
18. French AD, Dowd MK, Reilly PJ. *THEOCHEM* 1997;395:271.
19. Altona C, Sundaralingam M. *J. Am. Chem. Soc* 1972;94:8205. [PubMed: 5079964]
20. Altona C, Sundaralingam M. *J. Am. Chem. Soc* 1973;95:2333. [PubMed: 4709237]
21. Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, MacKerell AD Jr. *J. Comput. Chem* 2008;29:2543. [PubMed: 18470966]
22. Hatcher ER, Guvench O, MacKerell AD. *J. Chem. Theory. Comput* 2009;5:1315.
23. Guvench O, Hatcher ER, Venable RM, Pastor RW, MacKerell AD. submitted.
24. MacKerell AD Jr, Feig M, Brooks CL III. *J. Am. Chem. Soc* 2004;126:698. [PubMed: 14733527]
25. MacKerell AD Jr, Feig M, Brooks CL III. *J. Comput. Chem* 2004;25:1400. [PubMed: 15185334]
26. 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;102:3586.
27. Feller SE, MacKerell AD Jr. *J. Phys. Chem. B* 2000;104:7510.
28. Feller SE, Gawrisch K, MacKerell AD Jr. *J. Am. Chem. Soc* 2002;124:318. [PubMed: 11782184]
29. Foloppe N, MacKerell AD Jr. *J. Comput. Chem* 2000;21:86.
30. MacKerell AD Jr, Banavali NK. *J. Comput. Chem* 2000;21:105.
31. Vorobyov I, Anisimov VM, Greene S, Venable RM, Moser A, Pastor RW, MacKerell AD Jr. *J. Chem. Theory. Comput* 2007;3:1120.
32. Guvench O, MacKerell AD. *J. Mol. Model* 2008;14:667. [PubMed: 18458967]
33. Brooks BR, Brooks CL III, MacKerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RV, Woodcock HL, Wu X, Yang W, York DM, Karplus M. *J. Comp. Chem* 2009;30:1545. [PubMed: 19444816]

34. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA.; Vreven, T., Jr; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, K.; Kitao, O.; Nakai, H.; Klene, M.; Li, TW.; Knox, JE.; Hratchian, HP.; Cross, JB.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Ayala, PY.; Morokuma, K.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Zakrzewski, VG.; Dapprich, S.; Daniels, AD.; Strain, MC.; Farkas, O.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Ortiz, JV.; Cui, Q.; Baboul, AG.; Baboul, AG.; Clifford, S.; Cioslowski, J.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Challacombe, M.; Gill, PMW.; Johnson, B.; Chen, W.; Wong, MW.; Gonzalez, C.; Pople, JA. Gaussian 03; Revision B.04 ed. Pittsburgh, PA: Gaussian, Inc.; 2003.
35. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. *J. Comput. Chem* 1983;4:187.
36. MacKerell, AD., Jr; Brooks, B.; Brooks, CL., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. Schleyer, PvR; Allinger, NL.; Clark, T.; Gasteiger, J.; Kollman, PA.; Schaefer, HF., III; Schreiner, PR., editors. Vol. Vol. 1. Chichester: John Wiley & Sons; 1998. p. 271
37. Allen, MP.; Tildesley, DJ. *Computer Simulation of Liquids*. Oxford: Oxford University Press; 1987.
38. MacKerell AD Jr, Karplus M. *J. Phys. Chem* 1991;95:10559.
39. Allen FH. *Acta Crystallogr. Sect. B-Struct. Sci* 2002;58:380.
40. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J. Chem. Phys* 1983;79:926.
41. Hoover WG. *Phys. Rev. A* 1985;31:1695. [PubMed: 9895674]
42. Nosé S. *Mol. Phys* 1984;52:255.
43. Feller SE, Zhang YH, Pastor RW, Brooks BR. *J. Chem. Phys* 1995;103:4613.
44. Hockney, RW. The potential calculation and some applications. In: Alder, B.; Fernbach, S.; Rotenberg, M., editors. *Methods in Computational Physics*. Vol. Vol. 9. New York: Academic Press; 1970. p. 136
45. Darden T, York D, Pedersen L. *J. Chem. Phys* 1993;98:10089.
46. Steinbach PJ, Brooks BR. *J. Comput. Chem* 1994;15:667.
47. Ryckaert JP, Ciccotti G, Berendsen HJC. *J. Comput. Phys* 1977;23:327.
48. Galema SA, Høiland H. *J. Phys. Chem* 1990;95:5321.
49. Houseknecht JB, Altona C, Hadad CM, Lowary TL. *J Org Chem* 2002;67:4647. [PubMed: 12098271]
50. Houseknecht JB, Lowary TL, Hadad CM. *J. Phys. Chem. A* 2003;107:372.
51. Seo M, Castillo N, Ganzynkiewicz R, Ganiels CR, Woods RJ, Lowary TL, Roy P-N. *J. Chem. Theory. Comput* 2008;4:184.
52. Taha HA, Castillo N, Roy P-N, Lowary TL. *J. Chem. Theory. Comput* 2009;5:430.
53. Woodcock HL, Moran D, Pastor RW, MacKerell AD Jr, Brooks BR. *Biophys. J* 2007;93:1. [PubMed: 17554075]
54. Guvench O, MacKerell AD Jr. *J. Phys. Chem. A* 2006;110:9934. [PubMed: 16898697]
55. Humphrey W, Dalke A, Schulten K. *J. Mol. Graph* 1996;14:33. [PubMed: 8744570]

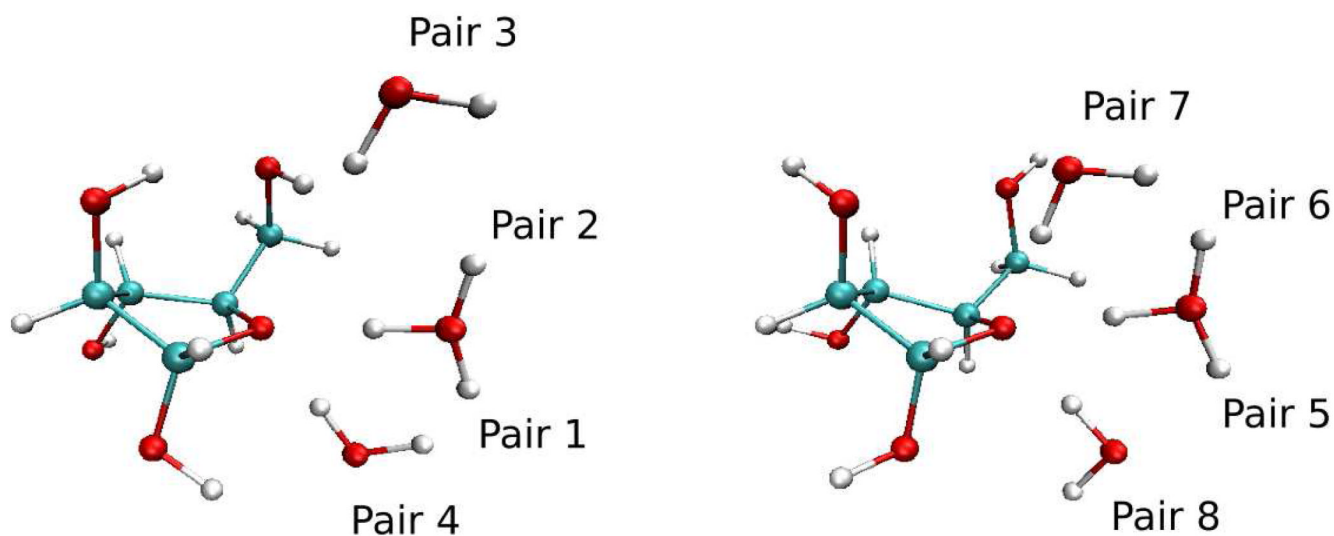


**Figure 1.** Pseudorotational wheel developed by Altona and Sundaralingam<sup>19,20</sup>

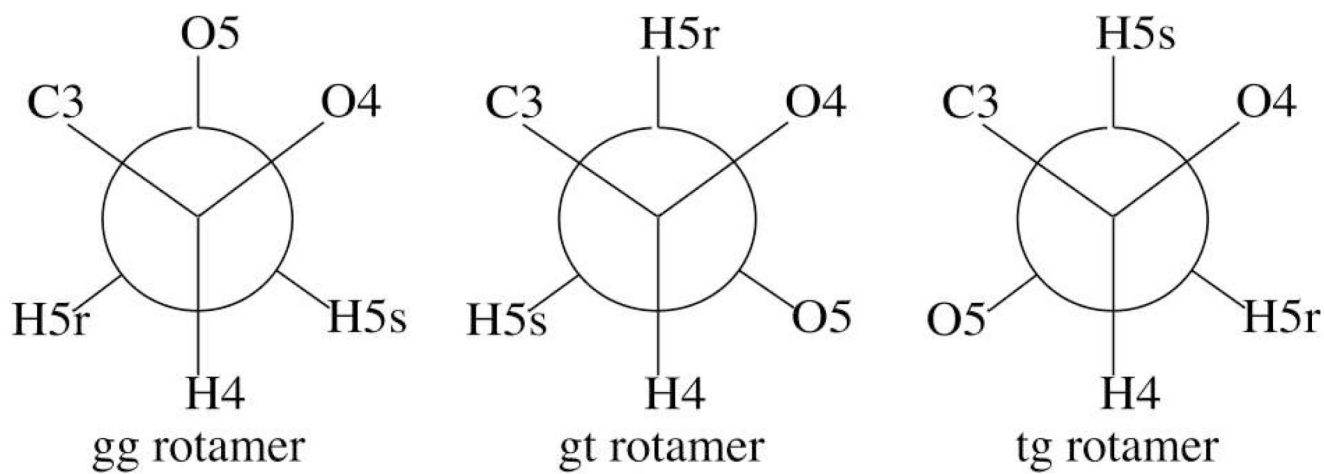


**Figure 2.**  
Furanose compounds parametrized for the CHARMM carbohydrate force field.

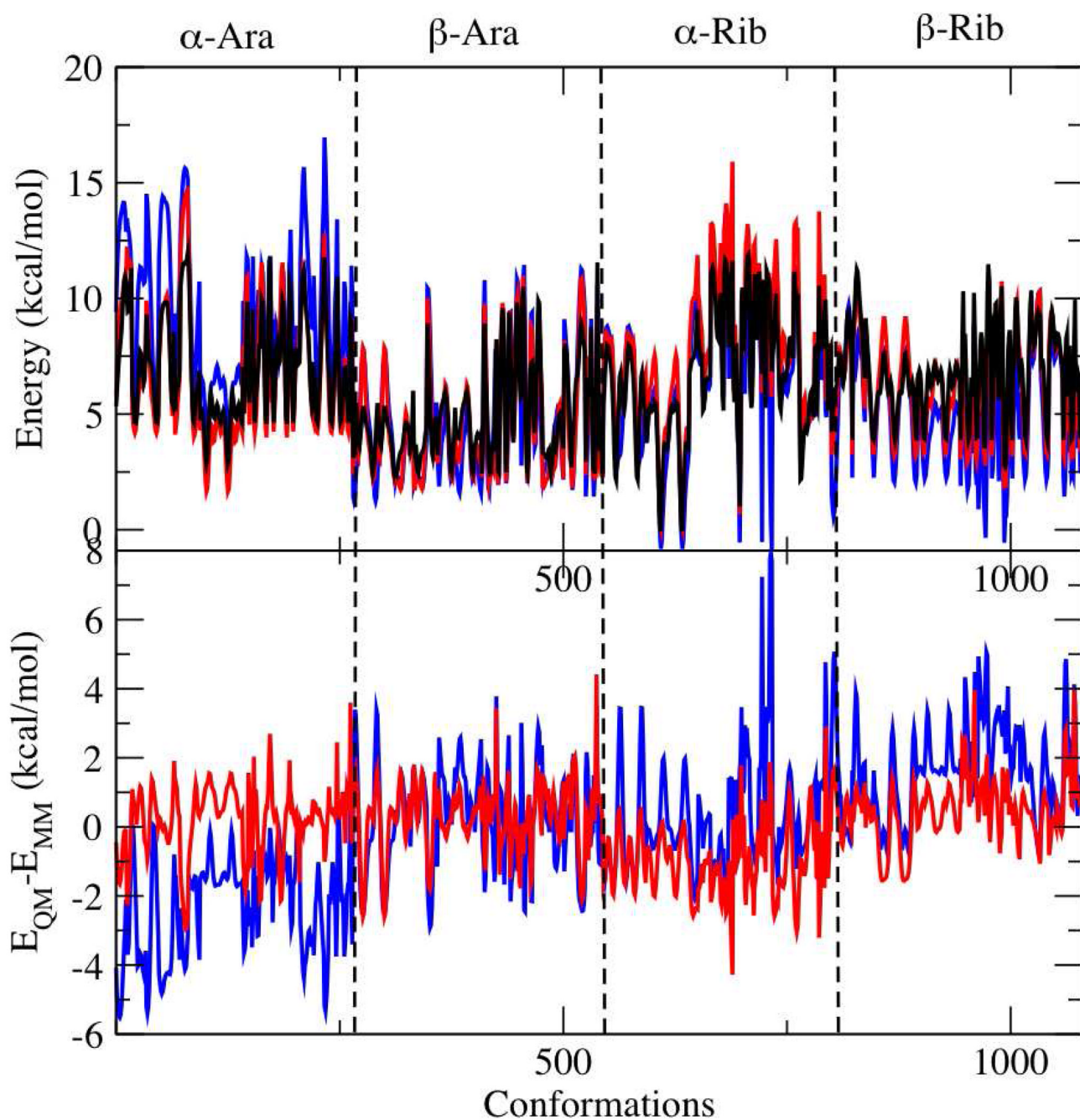




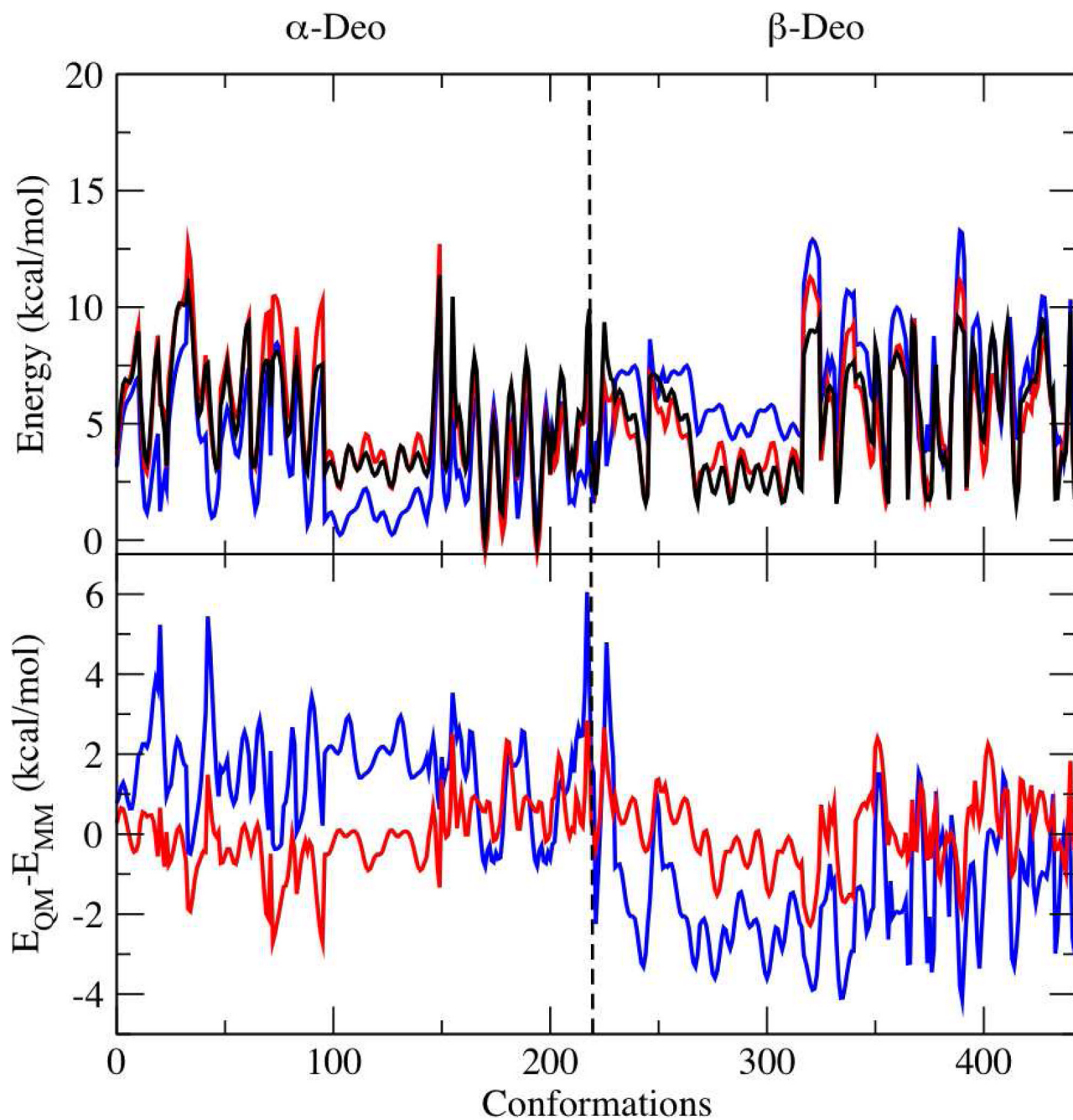
**Figure 3.** Schematic of the water-arabinofuranose interaction pairs. In pairs 5 through 8 the hydroxyl hydrogens are restrained away from the ether oxygen. For simplicity the water molecule orientations are shown simultaneously; actual calculations employ only a single water molecule. Molecular graphics were prepared with VMD.<sup>55</sup>



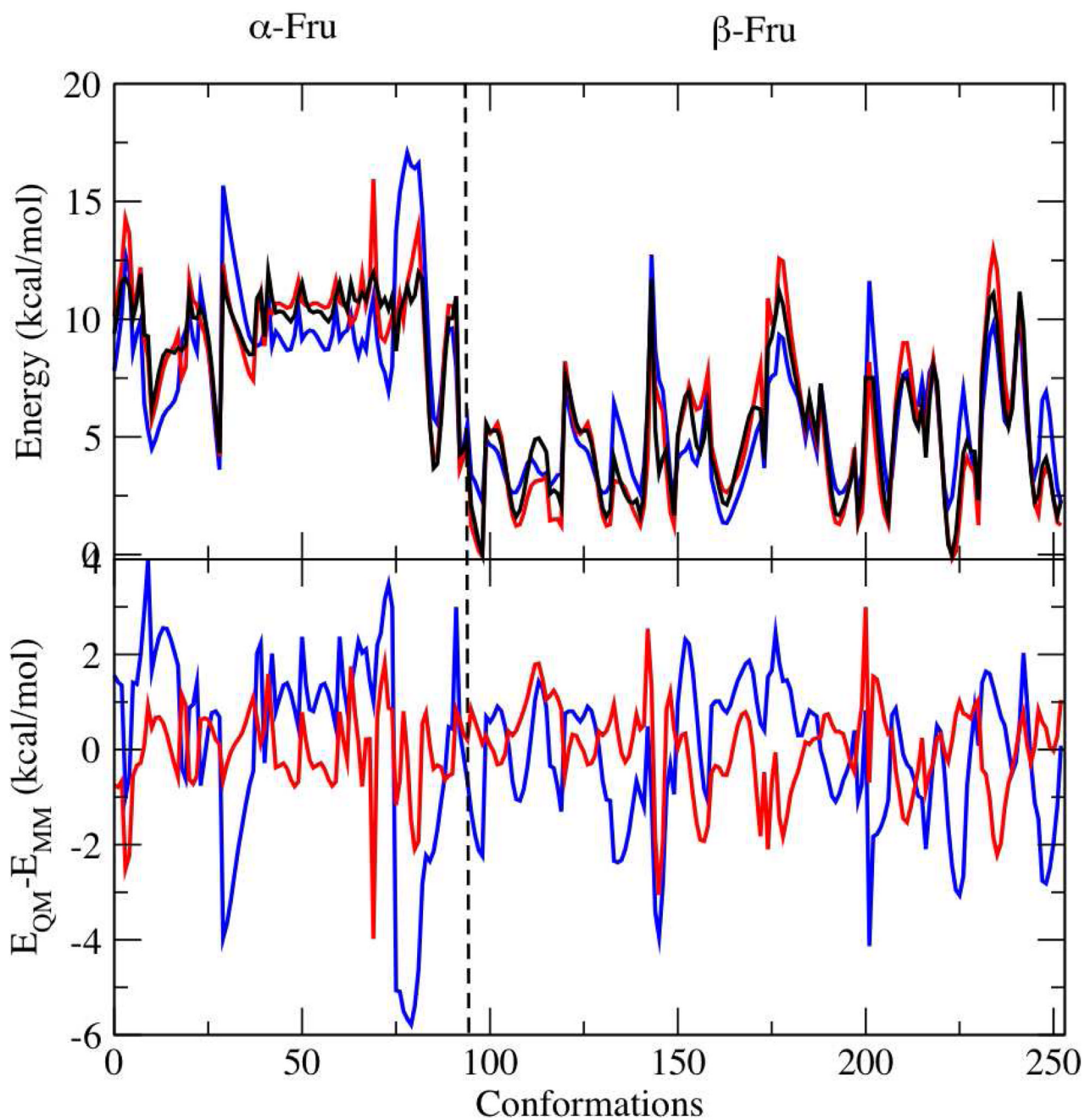
**Figure 4.**  
Definition of the gg, gt and tg exocyclic C3-C4-C5-O5 rotamer conformations.



**Figure 5.** Potential energy scans of approximately 1100 conformations of  $\alpha$ -arabinofuranose ( $\alpha$ -Ara),  $\beta$ -arabinofuranose ( $\beta$ -Ara),  $\alpha$ -ribofuranose ( $\alpha$ -Rib),  $\beta$ -ribofuranose ( $\beta$ -Rib). QM potential energies are cut off at 12 kcal/mol. QM is black, MM using optimized parameters is red and MM using unoptimized parameters is blue. Differences in energy ( $E_{QM} - E_{MM}$ ), shown in the bottom plot, for MM using optimized parameters is red and MM using unoptimized parameters is blue.

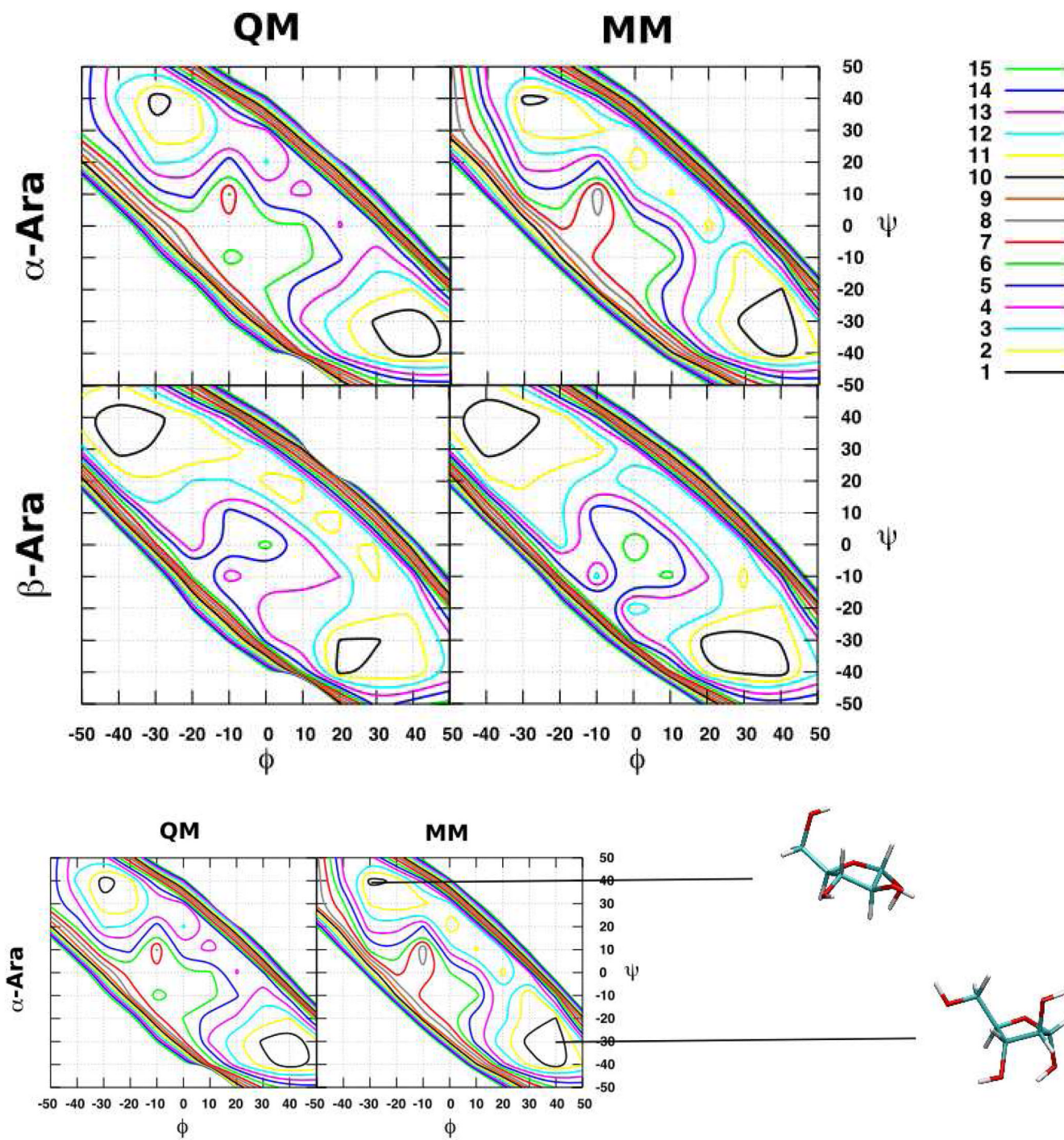


**Figure 6.** Potential energy scans of approximately 450 conformations of  $\alpha$ -deoxyribofuranose ( $\alpha$ -Deo) and  $\beta$ -deoxyribofuranose ( $\beta$ -Deo). QM potential energies are cut off at 12 kcal/mol. QM is black, MM using optimized parameters is red and MM using unoptimized parameters is blue. Differences in energy ( $E_{QM} - E_{MM}$ ), shown in the bottom plot, for MM using optimized parameters is red and MM using unoptimized parameters is blue.



**Figure 7.**

Potential energy scans of approximately 250 conformations of  $\alpha$ -fructofuranose ( $\alpha$ -Fru) and  $\beta$ -fructofuranose ( $\beta$ -Fru). QM potential energies are cut off at 12 kcal/mol. QM is black, MM using optimized parameters is red and MM using unoptimized parameters is blue. Differences in energy ( $E_{QM} - E_{MM}$ ), shown in the bottom plot, for MM using optimized parameters is red and MM using unoptimized parameters is blue.



**Figure 8.**  
3-dimensional potential energy scans of the  $\phi = \text{O4-C1-C2-C3}$  and  $\psi = \text{C1-C2-C3-C4}$  ring pucker torsions of  $\alpha$ -arabinofuranose ( $\alpha$ -Ara) and  $\beta$ -arabinofuranose ( $\beta$ -Ara).

Table 1  
 Comparison of water interaction energies and O---H distance by HF/6-31G(d) QM calculations and CHARMM force field for the pair conformations shown in Figure 3.

		$E_{QM}$ (kcal/mol)	$E_{MM}$ (kcal/mol)	$E_{MM} - E_{QM}$ (kcal/mol)	$R_{QM}$ (Å) <sup>a</sup>	$R_{MM}$ (Å)	$R_{QM} - R_{MM}$ (Å)
$\alpha$ - Arabinofuranose	Pair 1	-4.34	-3.98	0.36	1.82	1.72	-0.10
	Pair 2	-3.50	-2.55	0.95	1.87	1.78	-0.09
	Pair 4	-5.53	-4.64	0.89	1.78	1.73	-0.05
	Pair 5	-6.21	-6.49	-0.28	1.84	1.74	-0.10
	Pair 6	-6.67	-7.32	-0.65	1.82	1.72	-0.10
	Pair 7	-5.09	-4.18	0.91	2.82	2.95	0.13
	Pair 8	-6.58	-7.78	-1.21	1.89	1.77	-0.12
	Pair 1	-4.04	-4.29	-0.25	1.80	1.69	-0.11
$\beta$ - Arabinofuranose	Pair 5	-4.87	-4.67	0.20	1.94	1.82	-0.12
	Pair 6	-6.11	-6.87	-0.76	1.89	1.75	-0.13
	Pair 7	-6.31	-6.47	-0.16	2.23	2.18	-0.05
	Pair 8	-3.36	-2.58	0.78	1.91	1.83	-0.08
	Pair 4	-1.51	-1.41	0.10	1.95	1.78	-0.17
	Pair 5	-6.30	-7.14	-0.84	1.98	1.81	-0.17
	Pair 6	-5.21	-5.84	-0.63	2.00	1.83	-0.17
	Pair 7	-5.72	-5.87	-0.15	2.01	2.14	0.13
$\beta$ -Ribofuranose	Pair 8	-3.86	-4.14	-0.28	2.08	1.94	-0.14
	Pair 3	-1.16	-1.25	-0.09	2.18	2.08	-0.10
	Pair 5	-5.99	-5.84	0.15	1.84	1.75	-0.09
	Pair 6	-6.79	-7.54	-0.75	1.81	1.71	-0.10
	Pair 7	-5.62	-4.67	0.94	2.69	2.81	0.12
	Pair 8	-5.94	-6.04	-0.10	1.84	1.76	-0.08
	Pair 2	-4.04	-3.09	0.95	1.87	1.77	-0.10
	Pair 5	-6.37	-6.51	-0.15	1.83	1.73	-0.10
$\alpha$ -Xylofuranose	Pair 6	-5.57	-5.48	0.09	1.84	1.74	-0.10
	Pair 7	-5.60	-3.98	1.62	2.54	2.78	0.24
	Pair 8	-5.97	-6.66	-0.69	1.95	1.83	-0.11

	$E_{\text{QM}}$ (kcal/mol)	$E_{\text{MM}}$ (kcal/mol)	$E_{\text{MM}} - E_{\text{QM}}$ (kcal/mol)	$R_{\text{QM}}$ (Å) <sup>a</sup>	$R_{\text{MM}}$ (Å)	$R_{\text{MM}} - R_{\text{QM}}$ (Å)
Average			0.04			-0.07
RMSE			0.68			0.12

<sup>a</sup> $E_{\text{QM}}$  values are scaled by 1.16, and 0.2 is subtracted from the RQM values



**Table 2**  
Crystal lattice parameter and volumes calculated from crystal simulations.<sup>39</sup>

Compound	CSD ID	R factor	A	% error	B	% error	C	% error	Volume	% error
Me- $\alpha$ -Arabinofuranoside	JUQSET	0.06	12.26		6.83		4.60		378.17	
			12.92	5.41	6.79	-0.62	4.60	-0.08	390.62	3.29
Me- $\beta$ -Arabinofuranoside	QIBTIE	0.10	5.90		10.79		12.15		773.748	
			6.71	13.68	9.41	-12.78	12.42	2.23	783.66	1.28
Me- $\alpha$ -Lyxofuranoside	JUQSOD	0.03	10.34		15.50		4.63		741.18	
			10.49	1.45	15.83	2.10	4.59	-0.78	762.54	2.88
Me- $\beta$ -Ribofuranoside	ZOWJAW	0.03	4.86		24.16		12.88		1511.84	
			4.78	-1.59	25.82	6.88	12.99	0.88	1603.90	6.09
Me- $\alpha$ -Xylofuranoside	DMANTL07	0.04	6.22		8.14		7.42		369.11	
			6.50	4.53	8.20	0.72	7.49	1.00	389.82	5.61
Average				4.70		-0.74		0.65		3.83

**Table 3**  
Calculated solution densities at different concentrations of monosaccharides in a box of water with 1100 water molecules at  $T = 298\text{K}$  and  $P = 1$  atm.

Compound	Molality (mol/kg)	$N_{\text{solite}}$	Expt (g/cc) <sup>48</sup>	Calc (g/cc)	% error
Arabinofuranose	0.10	2	1.003	1.017	1.40
	0.20	4	1.008	1.021	1.29
	0.30	6	1.014	1.025	1.08
Fructofuranose	0.10	2	1.004	1.018	1.39
	0.20	4	1.011	1.023	1.19
	0.30	6	1.017	1.028	1.08
Lyxofuranose	0.20	4	1.008	1.021	1.29
	0.30	6	1.013	1.025	1.18
	0.36	7	1.017	1.027	0.98
Ribofuranose	0.10	2	1.003	1.017	1.40
	0.20	4	1.008	1.021	1.29
	0.30	6	1.013	1.025	1.18
Xylofuranose	0.10	2	1.008	1.017	0.89
	0.20	4	1.013	1.021	0.79
	0.30	6	1.017	1.025	0.79
Average					1.15

Table 4

Comparison of computed pseudorotational angles and populations of the North (N) and South (S) conformers to experimental estimates from a combined NMR and PSEUROT analysis.<sup>50</sup>

Pseudorotation angle and amplitude	Expt. P <sub>N</sub>	Calc. P <sub>N</sub>	Expt. P <sub>S</sub>	Calc. P <sub>S</sub>	Difference	Avg. Φ <sub>m</sub>
Me-α-arabinofuranoside	44	61	123	128	5	40
Me-β-arabinofuranoside	353	345	162	183	21	39
Me-α-ribofuranoside	119	47	125	135	10	41
Me-β-ribofuranoside	338	351	85	198	113	40
Me-α-lyxofuranoside	20	41	219	125	-94	41
Me-β-lyxofuranoside	349	340	305	209	-96	40
Me-α-xylofuranoside	324	20	124	125	1	40
Me-β-xylofuranoside	348	342	265	225	-40	38

Pseudorotation angle populations	Expt. %N	Calc. %N	Expt. %S	Calc. %S	Difference
Me-α-arabinofuranoside	39	31	61	69	8
Me-β-arabinofuranoside	87	62	13	38	25
Me-α-ribofuranoside	4	17	96	83	-13
Me-β-ribofuranoside	86	86	14	14	0
Me-α-lyxofuranoside	65	59	35	41	6
Me-β-lyxofuranoside	77	99	23	1	-22
Me-α-xylofuranoside	8	81	92	19	-73
Me-β-xylofuranoside	66	98	34	2	-32

**Table 5**  
Exocyclic rotamer population compared to experimental NMR.<sup>7,1,1,51,52</sup>

	Expt. <i>gg</i>	Calc. <i>gg</i>	Diff.	Expt. <i>gt</i>	Calc. <i>gt</i>	Diff.	Expt. <i>tg</i>	Calc. <i>tg</i>	Diff.
Me- $\alpha$ - arabinofuranoside	48	49	1	38	47	9	14	4	-10
Me- $\beta$ - arabinofuranoside	35	30	-5	57	64	7	8	6	-2